Acta Linguistica Hungarica

AN INTERNATIONAL JOURNAL OF LINGUISTICS

ACTA LINGUISTICA HUNGARICA

AN INTERNATIONAL JOURNAL OF LINGUISTICS

ACTA LINGUISTICA HUNGARICA

AN INTERNATIONAL JOURNAL OF LINGUISTICS

PREFACE

The present issue of Acta Linguistica Hungarica contains four papers directly or indirectly related to semantics.

Gábor Alberti's paper is devoted to the problem of compositionality in its relation to Discourse Representation Structures (DRS). It is pointed out that DRS can be made compositional by constructing a strictly compositional syntax. Furthermore, DRS can be simplified if the content of discourses is embedded into a model of the hearer's information state. Such a level of representation provides a uniform treatment of conditionals, canonical scenarios, universal, modal, non-factive and negative contexts.

László Hunyadi addresses the question of how and to what extent syntax, prosody, communicative and logical functions are related. The paper outlines the metrical syntax of Hungarian, the main claim being that there is a close relationship between basic sentence structure and prosodic structure. In the sentence three prosodically and communicatively distinct parts can be identified. The communicative structure of the sentence determines the relative scope of operators and is thus closely related to logical interpretation.

Márta Maleczki investigates the problem of indefinite subjects. Indefinite subjects may have a weak (non-specific) and a strong (specific) interpretation. It is shown that the specifying property of predicates can be derived from the telic or place-bounded character of the predicate. However, subjects, too, may contribute to the interpretation. This means that for the full range of interpretation possibilities a more general criterion is called for. This criterion, called General Specifying Criterion, states that each statement must have at least one specifying feature. Specifying features for subjects are (i) strong determiners and (ii) strong interpretation of weak subjects; specifying features for predicates are (i) telicity and (ii) locatedness.

Ildikó Tóth discusses the licencing conditions for negative polarity items (NPI) in Hungarian. Two types of NPIs are distinguished and described. It is shown that they involve different licencing mechanisms, both crucially depending on their indefiniteness. In other words, the author claims that analyzing negative polarity

items not as quantifiers but as expressions associated with free variables which can be bound by a nonveridical operator is a desirable move towards understanding their behaviour. The proposal put forward in the paper relies on inherent features of negative polarity items and is in several respects superior to earlier treatments of these items.

We would like to express our gratitude to Anna Szabolcsi who provided help in devising the contents of this issue and who has also been involved in the refereeing procedure. Without her invaluable help this particular collection of papers would never have been compiled and presented as it is here.

The editor

GENERATIVE ARGUMENT STRUCTURE GRAMMAR: A STRICTLY COMPOSITIONAL SYNTAX FOR DRS-TYPE REPRESENTATIONS

GÁBOR ALBERTI

Abstract

This paper presents arguments in favour of the **representational** character of the (original) Kamp–Heim Theory, which is frequently criticized nowadays because of the uncertain status of DRSs and the absence of **compositionality** in the strictest sense. I point out that there is a natural syntax according to which the simplest class of DRSs can be constructed in a compositional way; and logical-formula-like DRSs can be dispensed with if the **hearer's information state** is represented as one huge complex DRS instead of assigning DRSs to discourses. This representational level sheds new light on stubborn problems due to the fact that referents, propositions and worlds are defined by simultaneous recursion.

1. Compositionality and representationalism

1.1. Compositionality

This article is devoted to a crucial question of dynamic semantics, viz. the principle of **compositionality**.[1]

The principle comes from the Western tradition of logic, which is based on the approach that the semantic interpretation of any statement or predicate logical formula is obtained via a systematic semantic procedure interpreting its parts and the logical symbols connecting them in models. The principle of compositionality, attributed to Frege, is a way of articulating this systematic correspondence between syntax and semantics (Partee *et al.* 1990):

D1 **The Principle of Compositionality:** The meaning of a complex expression is a function of the meanings of its parts and of the syntactic rules by which they are combined.

Partee *et al.* (1990) say that "construed broadly and vaguely enough, the principle is nearly uncontroversial," but they add that "Montague's precise version of it

places rather severe constraints on admissible systems of syntax and semantics."
Montague (Dowty *et al.* 1981) was the one "whose papers showed once and for all
that the model-theoretic approach towards natural language was viable. These
papers also demonstrated that the model theories for natural and those for symbol-
ic languages have a great deal in common" (Kamp–Reyle 1993).

1.2. Discourse semantics

In the extension of model-theoretic semantics from the sentential to the discourse
level, however, theories have emerged (e.g. Kamp 1981; Kamp–Reyle 1993; Heim
1982; 1983) that are expressly non-compositional. Kamp and Reyle (1993), for
instance, only insist on "...uncovering the systematic [!] correlation between mean-
ing and syntactic form," instead of insisting on **compositionality** in a strict sense.

The simple two-sentence discourse in (1a) below illustrates the nature of the
problem with compositional representation in a simple predicate logic:

(1)   (a) A man walks in the park. He whistles.
      (b) ∃x[man(x) & walk_in_the_park(x)] & whistle(x)

While the male pronoun in the second sentence clearly refers to the walking man
mentioned in the first sentence, the formula in (1b), which is the straightforward
predicate logical representation of the two sentences, cannot express this meaning.
The last occurrence of variable x has nothing to do with the second and third occur-
rences of x because they are **bound variables** (bound by the existential quantifier ∃,
in the scope of which they are) whereas the last x is a **free variable**. In traditional
predicate logic there is no way to identify a free variable with a preceding bound
variable (i.e. to make sure that they take the same value) even if they are occur-
rences of the same symbol. And the problem is not only technical at all. The for-
mula associated with the first sentence expresses only the existence of at least one
walking man, which is correct in a truth-conditional sense since the reference to a
walking man does not exclude either other walking men, or men that are not walk-
ing, or young male persons, from the model. It cannot be explained, then, why *he*
can refer to an arbitrary member of a (perhaps large) group of male people as well
as if there were only one in the model.

The solution that the **Kamp–Heim Theory** (e.g. Kamp 1981; Kamp–Reyle
1993; Heim 1982; 1983) offers depends on the representation of partial models that
contain **discourse referents** (Karttunen 1976). The partial model assigned to the
first sentence of (1a) is a very small world with two referents, the former being a
man, the latter a park, and the former walking in the latter:

(1) (c) $<\{u, v\}, \{man(u), park(v), walk\_in(u,v)\}>$

In this approach *he* can easily find the referent u for it is the only male object in the small world; now the existence of a lot of walking men in the complete world is irrelevant. The **discourse representation structure** (DRS) of the two-sentence discourse under examination is the following:

(1) (d) $<\{u, v\}, \{man(u), park(v), walk\_in(u,v), whistle(u)\}>$

Though Kamp himself called these DRSs "partial models, typically with small finite domains" in the introduction of his 1981 paper and retains this approach later again (Kamp–Reyle 1993), DRSs behave as formulas of a special logical syntax at the same time. These formulas, for instance, can contain logical connectives (e.g. $\Rightarrow$, v; see **1.1**). Thus, simple DRSs can still be assimilated to partial models but complex DRSs can only be read as formulas (because of the logical connectives mentioned above).

1.3. Compositionality vs. representationalism

As pointed out by Groenendijk and Stokhof (1989), DRSs as formulas are non-compositional, at least in the strict Montagovian sense (which is a widely accepted specification of the informal formulation in D1). The DRS in (1d) serves as an illustration: the (unordered!) set of statements (about certain referents standing in certain relations and similar facts) does not constitute units that would correspond to units of a syntax.

Groenendijk and Stokhof (1989) argue that there are methodological, philosophical, empirical, computational and practical reasons "to be interested in trying to keep to compositionality," not to mention the fact that "a semantic theory such as Montague grammar, and an approach like Kamp's discourse representation theory, are hard to compare. ... One of the main obstacles is that the latter lacks, c.q. has abolished, the principle of compositionality, which is so central a feature of the former."

The authors have developed a compositional alternative to Kamp's discourse semantics, in which the formula in (1b) above, compositional according to a traditional predicate logical syntax, provides an appropriate reading.[2] What is not traditional in the method is a computational approach known from the semantics of programming languages, according to which a variable bound by an existential quan-

---

[2] There are also other directions of the efforts to find alternatives to DRT where traditional assumptions are insisted on to a greater extent (e.g. E-type approaches; Heim 1990).

tifier retains its last value after the existential formula's being verified. In the case
of formula (1b), for instance, if it is the value *Peter* ($x=Peter$) that has been chosen
to verify the existential statement, then this value remains that of the free occur-
rence of x, yielding the statement *whistle(Peter)*. In this **dynamic predicate logic**
(DPL) thus certain formulas receive a **dynamic interpretation**, which has been
intended to substitute for representations. To sum up, DPL could incorporate most
of the dynamic intuitions of the original Discourse Representation Theory but at the
cost of abandoning DRS-type representations altogether.

A non-representational theory, however, will face to problems concerning def-
inite descriptions without any explicit antecedent. Kálmán's (1990) examples serve
as an illustration:

(2)   (a) Joe got married yesterday. The priest spoke very harshly.
       (b) Joe got married yesterday. ???The dog barked very loudly.

Kálmán attributes the radical difference between (2a) and (2b) to the fact that, under
standard assumptions on cultural background, the hearer will know what priest the
speaker has in mind in (2a), whereas (s)he will have no idea of what *the dog* in (2b)
refers to. In a representational semantics a phenomenon like this can be accounted
for by turning DRSs into greater DRSs on the basis of the hearer's encyclopedic
knowledge.[3] The DRS in (2c) below, for instance, **licenses**[4] the (simplified) DRS
in (2d), which already contains a priest (marked with referent w) that can be
referred to:

(2)   (c) <{u}, {get_married(u)}>
       (d) <{u, v, w}, {get_married(u), marry(u, v), make_man_and_wife(w,u,v), priest(w)}>

To sum up, representational semantic systems seem to be easier and more effective
to account for linguistic phenomena but at the cost of being non-compositional (at
least in the strict sense that many insist on).

---

[3] A solution like this is not in accordance with the spirit of the non-representationalist DPL
because the cornerstone of this theory is that a value is retained if it has been taken explicitly. The
priest referent, however, cannot be taken explicitly after the utterance of the first sentence. Nor can it
be taken even due to some implicit mechanism because the existence of a priest is not a logical con-
sequence, but only a **licensed** possibility (see also the following footnote).

[4] This is not a **logical inference** because a civil official also has the right to marry couples, but
a possibility available in the hearer's encyclopedic knowledge, which is obviously relevant to lan-
guage usage. There is no "relevant dog" associated with marriage, however, at least in our culture.

1.4. Dynamic semantic systems

The fundamental papers on discourse semantics discussed so far share the following three properties.

I. The authors regard the meaning of a sentence not lying (only) "in its truth conditions but rather in the way it changes (the representation of) the information of the interpreter" (Groenendijk–Stokhof 1990). This approach gives the **dynamic** character to these semantics.

II. Nevertheless, the picture of the hearer's **information state** is oversimplified and in certain areas simply counter-intuitive. Kamp and Reyle (1993, 93) admit having adopted "models [in which DRSs are to be evaluated] in order to avoid inopportune questions about possible worlds," and they add that "this is standard practice in formal logic." The problem with regarding information states as **models** (Partee *et al.* 1990) is that models are **total** semantic objects in the sense that it is supposed to be known in the case of a model for each n-tuples of referents belonging to its universe whether or not they stand in a certain relation. "Models leave no relevant information undecided" (Kamp–Reyle 1993). Hearers, however, practically always have a partial knowledge. As Groenendijk and Stokhof (1989) also use models as information states, they are to regard atomic statements (e.g. *Peter loves Mary*) as **tests**, which means that an assertion heard is supposed to be either corroborated or rejected. The typical case is excluded: to regard an assertion as a **new** piece of information to the hearer.[5]

III. Although compositionality plays the main role in Groenendijk and Stokhof's (1989) paper, a relevant aspect of the problem is totally ignored, viz. **syntax**. What any version of the Principle of Compositionality (D) declares is nothing else but a tight systematic correspondence between syntax and semantics. Compositionality of a semantic system is always to be interpreted as 'compositionality according to some kind of syntax.' The syntax that Groenendijk and Stokhof (1989) consider is the most traditional predicate logical syntax whose compositional construction from a generative syntax of natural language (e.g. Partee *et al.* 1990) is far from simple even in small fragments of the English language (λ-abstraction, proliferation of types). As for Kamp and Reyle (1993, 18), their "choice of syntactic theory has been guided by opportunism." Thus the choice of syntax seems to be another oversimplified part of dynamic semantics.

---

[5] It is not evident that the above mentioned version of DPL can be embedded in a richer version where phenomena like this have some kind of dynamic effect.

1.5. Proposals: A compositionally adequate syntax for DRT, DRS-like information states

My main purpose is to prove that a dynamic semantics based on DRS-like representations need not necessarily be thought to be non-compositional. It is not excluded at all that there is an alternative to the anti-representationalist move. Compositionality is a matter of the relationship between syntax and semantics, so non-compositionality of a theory can be fixed in at least two ways: by changing the semantics or by changing the syntax. I argue that what is required is an appropriate syntax, relative to which DRS-like representations are compositional, in the strictest sense. The attractive discourse representation structures themselves inspired me first to raise the possibility of such a syntax (Alberti 1990). The version to be presented in this paper has been called **Generative Argument Structure Grammar**, (GASG) because of the distinguished role of a comprehensive lexical characterization of argument structures.

In this approach a (simple) DRS is to be regarded as the semantic side of a list of lexical relation names with referents, which is nothing else but a list of statements concerning facts that certain referents stand in certain relations; whereas the syntactic side of this list of lexical relation names is the surface form of the sentence, which is built up by (lexical features of) these relation names themselves. No phrase structure rules are required since each lexical item determines its own environment. Thus, GASG can be regarded as a special kind of categorial grammar (Oehrle *et al.* 1988), as the grammaticality and interpretability of sequences of (inflected) words depend on whether they satisfy the environmental requirements described in the lexical characterizations of each other (concerning not only conventional syntactic and categorial factors but other external features as well, such as morphological and intonational ones).

A sketchy analysis of a Hungarian sentence in Section 2 serves the purpose of outlining and exemplifying the leading ideas of this generalized categorial grammar. In that section I attempt to convince the reader that these ideas are viable and immediately motivated by the desire to cater to DRT. The reader may consult the Appendix if (s)he is interested in details.[6]

Another integral part of my approach (Section 3) is the claim that we should return to Kamp's (1981) original intuition on DRSs, i.e. they are **partial models**, rather than formulas of a special logical syntax or of anything else whose theoreti-

---

[6] A syntactic review of a couple of hot topics of Hungarian generative linguistics (arguments and adjunct, bracketing problems, neutral sentences, definiteness effects, focus constructions, elliptical constructions) is available in Alberti (1996). Alberti (1998) provides an improved version of GASG.

cal status is quite uncertain (Kálmán p.c.). The cost is that it is the hearer's information state that is to be defined as a single huge richly structured DRS, being built through the hearer's whole life, in which a discourse is embedded from sentence to sentence where each sentence feeds a couple of partial models (i.e. simple DRSs) to this gigantic DRS. Thus a discourse will not be assigned a usual box structure; instead, it is the hearer's information state that can be regarded as a gigantic complex DRS, which contains a partially ordered system of worlds.

I argue that logical connectives can entirely be dispensed with due to a simultaneously recursive definition of referents, statements about relations among these referents, and the structure of possible worlds at the hearer's disposal. This construction can be regarded as a powerful generalization of the usual separation of variables from constants in that constants that belong to a particular world can serve as variables to a formula that belongs to another world (in a precise sense to be defined in Section 3). Hence, defining the hearer's information state, which might seem to be a burden to us at first sight, sheds new light on stubborn problems. A series of famous examples of dynamic semantics will illustrate this point.

At the beginning of Section 2 I am going to mention an example which simultaneously illustrates almost every basic problem discussed in this paper. Its entire formal analysis (from syntax to embedding in the hearer's information state) is available in the Appendix.

2. Generative Argument Structure Grammar

This section is devoted to a brief demonstration of a grammar whose leading ideas are motivated by the desire to cater to DRT directly, on the one hand, and appear independently in different current linguistic theories, on the other. It will be pointed out that this grammar corresponds to DRS-like representations in the same way as constituent structure trees correspond to predicate logical formulas and Montagovian representations.

The principles will be exemplified by a partial analysis of the first clause of the sentence in (3) below (an entire analysis of the entire sentence is available in the Appendix). Here I attempt to provide technical details just enough to convince the reader that the grammar is viable and is to be regarded as a serious alternative to current generative grammars. Nevertheless, a certain amount of (non-conventional) formalism concerning syntax and morphology is to be introduced because it is relevant to semantics that there is a compositionally adequate syntactic, or rather, **external**, counterpart to an attractive semantic theory.

(3)   Ha 'Mari-nak 'nősül egy 'régi 'udvarló-ja, akkor 'fel-keres-i a 'menyasszony-t...
      if Mari-dat marryawoman-3sg an old suitor-poss3sg then pref$_{up}$-seek-3sg-defObj the fiancée-acc
      'If a former boyfriend of Mary's is getting married, she visits the fiancée...'
      (e.g. in order to talk to her about the man's bad habits)

2.1. The condition of lexical inclusiveness

The task of the syntax-semantics interface in this context is to establish "... the right correlations between schematic discourse referents [i.e. arguments of lexical entries' conceptual structures] and real discourse referents" (Kamp–Rossdeutscher 1994, 159). The authors also say that "...in essence, [its mechanism] will always be the same and will make use of the same interface information that our lexical entries encode." It is to be noted that the condition of **Lexical Inclusiveness** of the Minimalist Hypothesis (Chomsky 1995) declares the same principle: output representations consist of nothing beyond properties of items of the lexicon (i.e. lexical features); in other words, the interface levels consist of nothing more than arrangements of lexical features.

2.2. A homogeneous representational level of different kinds of external grammatical features

Hence, phrase structure rules, held formerly as basic source of creativity, have become redundant, and have been expressly suggested to be eliminated (Chomsky 1995). Movement, another central concept of transformational generative linguistics, has remained an integral part of the theory in the following form: it is assumed to be driven by morphological checking requirements, that is, certain lexical features are to check each other in syntax in special checking configurations. In the light of the condition of Inclusiveness, this approach is to be regarded as an explicit **syntactic encoding** of morphological information. Similar considerations have led me to the conclusion that intonational information (as well as other kinds of external information) is also assumed to be represented in a syntactically encoded form (Alberti 1996).

      In this model, thus, syntax (i.e. constituent structure trees with traces and empty functional heads) serves as a uniform level of representation of pieces of **external information** of different kinds (morphological, intonational, categorial, and, of course, syntactic kinds). In this respect what the Minimalist Theory manifests is a radically homogeneous arrangement of external linguistic information. A similarly radical alternative on the opposite side would be a level of representation where each piece of morphological, intonational, or syntactic information is represented as a piece of morphological, intonational, or syntactic information, respectively. That is what I am going to suggest now.

2.3. Unordered structure

The question as to which kind of representation is more suitable for the separation of **grammatical** sequences of words of a language from ungrammatical sequences is irrelevant now. Both representations are to contain the same information on external features of sequences of inflected words (called **numerations** by Chomsky 1995). What is relevant now, however, is that a (simple) DRS contains an **unordered** set of statements about referents, i.e. no order is imposed on its elements (Kamp–Reyle 1993, 121), whereas constituent structure trees, as well as predicate logical formulas, are linearly ordered constructions. It is not an accident, thus, that the possible compositional semantic counterparts of transformational generative syntactic representations are hypothesized to be based on traditional predicate logical formulas (Groenendijk–Stokhof 1989, 1990).

Obviously, DRS-like representations require an entirely different syntax. The version I am proposing is based on the idea that the syntactic, or rather, **external**, structure of a sentence consists of an unordered (!) set of lexical items that satisfy the environmental requirements desribed in the lexical characterizations of each other. The grammar is characterized by a **parallel** semantic and external description of the argument structure of words. This fact is the explicit guarantee for compositionality and the justification for the name **Generative Argument Structure Grammar**. As for classification, it seems that GASG belongs to the family of categorial grammars but differs from the classical versions in that the 'category' of a word contains references to certain features of not only its neighbors in a sentence but any word. A verb, for instance, may require the presence of a word marked with, say, the dative case in the sentence, without being sensitive to its precise position. Naturally, it is not excluded either that a word is sensitive to certain features of its left neighbor.[7]

2.4. Co-predication

In GASG a sentence can be assigned a DRS, on the one hand, which is an unordered set of statements concerning relations among referents, and an external representation, on the other hand, which is an unordered set of statements concern-

---

[7] Another deviation from classical categorial theories is that the satisfaction of categorial requirements of words in a sentence is not to be verified by building a constituent structure tree. In this respect, and in several other respects, GASG is similar to Kálmán and Rádai's (1996) 'construction grammar.' The only clear difference is that this construction grammar is based on external constructions or patterns stored in the lexicon whereas in my system constructions manifest themselves as lexical redundancy rules. As for catering to DRT, it is not evident to me that each construction is (to be) associated with some semantic content whereas in GASG each external relation between two words is assumed to be associated with a semantic relation. As has been mentioned, it is this strict parallelism that is the guarantee for compositionality.

ing relations among words of a sentence. What an external relation between two words expresses is that the one satisfies the description of a potential participant in the lexical characterization of the other. Thus there is an immediate mapping between semantic relations and external relations. The cornerstone of the implementation of a system like this is an appropriate formulation of these bilateral relations. What does it mean that two words **stand in a semantic relation** (encoded by an external relation)? The starting-point is **predicate-argument relation** as usual, but instead of saying that a word selects another word as its argument, it is to be said that there are two words **predicating of one and the same referent**. A relation between words like this will be referred to as **co-predication**.[8]

2.5. Starting-point of analysis: an underspecified DRS

As has been promised, a fragment of the analysis of the sentence in the Appendix illustrates here the crucial points (the strange indices are to be attributed to this incompleteness). A Hungarian sentence has been selected primarily in order to demonstrate the good capacity of GASG for processing morphological information.[9]

(4)   (a) 'Mari-nak 'nősül egy 'régi 'udvarló-ja.
            Mari-dat marryawoman-3sg an old suitor-poss3sg
            'A former boyfriend of Mary's is getting married.'

The meaning of a sentence depends on its words and the grammatical relations among these words so the first task is to collect the lexical relation names that belong to the words. As for the grammatical relations, their role is to be determined later, on the basis of the (morpho-)lexical characterizations of the inflected words, according to the principle of Lexical Inclusiveness (**2.1**). The set of the lexical relation names can already be regarded as a DRS, which is still underspecified: the

---

[8] The relation that a verb bears to its object in DRT, for instance is an obvious instance of co-predication. Both the verb and the object are to be represented as predicators, and their relation manifests itself in the fact that the argument slot of the predicator that belongs to the object is occupied by the same referent as one of the argument slots of the predicator that belongs to the verb. Williams (1995, 122–3) also refers to a similar relation between words in a transformational generative framework: "In this conception, one thinks of the theta roles as referring, and the NPs that they are assigned to as 'conditioning' that reference. Coreference, 'linking,' and the other binding theory relations would be relations among theta roles [which roughly correspond to referents] [instead of relations among NPs]."

[9] The linguistic data and observations on Hungarian encoded in the lexical characterizations to be discussed are partly evident, and partly due to researches in the 80's primarily in the framework of the Government and Binding Theory (Chomsky 1981). A book edited by Kiefer and É. Kiss (1994) provides a good summary on the latter part.

argument slots are filled with different objects, say, variables. The task of the syntax-semantics interface (2.1) is to substitute discourse referents for these variables. What depends on the grammatical relations of the sentence is which variables are to be replaced with the same discourse referent; and this question can be decided just on the basis of external relations, i.e. instances of co-predication (2.5). The underspecified DRS contains the information that somebody is named Mary, somebody (a man) gets married, and somebody is a boyfriend, in addition to pieces of information that cannot be interpreted at this level.[10]

(4)   (b)   {MARI(v_3), MARRY-A-WOMAN(v_4), AN(v_5), FORMER(v_6), BOYFRIEND$_{poss}$(v_7, v_8)}

2.6. Lexical characterizations

Instances of co-predication are to be revealed on the basis of external relations, and it is the lexicon that mediates between the former and the latter. The characterization that belongs to a lexical item is to contain the elementary DRS referred to by this item, on the one hand, and some kind of reference to the words in a potential sentence with which they are to stand in external relations, on the other. Each external relation "predicted" in the morpholexical characterization which is found in an actual sentence explicitly refers to a semantic relation, i.e. an instance of co-predication. It is in this way that the lexicon ensures compositionality.

First of all, however, let us consider the lexical characterizations in the order of the words in the sentence to be analyzed.

(5)   (a)   MARI(v_3)
           'Mari-dat: $<t^3>$
           CAT.N.PROPN(t^3), LEG.REF.SPEC.DEF(t^3), MOR.CASE.DAT(t^3), INT.STRESS(t^3)

The inflected word *Marinak* 'Mari-dat' is characterized as follows. As for its category (CAT), it is a noun (N), and specifically a proper noun (PROPN). What makes it **legitimate** (LEG) in a sentence (Alberti 1997a) is that it is a referential expression (REF), and specifically specific (SPEC) and, moreover, definite (DEF). As for morphology (MOR), it is marked with the dative (DAT) case (CASE). Finally, as for its intonational (INT) state, its actual occurrence is stressed (STRESS).

---

[10] I follow Kamp Reyle (1993) in treating proper names as predicates, which is often regarded as an objectionable practice unless it is made clear that the predicate in question semantically requires uniqueness and guarantees rigidity. In GASG, these conditions are satisfied in the course of the embedding of the semantic content of the sentence in the gigantic DRS that serves as the hearer's information state. It is at that level, too, that the uninterpretable pieces of information mentioned begin to function.

(5)   (b) MARRY-A-WOMAN(v_4)
          'marryawoman-3sg: $<t^4; ..., t^4{}_3, t^4{}_{31}>$
          CAT.V.INTR(t^4), FINITE(t^4), INT.STRESS(t^4), rec[INT.STRESS(t^4)],
          CAT.N($t^4{}_{31}$), MOR.CASE.NOM($t^4{}_{31}$), rec[INT.STRESS($t^4{}_{31}$)],
          MOR.PER.3($t^4{}_{31}$), MOR.NUM.SG($t^4{}_{31}$),
          LEG($t^4{}_3$), ~SYN.PREC($t^4{}_{31}$, $t^4{}_3$), SYN.NEAR($t^4{}_3$, $t^4{}_{31}$), ...

The lexical characterization of the finite (FINITE) intransitive verb (CAT.V.INTR) provides more information on the grammar. In addition to its **own word** (marked with t^4), there are references to other words to be found in its sentence. In this fragmentary analysis only the subject is discussed.[11]

Let us go on from row to row. The third formula in r1 means that the actual occurrence of this word is stressed (previous footnote!) while the similar fourth formula declares the general fact that being stressed is the default state of this verb or, in other words, it is a recessive condition on it, which can be overridden (e.g. in a focused sentence).

Rows 2–3 provide information on a noun (r2.1), marked with $t^4{}_{31}$, which is in Nominative (r2.2) and stressed as a default (r2.3). Further, it is to be a 3sg noun (r3).

The potential word marked with $t^4{}_3$ might seem to be a more mysterious thing. It is the word that **legitimize** (r4.1) the nominal element, which would remain predicative otherwise, viz. it manifests the requirement that the subject is to be referential. It would play the role of D in a DP in a transformational generative framework. An article can play this role in a sentence, but it may also occur that a proper noun plays both the role of the determiner ($t^4{}_3$) and that of the noun ($t^4{}_{31}$). Conditions r4.2–3 show that both cases have been taken into consideration: the potential noun does not precede the potential determiner and is near to it. These syntactic conditions are to be interpreted in the case of a separate determiner as follows: the noun immediately follows the determiner.

(5)   (c) AN(v_5)
          'a(n) : $<t^5; t^5{}_1>$
          CAT.DET.ART.IND(t^5), LEG.REF.NON-SPEC(t^5),
          ~INT.STRESS(t^5), ~INT.STRESS(t^5),
          CAT.N($t^5{}_1$), SYN.PREC(t^5, $t^5{}_1$), SYN.NEAR(t^5, $t^5{}_1$), MOR.NUM.SG($t^5{}_1$)

The third word is an indefinite article (r1.1), which is ambiguous (in Hungarian) wrt. specificity. This particular one here is assumed to be the non-specific version

---

[11] A few remarks are required at this point. See the **Notes on lexical characterizations in GASG** in the Appendix.

(r1.2), which has a separate lexical characterization (which is, naturally, closely related to that of the specific version in the morpholexical inheritance network (see footnote 11). Condition r1.2 also shows that this word is able to ensure a referential legitimacy to a nominal expression.

The funny second row says that this article is to be unstressed (in a grammatical sentence) (r2.1), and its actual occurrence is unstressed indeed. Otherwise, a sequence of words can be judged to be ungrammatical without delay.

An article requires the presence of a noun: t^5_1 refers to this potential noun (r3.1). The own word is to precede the noun (r3.2), which is required to be in singular (r3.4). Condition r3.3 is to be interpreted roughly as follows: the article is to be as 'near' to the noun as possible but if another word is **dominantly** required to be near to the noun then this latter word is to be inserted between the article and the noun (I mean an adjective).

(5) (d) FORMER(v^6)
     'former: $<t^6; t^6_1>$
     CAT.A(t^6),
     CAT.N(t^6_1), SYN.PREC(t^6, t^6_1), dom[SYN.NEAR(t^6, t^6_1)], ...

The (attributive) adjective (r1) requires the presence of a noun (r2.1), which follows it (r2.2). The requirement that the potential noun is to be near to the own word of the adjective (r2.3) is a dominant requirement, which is the expression of the observation that an adjective inserts between the article and the noun in a Hungarian DP, as has been mentioned just in the previous paragraph.

(5) (e) BOYFRIEND$_{poss}$(v^7, v^8)
     'boyfriend-poss3sg: $<t^7; t^7_1, t^7_{11}>$
     CAT.N(t^7), MOR.CASE.NOM(t^7),
     LEG(t^7_1), CAT.N(t^7_{11}), ~SYN.PREC(t^7_{11}, t^7_1), SYN.NEAR(t^7_{11}, t^7_1),
     MOR.PERS.3(t^7), MOR.NUM.SG(t^7), MOR.PERS.3(t^7_{11}),
     [MOR.CASE.NOM(t^7_{11}) ∨ MOR.CASE.DAT(t^7_{11})]
     dom[MOR.CASE.NOM(t^7_{11}) ⇒ (SYN.PREC(t^7_{11}, t^7) & SYN.NEAR(t^7_{11}, t^7))],
     rec[MOR.CASE.DAT(t^7_{11}) ⇒ (SYN.PREC(t^7_{11}, t^7) & SYN.NEAR(t^7_{11}, t^7))], ...

The last word is a noun again (r1.1) marked with the nominative case (r1.2). Due to its possessive suffix, its lexical characterization contains reference to a DP (in conventional terms), viz. both a determiner (t^7_1) and a noun (t^7_{11}) are required (r2.1–2). The determiner is to precede the noun and to be as near to it as possible unless they coincide (r2.3–4). The own word belongs to the possession, which is in 3sg (r3.1–2).

The potential possessor is required to be in its third person form, which is indicated by the possessive suffix. The number of the possessor is not determined. The complex formula in r4 says that the possessor is required to be marked with either nominative or dative. The (even more complex) last two formulas contain the information that the possessor immediately precedes the possession as a default but other factors (e.g. topicalization) may override this requirement if, and only if, the possessor appears in its dative form. Thus the external relation between possession and possessor in Hungarian is primarily encoded by morphological features.

2.7. Semantic relations and conglomerates of external means

It is worth mentioning at this point an advantage of the homogeneous representation of different kinds of external grammatical features (2.2): it is expressed explicitly that an external relation between two words is encoded by a structured conglomerate of pieces of information of different external kinds (syntactic, morphological, intonational, etc.). In this approach the difference between languages wrt. the correspondence of different conglomerates of external information to a fixed semantic relation can be regarded as an irrelevant superficial difference (e.g. the difference between configurational and non-configurational languages). What is relevant is that there must be enough information for the appropriate indication of external, and hence semantic, relations.

I hypothesize that there is a (presumably universal) set of semantic relations on the one hand, and there is an (also presumably universal) set of external relations on the other, and languages realize different mappings from one to the other, or rather, there is a mapping, for each language, from semantic relations to **sets** ('conglomerates') of external relations. This picture shows "the fundamental, or we might say, 'material' unity of different means of language" (Szépe 1964) where immediate syntactic relations (actual word order) do not seem to play a distinguished role. They are prepared to cooperate with non-syntactic external relations within a language and to encode a semantic relation encoded by non-syntactic relations in another language. This homogeneous model of external relations promises a grammar characterized by a flexible interaction of syntactic, morphological, intonational, categorial, etc. information, all these stored in the lexicon.[12]

---

[12] This homogeneous model of external relations differs from that of Lexical Functional Grammar (LFG; e.g. Bresnan 1982; Wescoat 1987; Bresnan Kanerva 1989). In LFG it is established that external structures vary across languages while internal [semantic] structures are largely invariant across languages, and this observation has led them to the conclusion that the amount of transparency [the idea that the internal and external structures must have the same form] might vary with the language type. This conclusion is undesirable from a metatheoretical point of view because it implies that the child's task in the course of language acquisition may differ in difficulty, or at least in

2.8. Compositional correspondence between GASG and simple DRSs

The parallel two systems of equations below serve as a formal demonstration of the compositional correspondence between GASG and simple DRSs. Let us consider the equations from row to row. The left column shows the satisfaction of the environmental requirements described in the lexical characterizations discussed above, i.e. the assignment of actual words to the potential ones.

$$
\begin{array}{ll}
(6) \quad & t^4_{\ 3} = t^5 \qquad\qquad v_4 = v_5 \\
& t^4_{\ 31} = t^7 \qquad\quad\ v_4 = v_7 \\
& t^5_{\ 1} = t^7 \qquad\qquad v_5 \quad v_7 \\
& t^6_{\ 1} \quad t \qquad\qquad\ v_6 \quad v_7 \\
& t_1 = t^3 \qquad\qquad v_8 \quad v_3 \\
& t^7_{\ 11} = t^3 \qquad\qquad v_8 \quad v_3
\end{array}
$$

$t^4_{\ 3}$ has been characterized in (5b) as a determiner-like element (r4.1) that precedes (r4.2) a noun marked with Nom. (r2.1–2). The indefinite article whose own word is t^5 (5c) meets this description. Hence, there is an external relation between the verb and the indefinite article, which corresponds to their semantic relation in the right column. The content of this relation is that 'a non-specific entity gets married.' This piece of information will be relevant in the course of the embedding of the immediate semantic content of the sentence in the hearer's information state.

$t^4_{\ 31}$ is required to be a 3sg noun marked with the nominative case (5b). Obviously, the own word of the boyfriend in (5e) meets these requirements. Hence, the argument slot of the verb and the first argument slot of the possession are occupied by the same referent, i.e. these two words are co-predicative. The semantic content is straightforward here: it is the boyfriend that gets married.

$t^5_{\ 1}$ is the noun in singular form that is preceded by the indefinite article (5c). What has already been known is corroborated: 'an' and 'boyfriend' are co-predicative. Thus the non-specific person is a boyfriend.

$t^6_{\ 1}$ is characterized as a noun just after the adjective (5d). This noun is the boyfriend again. Hence, 'former' and 'boyfriend' stand in the co-predication relation. The precise interpretation of this relation requires implicit lexical information

---

nature. My homogeneous theory predicts, however, that there is no difference between immediate syntactic relations and non-syntactic external relations in either respect, including transparency. It is not more (or less) difficult to grasp a syntactic external relation than a non-syntactic one. Nor is it more (or less) difficult for the child to learn the former kind of means than the latter kind. Babarczy's (1996) cross-linguistic research on language acquisition corroborates this hypothesis derived from the theory: it is not true at all that syntactic means are easier to learn than morphological means. English children are good at the former whereas Hungarian children at the latter.

in the course of the embedding of the sentence in the hearer's information state
since it cannot be regarded as a simple conjunction: 'x is boyfriend and x is former'.

$t^7{}_1$ and $t^7{}_{11}$ are characterized as a determiner and a noun marked with nomi-
native or dative, respectively (5e). The proper name in dative ('Mari-dat') is suit-
able for both roles (5b). Hence, the definite person named Mari is to be selected as
the possessor of the boyfriend. A precise and realistic characterization of the pos-
sessive construction, as usual, requires implicit encyclopedic information.

After summarizing the results concerning semantic relations (7), the under-
specified DRS in (4a) above can be replaced with a DRS (8) that can serve as an
input of the process in the course of which the semantic content of the sentence is
embedded in the hearer's information state.

(7) $v_3 = v_8 = v_9 = r_m$ (Mari)

         $v_4 = v_5 = v_6 = v_7 = r_s$ (the man that gets married)

(8) $\{\text{MARI}(r_m), \text{MARRY-A-WOMAN}(r_s), \text{AN}(r_s), \text{FORMER}(r_s), \text{BOYFRIEND}_{poss}(r_s, r_m)\}$

3. Information states

This section is devoted to the demonstration of an attempt to construct a realistic
model of the **hearer's information state** (1.4.II, 1.5) and the embedding of the
content of discourses in this structure from sentence to sentence. I am going to
argue that it is an indispensable task of dynamic semantics. One reason is that cer-
tain features of a sentence cannot be interpreted without considering the hearer's
lexical knowledge, cultural or encyclopedic knowledge, or the knowledge supposed
to be common to speaker and hearer due to their long acquaintance (1.3). Another
reason is that it is the cost of returning to Kamp's (1981) original intuition, i.e.
DRSs are small partial models, instead of accepting their uncertain theoretical sta-
tus in present approaches (Kamp–Reyle 1993). It is worth that cost, however, to
construct a level of representation like this because it enables us to capture relations
among referents, propositions and possible worlds in so effective a way that sheds
new light on stubborn problems of dynamic semantics.

3.1. Lexicon, referents, worlds

Let us scrutinize the arguments in support of the elaboration of the level of repre-
sentation mentioned above as the **hearer's information state**.

The starting-point according to the logic of this article is that the introduction of log-
ical connectives into DRSs in the course of the extension of the theory to new linguistic
areas creates a confusion about the theoretical status of DRSs, which is also relevant to the

possibility for constructing a strictly compositional syntactic counterpart to them (details in **3.1.3**). An alternative might be to insist on regarding DRSs as small partial models, whose intricate system of connections is to be given a separate level of representation.[13]

Another observation provides a key to an appropriate determination of this level. As was mentioned above, certain features of a sentence cannot be interpreted without considering particular pieces of the hearer's lexical, cultural, encyclopedic, or interpersonal knowledge. Obviously, they do not come from a discourse consisting of a few sentences but from a source that can be regarded as the hearer's information state. Thus the content of a discourse is to be embedded in this information state (from sentence to sentence), instead of assigning an independent representation to the discourse. In this approach a hearer is supposed to build a single gigantic DRS during his/her life, which is not a simple one, of course, but richly structured. It may enable us to reflect in some way structures of different speakers' different texts and the two-directional transfer of information between discourses that a hearer participates in and his/her knowledge: on the one hand, discourses feed information to the hearer's information state and, on the other, the old knowledge stored there is required to produce coherent interpretation of a discourse. Thus the hearer's information state can serve as a representation **in which** and **by which** discourses can be interpreted.

Another possibility provided by a level like this concerns a uniform representation and adequate arrangement of different sorts of information (lexical, cultural/encyclopedic, interpersonal, etc.), which are to be available at the same time in the course of the interpretation of a discourse (see Appendix).

Furthermore, the definition of the hearer's information state suggests an approach to the encyclopedic and interpersonal data (often required by the calculation of the coherent interpretation of discourses; **1.3, 3.1.3.5–6**) which is not so hopeless than regarding them as a huge data base. It is often held by linguists that phenomena whose explication requires encyclopedic data or other non-lexical sorts of data do not belong to semantics, or do not belong to linguistics at all. A straightforward reason for this approach is that lexical data can still be thought of as such that can be collected and arranged in a finite lexicon whereas non-lexical knowledge amounts to knowledge about an entire culture and perhaps the whole lives of the speaker's and the hearer's. Theoretically, the hearer's information state is to include this knowledge but in the present approach the linguist's task is not to fill it up with data but to construct the structure that makes it possible for the information content of discourses to be fed to it. Both lexical and non-lexical data come from discourses and are stored in the same form as discourses are represented.[13]

---

[13] I have adopted here some principle of Homogeneity of Lexical and Discourse Representations, which Kálmán (1990) attributes to Bartsch (1987, 2). See also Kálmán – Szabó (1990).

Finally I would like to mention a promising property of the structure of the hearer's information state as will be defined here, which can be elucidated by comparing it with a traditional predicate logical system. Referent-like objects (I mean individual constants and variables), relation (predicate) names, and worlds are defined separately. In the course of the definition of the hearer's information state, however, these three structured sets, together with the set of propositions, are determined by **simultaneous recursion**. Thus these sets will permanently be defined depending on each other. This construction is an effective explicit generalization of the traditional predicate logic, in a spirit inspired by (the early version of) DRT (Kamp 1981). Instead of two sorts of individual terms, for instance, as many different sorts of referents will be at our disposal as many worlds have been defined. Informally speaking, a referent that 'belongs to' a world plays the role of a constant in that world while looks like a variable from (certain) other worlds.

The following definition, which still resembles a fragment of that of a traditional predicate logic, is the basis of the definition of the hearer's information state. The remarks about parts of the former definition serve as preparation for the fairly complicated latter one.

D2   Suppose P_k is a finite (or empty) set for every natural number k (k = 0, 1, 2, ...), called the set of **k-ary lexical relation names**. Let P denote their union: $\cup\{P_i : i = 0, 1, 2, ...\}$, the set of **lexical relation names**. Suppose further that **R** is a denumerably infinite set, called the set of **referents**, and **W** is also a denumerably infinite set, that of **worlds**.

I would like to make a few remarks about the definition.

3.1.1. Lexicon

P is a huge lexicon where also such abstract items can be found as the focus operator or lexical items corresponding to, say, locative prepositions or inflections in language, in accordance with an attempt to consider semantic relations as instances of co-predication (Appendix). Elements of P_k refer to k-place relations. If $p \in P_k$, and $r_1, r_2, ..., r_k$ are referents from **R**, then the expression '$<r_1, r_2, ..., r_k> \in p$' is a well-formed proposition: we claim that these referents (in this order) stand in the relation denoted by p. The expression '$<r_1, r_2, ..., r_k> \in p$' (which will often be written as '$p(r_1, r_2, ..., r_k)$') is not a relation, only an **instance** of a relation, or a predication of certain referents' standing in a relation. The relation itself is a set of k-tuples of referents, i.e. a subset of R^k. It is relevant that in traditional model-theoretic semantics a model is supposed to store **total** information on relations (each k-tuple of referents is known to be inside or outside a k-place relation), whereas the hearer's information state in my approach typically contains **partial** information on

relations.[14] The application of traditional models as information states (Groenendijk–Stokhof 1990, Kamp–Reyle 1993) is an anachronistic relic of truth-conditional semantics in dynamic semantics, or at least a temporary situation, which results in such consequences as claiming that an assertion, say *Peter loves Mary*, is only a **test** to the hearer (i.e. (s)he is only able to judge its truth value but cannot embed its information content in his/her information state).

The elements of **P** are assumed here to be real lexical units of a language. As for major syntactic categories, nouns denote one-place relations, adjectives and adverbs denote one- or two-place relations, and verbs denote k-place relations where k ranges from 0 to 5. A proposition formed with a 0-place relation (e.g. *havazik* 'it snows') plays the role of an atomic statement in a partial model.[15]

3.1.2. Referents

The set **R** of referents is supposed to be similar to Landman's (1986) **pegs**: before use they contain no information, they are only **carriers** of information.

As is illustrated by the examples below, the relation among referents and propositions is two-directional: on the one hand, propositions can be constructed from referents (by combining them with lexical relation names) and, on the other, referents can be constructed from (or rather, assigned to) propositions.

(9) (a) The boy loves a pretty girl. He admitted it to her. His friend was surprised by it.

  (b) $r_3 = \{\text{BOY}(r_1), \text{LOVE}(r_1, r_2), \text{PRETTY}(r_2), \text{GIRL}(r_2)\}$

  $r_4 = \{\text{ADMIT}(r_1, r_3, r_2)\}$

  $\{\text{BE\_SURPRISED\_AT}(r_5, r_4), \text{FRIEND}(r_5, r_1)\}$

---

[14] 'Partial information on a relation' is simply a relation in a mathematical sense. Its partial character manifests itself only in its interpretation. '$<r_1, r_2, ..., r_k > \notin q$' does not necessarily mean that the given referents do not stand in relation q. What this formula means is that the hearer does not still know whether these referents stand in the given relation, or not. Hence, '$<r_1, r_2, ..., r_k > \in q$' and '$<r_1, r_2, ..., r_k > \in$ not-q' is a contradiction, as usual, where not-q is the relation that belongs to the negation of the expression that belongs to relation q. However, '$<r_1, r_2, ..., r_k > \notin q$' and '$<r_1, r_2, ..., r_k > \notin$ not-q' is not a contradiciton but only the expression of a fact that the hearer does not still know whether the given referents stand in the given relation, or not. Thus, relations q and not-q are to be defined separately.

[15] This paper is not about language acquisition (or the evolution of languages) so the set of lexical relation names used in the course of the recursive definition of the hearer's information state is assumed to be a fix subset of **P**. Nevertheless, it is not excluded to assume that the hearer's 'active vocabulary' expands via, say, substituting $p(r_1^{~1}, r_2^{~1}, ..., r_1^{~2}, r_2^{~2}, ..., ..., r_1^{~k}, r_2^{~k}, ...)$, where p is a new element of the set of lexical relation names, for the DRS $\{p^1(r_1^{~1}, r_2^{~1}, ...), p^2(r_1^{~2}, r_2^{~2}, ...), ..., p^k(r_1^{~k}, r_2^{~k}, ...)\}$ where each p^i belongs to an earlier stage of the hearer's vocabulary.

The first sentence demonstrates the typical case, which can be represented by even the simplest version of predicate logic, when persons or things are referred to, and something is asserted of them. In the second sentence the pronoun *it* refers to the proposition constructed on the basis of the first sentence (and not a person or a thing). Notice that this special referent does not differ from entity-type referents wrt. the occupation of argument slots of lexical relation names. It is asserted of two simple referents and the one that refers to a DRS that they stand in a certain relation. And a pronoun in the third sentence can refer to this DRS, the one expressed by the second sentence, again... The definition of the hearer's information state can capture this mutual relation between referents and propositions due to the technique of simultaneous recursion.

The famous example below also illustrates a case where certain referents' standing in a certain relation is referred to. The new element is that the DRS referred to is a complex one in the sense that it expresses a relation between two simple DRSs, marked with r^1 and r^2 here. According to the definition of the hearer's information state, the merchant's referent 'belongs to' another world than the farmer's one and the donkey's one (in a sense to be discussed precisely). What is demonstrated here is that even a DRS in whose construction different worlds have been involved can be referred to.

(10) (a) If a farmer owns a donkey, HE sells IT to a merchant.
          $r^0 = <r^1, r^2>$
          where $r^1 = \{\text{FARMER}(r^1{}_1),\ \text{OWN}(r^1{}_1, r^1{}_2),\ \text{DONKEY}(r^1{}_2)\}$, and
          $r2 = \{\text{SELL}(r^1{}_1, r^1{}_2, r^2{}_1),\ \text{MERCHANT}(r^2{}_1)\}$
      (b) ... Mary is surprised at THIS STRANGE CUSTOM.

To sum up, not only the construction of DRSs and worlds are based on referents but also that of referents is often based on DRSs and worlds. The hearer's information state is an effective means to truly reflect the intricate system of relations among referents, propositions, and worlds, due to its being defined by simultaneous recursion.

3.1.3. Worlds

The hearer's information state contains a subset of the set **W** of worlds, furnished with a structure, a **partial order**. They enable the hearer to store separately knowledge about the actual state of affairs, the past, the future, and the culture, canonical scenarios, lexical and encyclopedic units of knowledge, assumptions on the background knowledge of different speakers, as well as beliefs, wishes, and uncertain information that comes from other people. Members of the set **W** of worlds are similar to pegs, as well as referents, in that they are only **carriers** of information.

Before use they are empty, and the hearer feeds information to them. They are essentially represented as DRSs. They are frozen discourses whereas actual discourses are constructed from up-to-dated fragments of stored worlds.[16]

Below I am reviewing the data, mainly famous examples of discourse semantics embedded in a sketchy history of a particular direction line, which have motivated the world structure involved in the definition of the hearer's information state.

3.1.3.1. Truth-conditional model-theoretic semantics

The simple two-sentence discourse in (11) is a popular starting-point of papers on discourse semantics (e.g. Groenendijk–Stokhof 1990).

(11) A man walks in the park. HE whistles.

It demonstrates why truth-conditional semantics based on total models is unsatisfactory. **Meaning** in that model lies in truth conditions. The meaning of the first sentence thus equals the set of models where it is true. And it is true in models where there is **at least** one man walking in the park. What entity does *he* in the second sentence refer to then? Obviously, a male person should be looked for in the model. The theory predicts that unless exactly one male person can be found in the model, the discourse is ungrammatical. Although there **may** be more men in the model, the discourse is undoubtedly correct.

3.1.3.2. Dynamic semantics: small partial models

This simple phenomenon has led some researchers to say that "...the meaning of a sentence does not lie in its truth conditions but rather in the way it changes (the representation of) the information of the interpreter" (Groenendijk–Stokhof 1989). This idea is the cornerstone of dynamic semantics. One way to implement it is to create small partial models for discourses (Kamp 1981; Kamp–Reyle 1993; Heim 1982; 1983). In the partial model that can be associated with the first sentence of discourse (11), there is exactly one man, even if the speaker himself thinks that there are more men in the park. The pronoun *he* straightforwardly refers to the unique man in the small partial model. This approach is based on a special representation of partials models, called **discourse representation structures** (DRS), and hence the entire theory is often called a **representational** semantics.

DRSs also provide a solution to the problem of the famous **donkey sentences**:

(12) (a) If a farmer owns a donkey, HE beats IT.
      (b) If a farmer owns a donkey, HE beats IT. *HE hates IT.
(13) (a) Every farmer who owns a donkey beats IT.
      (b) Every farmer who owns a donkey beats IT. *He hates IT.

There is no evident solution available in traditional truth-conditional theories since a farmer may own more donkeys in the case of (12–13a); and which one has been referred to by the pronoun *it* in the second clause then? The discourse representational solution is really attractive. A small partial model should be constructed with a farmer and a donkey owned by him, and another one in which the farmer of the first model beats the donkey of the first model (Kamp–Reyle 1993):

(13) (c) $< <\{x,y\}, \{farmer(x), donkey(y), own(x,y)\}> \Rightarrow <\{ \}, \{beat(x,y)\}> >$

This complex DRS predicts (correctly) that sentence (12a) is true if the truth of the first subordinate DRS always implies the truth of the second subordinate DRS, and the truth value of the former should be checked for each <farmer, donkey> pairs. The crucial point here is that it is irrelevant how many times a certain farmer takes part in <farmer,donkey> pairs, i.e. how many donkeys particular farmers have.

The intuition is excellent but the use of the conditional symbol ($\Rightarrow$) creates a confusion about the status of DRS. Is it a partial model,[17] i.e. a semantic object, or a formula of a predicate logical language,[18] hence a syntactic object? This theoretical uncertainty entails other problems, that of **compositionality**, for instance. In Section 2 I argued against a categorical refusal (Groenendijk–Stokhof 1990) of the possibility of a syntax relative to which the construction of simple DRSs is compositional.

Therefore the introduction of complex DRSs and the attribution of a double character to them seem to be an unfavorable tendency. We should return to the (compositional) **partial model** interpretation and find another way to capture relations between DRSs. I have argued that this way is through a DRS-like representation of the hearer's information state. Thus what I suggest is that, instead of asso-

---

[17] "DRSs are partial information structures, not only in that they will typically assert the existence of only a small portion of the totality of individuals that are supposed to exist in the worlds of which they intend to speak, but also in that they will specify only some of the properties and relations of those individuals whose existence they do assert." (Kamp – Reyle 1993)

[18] "...each DRS can be regarded as a formula of [first-order] predicate logic in disguise- -a disguise, moreover, that neither is nor is intended to be a particularly effective one." (Kamp – Reyle 1993)

ciating a discourse with a complex DRS consisting of a logically structured set of simple DRSs, we should rather embed simple DRSs in the hearer's information state one by one, where they are assimilated by a partially ordered set of worlds. This latter structure is intended to substitute for logical connectives.

I argue that certain lexical items in the sentence to be processed (e.g. articles, connectives, tense and aspect, etc.) are responsible for the successful implementation of this assimilation. They control the embedding of a discourse (from sentence to sentence) in the hearer's information state. Their contribution to the extension of the amount of information at the hearer's disposal manifests itself in their **discourse organizing capacity**.[19]

3.1.3.3. Dynamic Predicate Logic: no representations

Another way to solve problems with the uncertain status of DRSs is to dispense with them and return to a more conventional predicate logic. Groenendijk and Stokhof (1989; 1990)[20] have supplied certain logical connectives with a limited **dynamic power**, and at this cost they have managed to save formulas in their original predicate logical shape:

(11) $\exists x[man(x) \& walk\_in\_the\_park(x)] \& whistle(x)$
(12) $\exists x[farmer(y) \& \exists y[donkey(y) \& own(x,y)]] \rightarrow beat(\mathbf{x,y})$
(13) $\forall x[[farmer(x) \& \exists y[donkey(y) \& own(x,y)]] \rightarrow beat(\mathbf{x,y})]$

The original problem was that the highlighted variables above are **free** occurrences, and there was no technique to force them to take the same values as the earlier **bound** occurrences of the same variables. This problem was not only a technical one but had something to do with **totality** of models. It is not at all sure that the value of x or y (a man or a donkey) that have been found to verify the given existential formula is really the one that also verifies the formula with the free occurrence of the given variable. The authors mentioned above applied an approach

---

[19] Obviously, the hearer's information state itself also affects the process of the embedding of sentences. It is partly the hearer that will decide (unconsciously) what to do with a discourse; which I find a realistic picture. **Misunderstanding**, for instance, a typical source of humor in comedies, is based on this fact: participants of conversations associate certain expressions with different referents. (Naturally, a speaker's referents always differ from another speaker's referents. Nevertheless, certain referents seem to refer to the same entity in the real world, say, a person.)

[20] I would like to mention again that there are other sorts of attempts to return to more conventional principles of logic (e.g. Heim 1990), too. As what is focused on here is how to get rid of certain theoretical problems with DRT, and what is the consequence of the introduction of the means offered, a detailed comparison between DRT and more conventional alternatives would go beyond the scope of this paper.

known from the semantics of programming languages: they have defined certain
logical connectives, viz. &, ∃ and →, so that after verifying (certain parts of) the
formula created by them the last value they have taken must be retained. Due to this
dynamic power, the bold face variables above **must** be coreferential with their ear-
lier bound occurrences.

It should be emphasized that only the three logical connectives mentioned are
supplied with dynamic power and, moreover, the conditional (→) has only an **inter-
nal** dynamic power, meaning that in a formula φ→ψ the variables of φ are to be iden-
tified with those of ψ, but later these values must not be retained. Conjunction (&),
however, is both internally and **externally** dynamic: in a formula φ→ψ the variables
of φ are to be identified with those of ψ, and then these values must be retained.

The data below in (11)–(14) provide evidence in support of this variation of
predicate logic:

(11) (a) A man walks in the park. HE whistles.

       (b) A man walks in the park. HE whistles. HE is happy.

       (c) It is not the case that a man walks in the park. *HE whistles

(12) (a) If a farmer owns a donkey, HE beats IT.

       (b) If a farmer owns a donkey, HE beats IT. *HE hates IT.

(13) (a) Every farmer who owns a donkey beats IT.

       (b) Every farmer who owns a donkey beats IT. * He hates IT.

(14)   Every farmer owns a donkey. *HE beats IT.

In the case of conjunction (which is externally as well as internally dynamic)
even the pronoun in the third sentence in (11b) has inherited the referent of 'a man'
in the first sentence and that of 'he' in the second one. (Static) negation (11c) can-
not extend the referent of 'a man' in the first sentence to that of 'he' in the second
one. In the case of the conditional formulas, which are only internally dynamic,
variables of the second clause, but not those of the third clause (which is the sec-
ond sentence), are able to accept earlier values (12). Finally, the universal quanti-
fier has no dynamic power (14).

Thus the basic idea is that certain logical connectives (&, ∃ and →) ensure
passing on values of bound variables (a possibility not available in traditional pred-
icate logic).

3.1.3.4. Modal subordination

Two kinds of problems seem to arise. (14b, d) illustrate cases where values are
passed on that have been predicted not to be able to be passed on. Whereas in
(16)–(17) in **3.1.3.5** values that have never been taken seem to have been passed on.

(14) (a) Every farmer owns a donkey. *HE beats IT.

    (b) Every player chooses a pawn. HE puts IT on square one.

    (c) Harvey courts a girl at every convention. *SHE is very pretty.

    (d) Harvey courts a girl at every convention. SHE is usually very pretty.

    (e) Jancsi hal-at akar fog-ni. *Lát-od (A HAL-AT) innen?

        John fish-acc want-3sg catch-inf see-2sg-def the fish-acc from-here

        'John wants to catch a fish. *Do you see IT / THE FISH from here?'

    (f) Mari gazdag ember-hez akar feleség-ül men-ni. Bankár kell legy-en (minimum)!

        Mari rich man-to want-3sg wife-as go-Inf banker must be-imp-3sg minimally

        'Mary wants to marry a rich man. HE must be a banker

    (g) Bár len-ne Mari-nak autó-ja! *Péter is fog-ja vezet-ni (AZ AUTÓ-T).

        I_wish be-cond-3sg Mari-gen car-poss-3sg Péter also be-fut-3sg-def drive-inf the car-acc

        'I wish Mary had a car. *Peter will drive it too.'

    (h) Bár len-ne Mari-nak autó-ja! Péter is vezet-het-né (AZ AUTÓ-T).

        I_wish be-cond-3sg Mari-gen car-poss-3sg Péter also drive-can-cond-3sg-def the car-acc

        'I wish Mary had a car. Peter could drive IT / THE CAR TOO.'

    (i) Az-t hisz-em, Péter-nek van autó-ja. ???Most a ház mögött van (AZ AUTÓ).

        that-acc believe-1sg-def Péter-gen is car-poss-3sg now the house behind is the car

        'I suppose Peter has a car. ???Now it is behind the house.'

    (j) Az-t hisz-em, Péter-nek van autó-ja. Feltehetőleg most a ház mögött van (AZ AUTÓ).

        that-acc believe-1sg-def Péter-gen is car-poss-3sg presumably now the house behind is the car

        'I suppose Peter has a car. Now IT/THE CAR is presumably behind the house. (I guess...)'

In (14b) the second sentence is held to be acceptable due to the fact that a **canonical scenario** is described. In (14d) *usually* saves the second sentence by clarifying that an arbitrary convention is referred to.[21] I agree with Karttunen (1976) that cases where universal quantification is involved (14a–d) are related to those where **modal** (14e–h) or **non-factive** (14i–j) verbs (expressions) are involved.[22] The common property is that non-actual worlds (of wishes, beliefs or possibilities) seem to have come into existence, whose referents can sometimes (second sentences) be

[21] Naturally, (14c) is unacceptable only with a reading where different girls are courted.

[22] Hungarian translations of Karttunen's (1976) examples make things clearer because in Hungarian specific indefinite (e.g. *egy halat* 'a fish.acc') and non-specific (e.g. *halat* 'fish.acc') nominal expressions can be distinguished (by using a determiner, on the one hand, or choosing a bare nominal expression, on the other) whereas in the original examples these two readings interfere with each other. Hungarian thus have a richer system of external means in this area, or rather, a different system, which lacks, however, forms like 'no car' ('Peter has no car'). I regard this phenomenon as a further example of different mappings from semantic relations to sets of external relations (where determiner use co-operates with verb selection: *halat (*meg)fog* vs. *(meg)fog egy halat* 'fish.acc (*preverb)+catch' vs. '(preverb)+catch a fish.acc.'

referred to, and sometimes cannot. Certain sentential adverbials or modal expressions in the second sentences (*usually, must, could, presumably, guess*) help make these sentences acceptable, that is, to enter non-actual worlds (called **modal subordination** by Roberts 1989), due to their discourse organizing capacity.

Groenendijk and Stokhof (1989; 1990) argue that what is required is an **externally dynamic** version of conditional, in addition to the **only internally** dynamic version. Moreover, as shown by (15b–c) below, sometimes the speaker seems to use an externally dynamic version of negation ~ (instead of the static one), and a both internally and externally dynamic version of disjunction ∨ (instead of the static one):

(15) (a) If a client turns up, you treat HIM politely. You offer HIM a cup of coffee and ask HIM to wait.

(b) It is not true that John does not own a car. IT is red and IT is parked in front of the house.

(c) Either there is no bathroom here, or IT is in a funny place. In any case, IT is not on the first floor.

There are no technical obstacles to elaborating a version of predicate logic like this but it will have a too great generative power.[23] Furthermore, this approach suggests a strategy in the course of which, in cases like those of the two-sentential discourses examined, the speaker ought to decide, when he utters the first sentence, which version of the alternative logical connectives to use (the dynamic or the static one). I do not think that a speaker necessarily makes a decision in advance. He may decide on adding a bit later some further information on the non-actual world. He should do nothing else but to insert a special expression in the second sentence, which has a discourse organizing capacity for referring back to the non-actual world similar to that of anaphors. The hearer's task is to understand this special kind of reference to a world and to include the information content of the sentence to the non-actual world, instead of his earlier model. My definition of information states is a formulation of this approach.

3.1.3.5. Encyclopedic knowledge

The problem with (16)–(17) is that values that have never been taken seem to have been passed on.

(16) (a) Joe got married yesterday. The priest spoke very harshly.

(b) Joe got married yesterday. ???The dog barked very loudly.

(17)   Ha 'Mari-nak 'nösül egy 'régi 'udvarló-ja, akkor 'fel-keres-i a 'menyasszony-t...

if Mari-dat marryawoman-3sg an old suitor-poss3sg then pref$_{up}$-seek-3sg-defObj the fiancée-acc

'If a former boyfriend of Mary's is getting married, she visits the fiancée...'

(e.g. in order to talk to her about the man's bad habits)

[23] It can generate lots of unacceptable discourses (an observation due to Kálmán (p.c.)).

No priest or fiancée is mentioned in the first sentence of the corresponding discourse. What makes then the definite descriptions, usually held to be able to refer back to **old information** (known from earlier text or somewhere else) **legitimate**? Our cultural knowledge that a wedding can be associated with a well-defined fiancée and a well-defined priest (and this association is not necessarily a logical inference but only some kind of **accessibility** between lexical items (Kálmán 1990; Kálmán–Szabó 1990)). And this knowledge cannot be supposed to be stored anywhere else but in the hearer's information state. I argue that embedding sentences in information states where certain pieces of encyclopedic knowledge have been activated (as a consequence of a so far unsuccessful search for an antecedent to legitimize a definite description) is to be preferred to pumping lexical information into discourse representations. A comparison of (17) with (16) shows that alternative worlds are also suited for being fed with extra encyclopedic information.

3.1.3.6. Common background knowledge

Finally, the illustration below is intended to prove that the hearer's information state should contain a great number of alternative worlds.

(18)   Speaker A/B/C, to hearer D: "I met Peter yesterday..."

Suppose speaker A said the given sentence to hearer D a day before speaker B did, and speaker C said it to another person but hearer D happened to hear it. Suppose further that A works for the same company, and B lives in the same house, as D. It may occur that though the name *Peter* refers to three different persons, D as a hearer feels no uncertainty. It refers to a man at work in the first case, a neighbor in the second case, and someone unknown to D in the third case.

How many persons named *Peter* then can be found in D's information state? There must be at least two. And how can he select the appropriate one in the course of a discourse? It should be assumed (again!) that the hearer uses three different partial worlds in the three cases. In the first case he selects a world that contains A's and his common background knowledge, in the second case: B's and his common background knowledge, whereas in the third case D presumably creates a new world. In the first two worlds each there is only one person with the name *Peter*[24] so D can easily find the appropriate referent. In the new world in the third case, D

---

[24] It is not excluded that there are more Peters in two persons' common background knowledge, while retaining that no problem arises with reference. The explanation lies in the fact that the common background knowledge itself consists of a great number of intricately related worlds. Out of which, however, there is a **salient** one with only a unique Peter, and the salience of this world also belongs to their common background knowledge.

will not begin to look for a person with the name *Peter* for he knows that the speaker does not suppose him to know the Peter in question. Hence, D introduces a new referent in the new world and waits for new pieces of information concerning this unknown Peter...

Thus among the relevant factors are the discourse organizing capacity of proper names, a rich system of alternative worlds to store information on the common background knowledge of different pairs of people, and the role that the hearer plays in a discourse.

3.2. The hearer's information state

Let us consider a set **P** of lexical relation names, a set **R** of referents and a set **W** of worlds. In what follows the **hearer's information state** is defined.

3.2.1. Preliminary remarks

The basis of the definition is simple: the hearer has referents and partial information on relations between these referents. The role of referents is to indicate that certain pieces of information concern the same entity.

The first complication, discussed in **3.1.2**, is that referents are permanently mixing with statements about referents, yielding that the set R of referents and the set Q of statements about relations between referents below are not disjoint sets and they are to be defined by simultaneous recursion. It also has turned out that many referents refer not to entities of the real world but products of discourses.

Another complication concerns worlds, thoroughly discussed in **3.1.3.** Without a well-organized network of worlds, one could find nothing in his memory and the speaker could refer to nothing effectively. Here it is assumed that the relation between worlds is a **strict partial order** ('precede'),[25] which is claimed to be enough for substituting for logical connectives. As was shown in **3.1.3.4**, it is also relevant to the hearer which point of his world structure is just **active**. Thus a **cursor** is required that points to the active world. This world structure thus is like a Windows editor.

The third complication concerns referents. It is a commonplace that definite descriptions (e.g. *the boy* below), personal pronouns (*he*), and proper names (*Peter*) typically refer to **old referents**, supposed by the speaker to be contained by the hearer's information state, whereas an indefinite description (*a pretty girl*) triggers the introduction of a **new referent**, i.e. the application of an **empty peg** (Landman

---

[25] According to the definition of **strict partial order** (e.g. Partee *et al.* 1990), I. no world precedes itself (**irreflexivity**), II. if a world precedes another world, then the latter does not precede the former (**asymmetry**), III. if a world precedes a second world, which precedes a third one, then the first one also precedes the third one (**transitivity**).

1986). In the definition below this latter process is divided into (i) the introduction of an empty referent into the set R_{new} of 'new referents', and (ii) the assignment of an assertion to this referent, resulting in it leaving the set of new referents for the set R_{old} of old referents.

(19) The boy / He / Peter caught sight of a pretty girl.

Examples like those in (16)–(17), however, are a warning to us that in certain cases the hearer's information state scarcely contains the referent to be searched because of its being referred to by a definite description. A referent like this should be searched in an **extension** of the hearer's information state (**3.3**), which can be constructed by an appropriate application of the hearer's lexical, encyclopedic, and/or interpersonal knowledge. The permanent definition of the accessible referents of formulas below will be relevant later to this latter definition.

There is also a fourth complication: we can refer to sets (or **plural individuals**), or more precisely, we can introduce set referents for sets on which we may have partial information. This question will be ignored.

Now let us consider the entire simultaneously recursive definition of the hearer's information state. Its pieces are going to be discussed afterwards.

D3 An **information state** of the hearer (interpreter) is an n-step information state $I = <R_{old}, R_{new}, Q, W, <, \varepsilon, w>$ for some natural number n.
An **n-step information state** is an $I = <R_{old}, R_{new}, Q, W, <, \varepsilon, w>$ septuple, to be defined by simultaneous recursion, where
R_{old}, a subset of **R**, is called the set of **old referents**;
R_{new}, also a subset of **R**, is the set of **new referents**;
Q is a set of **relations** (i-place relations between the referents in $R = R_{old} \cup R_{new}$, for different natural numbers i);
W, a subset of **W**, is called a family of words;
$<$ is a relation between worlds, a strict partial order ($w_1 < w_2$ is to be read as w_1 **precedes** w_2, and we say that w_1 **immediately precedes** w_2 if w_1 precedes w_2 and there is no world w_3 in W such that w_1 precedes w_3 and w_3 precedes w_2; furthermore, there is a **least element** in W, defined by the property of preceding every world except itself);
ε is a relation from referents (R) to worlds (W) (read: r **belongs to** w) such that if a referent belongs to a world, then it also belongs to each world preceded ('subsequent worlds');
and w is a distinguished element of W, the **cursor position**.
Remember P_m denotes the set of m-place lexical relation names, which is a subset of **P**.
A **one-step information state** $I = <R_{old}, R_{new}, Q, W, <, E, w>$ is of the following form where r_0 is an arbitrary element of **R**, and w_0 is an arbitrary element of **W**:

$R_{old} = \varnothing$

$R_{new} = r_0$

$Q = \varnothing$

$W = \{w_0\}$

$< = \varnothing$

$\varepsilon = \{<r_0, w_0>\}$

$w = w_0$.

If $I' = <R_{old}', R_{new}', Q', W', <', \varepsilon', w'>$ is an (n–1)-step information state (with $R' = R_{old}' \cup R_{new}'$), then **each** of the 7-tuples $I = <R_{old}, R_{new}, Q, W, <, \varepsilon, w>$ defined below is an n-step information state.

D3.1. **cursor movement backwards (to a world w'')**:

$R_{old} = R_{old}'$

$R_{new} = R_{new}'$

$Q = Q'$

$W = W'$

$< = <'$

$\varepsilon = \varepsilon'$

$w = w''$ where w'' is an arbitrary world ($w'' \in W'$) that precedes ($<'$) w'.

2. **cursor movement forwards (to a world w'')**:

$R_{old} = R_{old}'$

$R_{new} = R_{new}'$

$Q = Q'$

$W = W'$

$< = <'$

$\varepsilon = \varepsilon'$

$w = w''$ where w'' is an arbitrary world ($w'' \in W'$) preceded ($<'$) by w'.

3. **introduction of a new referent r''**:

$R_{old} = R_{old}'$

$R_{new} = R_{new}' \cup \{r''\}$ where $r'' \in \mathbf{R} \backslash R'$ (remember $R' = R_{old}' \cup R_{new}'$)

$Q = Q'$

$W = W'$

$< = <'$

$\varepsilon = \varepsilon' \cup \{<r'', u> : w' = u \text{ or } w' <' u\}$

$w = w'$

4. **expansion of a simple relation**:

$R_{old} = R_{old}' \cup \{r_1, r_2, ..., r_m\}$

$R_{new} = R_{new}' \backslash \{r_1, r_2, ..., r_m\}$

$Q = Q' \cup \{P(r_1, r_2, ..., r_m)\}$ where $r_1, r_2, ..., r_m \in R'$ and $P \in \mathbf{P}_m$ $(m \geq 0)$

$W = W'$

$< = <'$

$\varepsilon = \varepsilon'$

$w = w'$

The accessible referents of $P(r_1, r_2, ..., r_m)$ are: $r_1, r_2, ..., r_m$.

5. **expansion of a complex relation**:

$R_{old} = R_{old}'$

$R_{new} = R_{new}'$

$Q = Q' \cup \{\{q_1, q_2, ..., q_m\}\}$ where $q_1, q_2, ..., q_m \in Q'$ $(m \geq 2)$

$W = W'$

$< = <'$

$\varepsilon = \varepsilon'$

$w = w'$

The accessible referents of $\{q_1, q_2, ..., q_m\}$ are: all the accessible referents of $q_1, q_2, ..., q_m$.

6. **introduction of a referent r" for an instance of a relation**:

$R_{old} = R_{old}' \cup \{r''\}$ where $r'' \notin R'$, $r''=q''$, q'' is an arbitrary element of Q'

$R_{new} = R_{new}'$

$Q = Q'$

$W = W'$

$< = <'$

$\varepsilon = \varepsilon' \cup \{<r'', u> : w' = u \text{ or } w' <' u\}$

$w = w'$

The accessible referents of r'' are: the accessible referents of q''.

7. **introduction of a new world w"**:

$R_{old} = R_{old}'$

$R_{new} = R_{new}'$

$Q = Q'$

$W = W' \cup \{w''\}$ where $w'' \in W\backslash W'$, chosen arbitrarily

$< = <' \cup \{<u, w''> : u = w' \text{ or } u <' w'\}$

$\varepsilon = \varepsilon' \cup \{<r, w''> : <r, w'> \in \varepsilon'\}$

$w = w''$

8. **expansion of a conditional relation (and the introduction of a referent for it)**:

$R_{old} = R_{old}' \cup r''$ where $r'' = <q_1, q_2, ..., q_m>$ and $r'' \in R\backslash R'$ $(m \geq 2)$

$R_{new} = R_{new}'$

$Q = Q' \cup \{<q_1, q_2, ..., q_m>\}$ where each q_i is an element of R_{old}'

$(i=1,2,...,m)$ and belongs to a world w_i of W' which precedes $(<')$ w_{i+1}

$(i=1,2...,m-1)$, and $w' <' w_1$

$W = W'$

$< = <'$

$\varepsilon = \varepsilon' \cup \{<r'', u> : w' = u \text{ or } w' <' u\}$

$w = w'$

The accessible referents of r": $\emptyset$ (it has none).

9. **introduction of a set referent r":**

$R_{old} = R_{old}'$

$R_{new} = R_{new}' \cup \{r''\}$

$Q = Q' \cup \{\text{MEMBER}(r, r'')\}$ where MEMBER $\in$ P_2 (a distinguished binary lexical relation name)
and $r \in R_{old}'$ (arbitrary)

$W = W'$

$< = <'$

$\varepsilon = \varepsilon' \cup \{<r'', u> : w' = u \text{ or } w' <' u\}$

$w = w'$

The first remark about the long definition above concerns its **status**. I regard it as a definition of the "state" of the hearer, in whose development both linguistic and extralinguistic influences have played some role. As for linguistic influences, the essence of my approach is that a sentence heard turns the hearer's information state into a state that satisfies the criteria of being an information state of the hearer again. Moreover, there are intermediate states that are also 'information states of the hearer'. As for extralinguistic influences, the way they contribute to the hearer's information state presumably differs from the way linguistic factors exert their influence, but it is not obvious that the (structure of the) result is different.

A recursive definition requires an **initial step**. Here the initial step consists of the introduction of a referent and a world, the former belonging to the latter. The cursor points to the single world.

3.2.2. The cursor

The **cursor** can move backwards as well as forwards in the network of worlds (D3. 1–2). Here are the relevant rows of the parts of the definition concerning the cursor:

D3.1: $w = w''$ where w" is an arbitrary world (w"$\in$ W') that precedes ($<'$) w'.

D3.2: $w = w''$ where w" is an arbitrary world (w"$\in$ W') preceded ($<'$) by w'

This network, due to the rules that develop it (1, 2, 7), is **connected** in the sense that there are no isolated sets of worlds in it.[26] Hence, it is possible to get to an arbitrary

---

[26] A proper subset U of a set W of worlds is isolated iff ('if and only if') for each element u of U, there is no such element w in W that precedes or is preceded by u. The world structure in the hearer's information state can be represented in the same well-known way as the **dominance** relation between constituents of a sentence: by means of a constituent structure tree

world from any world by some combination of the two kinds of cursor movements. The argument against the immediate introduction of a single cursor movement rule that would ensure the possibility of a free jump between arbitrary worlds is that in certain periods of the development of the hearer's information state there seem to be only backward movements (see later). Otherwise, there must be situations when the hearer jumps immediately from a world to an **incommensurable** world: for instance, when another speaker begins to talk to him, the hearer should select (a world in) another common background (**3.1.3.6**).

3.2.3. New referents

The introduction of a new referent (r" in D.3.3) is associated with a constraint that the new referent will belong to every world that the world the cursor points to precedes, in addition to the world itself that the cursor points to:

D3.3: $\varepsilon = \varepsilon' \cup \{<r", u> : w' = u \text{ or } w' <' u\}$

This constraint is part of a general constraint (see D3.6–9), i.e. every referent of a world is accessible from any subsequent (or **later**) world but not vice versa (Kamp and Reyle (1993) formulate a similar constraint). In (14g–h), for instance (repeated below), a new "world of wishes" is supposed to have been introduced that is preceded by the original, say now, actual world.

(14) (g) Bár len-ne Mari-nak autó-ja! *Péter is fog-ja vezet-ni (AZ AUTÓ-T).
         I_wish be-cond-3sg Mari-gen car-poss-3sg Péter also be-fut-3sg-def drive-inf the car-acc
         'I wish Mary had a car. *Peter will drive IT too.'
     (h) Bár len-ne Mari-nak autó-ja! Péter is vezet-het-né (AZ AUTÓ-T).
         I_wish be-cond-3sg Mari-gen car-poss-3sg Péter also drive-can-cond-3sg-def the car-acc
         'I wish Mary had a car. Peter could drive IT / THE CAR too.'

The non-existing car only belongs to this new world whereas Peter is obviously to be interpreted as a participant of the real world. The second sentence of (14g) is intended to concern the real world (presumably because of the discourse organizing capacity of the simple future tense); that is why it is wrong: it is only the new world that the car referent belongs to. Whereas the second sentence of (14h), concerning the new world (of wishes), due to the conditional mood, is correct: hence, Peter's referent also belongs to the new world. Continuations of the common first sentence in (14g–h), where the real world is concerned but there is no reference to referents that only belong to the world of wishes, would also be correct (e.g. 'Unfortunately, neither Peter nor Mary earns much').

3.2.4. Simple and complex relations

D3.4 is the only rule that feeds information to the system, the other rules organize and/or reorganize it. The new piece of information is a proposition concerning a relation: it is asserted of a definite number of referents that they stand in the given relation.

D3.4: $Q = Q' \cup \{P(r_1, r_2, ..., r_m)\}$ where $r_1, r_2, ..., r_m \in R'$ and $P \in P_m$ $(m \geq 0)$

As worlds are partial models, which means that we have only partial information on relations, a new statement about a relation is an "expansion" of (our information on) the given relation. If a relation name has never been used before, its expansion means that our information on this relation expands from the empty set $(\varnothing)$ to a singleton. Once a new referent turns out to stand in a relation, it leaves the set of new referents for the set of old referents, since it is already suitable for being referred to by means of just this relation

Notice that it is not assumed that the referents mentioned in D3.4 necessarily belong to the active world (the world the cursor points to) so there may be referents that only belong to subsequent worlds. This approach enables us to make a distinction between arguments and non-argument-like expressions (e.g. bare nominals; this latter ones will not be associated with referents in the active world, see (23) **3.3**).

D3.5 provides the means to produce situations, which can also be associated with referents by D3.6. Then situations can already be referred to.[27]

D3.5: $Q = Q' \cup \{\{q_1, q_2, ..., q_m\}\}$ where $q_1, q_2, ..., q_m \in Q'$ $(m \geq 2)$

D3.6: $R_{old} = R_{old}' \cup \{r''\}$ where $r'' \notin R'$, $r'' = q''$, q'' is an arbitrary element of Q'

(9)   (a) The boy loves a pretty girl. He admitted it to her. His friend was surprised by it.
       (b) $r_3 = \{BOY(r_1), LOVE(r_1, r_2), PRETTY(r_2), GIRL(r_2)\};$
       (c) $< r_1, r_2 > \in q$
       (d) $r_4 = \{ADMIT(r_1, r_3, r_2)\}$
       (e) $\{HE(v_1), ADMITTED(v_1, v_2, v_3), IT(v_2), SHE(v_3)\}$
       (f) $\{BE\_SURPRISED\_AT(r_5, r_4), FRIEND(r_5, r_1)\}$

---

[27] Identifying this 'association' of situations and referents with an identity relation between them (D3.6) is undoubtedly a simplification which is somewhat harmful unless situations are furnished with sufficient (spatio-) temporal information.

(9c) above shows the complex relation, which can also be regarded as a (non-lexical) relation (q above) between r1 and r2, about which we have partial knowledge. Those referent pairs stand in this relation whose first member is a boy, second member is a girl and pretty, and the former loves the latter. Thus D3.5 is an effective way to extend the lexicon without including non-lexical elements in **P**.

D3.6 enables us to refer to a situation by simply including a statement about a relation between certain referents in the set of old referents. Thus entity-type referents mix with statements about them. Perhaps it is this mixture of levels that makes language so effective, and so intricate, a device.

The example above illustrates that a statement's referent can be part of a statement again (the discourse referent r_3 in (9d) and the other discourse referent r_4 in (9f)). A grammatical analysis in the style demonstrated in Section **2** assigns the second sentence of (9a) the underspecified DRS in (9e), which expresses that a male person admitted something to a female person. Due to the special discourse organizing capacity of definite pronouns, the temporary referents v_1, v_2, and v_3 can be identified with the referent assigned to the boy mentioned in the first sentence, the referent of the whole situation described by the first sentence, and the referent of the pretty girl mentioned also in the first sentence, respectively. We have constructed the DRS in (9f). The DRS in (9d), which belongs to the third sentence in (9a), can be calculated likewise.

3.2.5. The tree of worlds

The introduction of a new world (w" in D3.7) could be compared to the growing of a twig from a thicker twig, or a bough, or the trunk of the tree of worlds. This new twig is assumed to be preceded by just the last active world (w') and the worlds preceding it. The new world inherits all referents of the world that it has grown from (which implies that it inherits all referents of the preceding worlds; but later it can be assigned further referents that are not necessarily assigned to the preceding worlds).

D3.7: $W = W' \cup \{w"\}$ where $w" \in W\backslash W'$, chosen arbitrarily

$< = <' \cup \{<u, w"> : u = w'$ or $u <' w'\}$

$\varepsilon = \varepsilon' \cup \{<r, w"> : <r, w'> \in \varepsilon'\}$

I conjecture that the family of worlds actually has a trunk, i.e. a least element that precedes every other world. This trunk should correspond to what the hearer regards as the real world. Presumably, the trunk contains the hearer's lexical knowledge whereas boughs contain different parts of his cultural or encyclopedic knowledge, and thicker twigs the common backgrounds associated with his acquaintances.

3.2.6. Conditionals

Whereas in D3.5 a **set** of statements about relations has been introduced, in D3.8 a (totally ordered) **sequence** of them is constructed. It receives a referent that belongs to the active world, and the referent of the first member of the sequence belongs to a subsequent (later) world, then the referent of the second member belongs to an even later world etc.

D3.8: $Q = Q' \cup \{<q_1, q_2, ..., q_m>\}$ where each q_i is an element of R_{old},

(i = 1,2,...,m) and belongs to a world w_i of W' which precedes ($<'$) w_{i+1}

(i = 1,2...,m-1), and $w' <' w_1$

This special construction is intended to capture the essence of **conditionals**, which have played a central role in the development of dynamic semantics (see (12)). Let us analyze the following examples:

(20) (a) If a farmer owns a donkey, HE sells IT to a merchant.

$r^0 = <r^1, r^2>$ where $r^1 = \{FARMER(r^1_1), OWN(r^1_1, r^1_2), DONKEY(r^1_2)\}$, and

$r^2 = \{SELL(r^1_1, r^1_2, r^2_1), MERCHANT(r^2_1)\}$

where $w^0 < w^1 < w^2$, $r^i_{(j)} \varepsilon w^i$

(b) ... *HE will get little money for IT. $(w^0; HE(r^1_1), IT(r^1_2))$

(c) ... Mary is surprised at THIS STRANGE CUSTOM. $(w^0; CUSTOM(r^0))$

(d) ... Or HE keeps on beating IT. $(w^1; HE(r^1_1), IT(r^1_2))$

(e) ... Although HE usually gets little money for IT. $(w^2; HE(r^1_1), IT(r^1_2))$

Suppose w^0 is the world that the discourse referent of the entire conditional sentence in (20a) belongs to in the hearer's information state after the embedding of the sentence. World w^0 can be called the 'actual world'. Two other worlds are to be introduced in order to place the DRSs belonging to the two clauses of the conditional sentence: w^1 is preceded by w^0 and precedes w^2; further, a referent with a superindex i always belongs to world w^i, for i = 0, 1, or 2. Thus, only the referent of the entire conditional belongs to world w^0 (besides Mary's referent in (20c)). The DRS of the first clause (r^1) and the two participants mentioned in this clause, the farmer (r^1_1) and the donkey (r^1_2), have referents that belong to world w^1. The referents that belong to (only) world w^2 are that of the DRS of the second clause and that of the merchant. The word 'only' is intended to remind the reader of the assumption that later worlds inherit all referents from preceding worlds. The referent of the donkey, for instance, belongs to world w^2 as well.

The continuation in (20b) is incorrect since it is to be interpreted in the actual world (w^0), presumably because of the simple future tense, but the potential refer-

ents of *he* and *it*, the farmer and the donkey, cannot be found in that world, since they belong to later worlds only. Continuation (20c) is correct, however, provided that Mary is known. The explication is that this sentence is to be interpreted in the actual world, too, but the expression *this strange custom* refers to the discourse referent of the entire conditional, which also belongs to the actual world (by contrast with its clauses). The case of continuation (20e) is fairly simple: this sentence is to be interpreted in w^2, due to *although*, which clearly connects the selling situation to the content of the second sentence, and *usually*. *He* and *it* refer to the farmer and the donkey, whose referents have been said to belong to world w^1, so they also belong to the later world w^2.

As for the continuation in (20d), the connective *or* triggers the introduction of a new world w^3, which is to be an alternative to w^1 in the sense that w^3 is immediately preceded by w^1 (as well as w^2) but w^3 and w^2 are incommensurable worlds in the tree structure. As w^3 inherits the referents of the farmer and the donkey from the preceding w^1, sentence (20d) is easy to interpret.

As for the discourses with canonical scenarios, universal expressions, modal or non-factive elements in (14) and (15a), I argue that they should be analyzed essentially the same way as conditionals. These factors, as well as the conditional structure, triggers the introduction of non-actual worlds, due to their special capacity for discourse organizing. The semantic content of the sentences in question is embedded in these non-actual worlds and can be retrieved by means of similar elements with an appropriate discourse organizing capacity.

An entire solution to the illustrated problems require a formal analysis of the discourse organizing factors, which would go beyond the scope of this paper. The claim here is that the treatment of the different phenomena collected above is to be based on the same mechanism (described in the previous paragraph and formalized in connection with the conditional sentence). Let us review the basic types:

(14) **canonical scenario:**
    (b) Every player chooses a pawn. HE puts IT on square one.
    universal quantifier:
    (c) Harvey courts a girl at every convention. *SHE is very pretty.
    (d) Harvey courts a girl at every convention. SHE is usually very pretty.
    modal expression:
    (g) I wish Mary had a car. *Peter will drive IT too.
    (h) I wish Mary had a car. Peter could drive IT / THE CAR too.
    non-factive expression:
    (i) I suppose Peter has a car. ???Now it is behind the house.
    (j) I suppose Peter has a car. Now IT/THE CAR is presumably behind the house.

The first sentence in (14b) triggers the introduction of a non-actual world, which represents a canonical scenario, presumably due to its special style and content. The continuation is to be interpreted in the world of the canonical scenario (because there is no change in tense, mood, and aspect, and, perhaps, intonation), so *he* can be regarded as referring to the player who has chosen a pawn. In (14d) it is obviously due to *usually* that the second sentence is to be interpreted in the non-actual world where Harvey's behavior at an arbitrary convention (with a single girl courted by him) is described. In (14h) it is to be attributed to the conditional form *could* that the car in the world of wishes can be referred to in the second sentence. Finally, in (14j), it is presumably due to *presumably* that the car mentioned in the first sentence, whose existence is only a supposition, can be referred to in the second sentence.

3.3. Sentence embedding in (extended) information state

I would like to repeat here that my purpose is to save the **representational** nature characteristic of the beginning of dynamic semantics (Kamp–Heim Theory), often criticized nowadays because of facts resulting from the uncertain status of DRSs and the absence of compositionality in the strictest sense. I am arguing that there is a natural syntax according to which the simplest class of DRSs can be constructed in a **compositional** way whereas logical-formula-like DRSs can be dispensed with. What needs to be done is to represent the **hearer's information state** as one huge intricate DRS instead of assigning DRSs to discourses. Further, I would like to point out that this approach sheds new light on stubborn problems due to the fact that referents, propositions and worlds are defined by simultaneous recursion.

3.3.1. Preliminary remarks

In **3.2** the basic structure of the hearer's information state was defined. Now I would like to define how a sentence said to the hearer can be embedded in a structure like this. Examples in (16)–(17) show that the first step should be the preparation of the information state for the new sentence. Or rather, the new sentence triggers some reorganization of the information state. Or even more precisely, the hearer believes in the **coherence** of the discourse performed (Kálmán 1990; Kálmán–Szabó 1990) so he is prepared for reorganizing the structure of his information state to a certain extent unless he could embed the sentence in its original version. In other words, he **extends** his information state in certain areas.

The definition below is intended to determine the structural limits of this extension or reorganization at the moment of hearing the sentence. Thus no further information is supposed to get in the system, either linguistic or extralinguistic;

only reorganizing steps are allowed besides the very piece of information that a sentence should be processed.

Consequently, an unlimited use of the steps defined in D3.2 is excluded. Nevertheless, six of them can be used: those that can feed no information to the system. The fifth rule (in cooperation with the seventh rule) creates an **absorbing** world: a point for the new sentence to cling to before embedding. The last three rules practically produce referents, so an **extension** of the hearer's information state may offer more referents to lexical items of the sentence to be embedded than the initial information state.

Notice that the **extension** of the hearer's information state is an information state again, which could be constructed by means of the original definition of the hearer's information state (D3). A consequence of this fact is that the **available referents** of propositions in Q (in D) need not be defined again but are declared now to be the same as if the given propositions were produced by D3.

D4 An **extension** of the hearer's information state $I^* = <R_{old}{}^*, R_{new}{}^*, Q^*, W^*, <^*, \varepsilon^*, w^*>$ is an $I = <R_{old}, R_{new}, Q, W, <, \varepsilon, w>$ m-step extension of the hearer's information state for some natural number m such that $W\backslash W^*$ contains exactly one element, called **absorbing world**.

An **m-step extension** $I = <R_{old}, R_{new}, Q, W, <, \varepsilon, w>$ of the hearer's information state $I^* = <R_{old}{}^*, R_{new}{}^*, Q^*, W^*, <^*, \varepsilon^*, w^*>$ is an information state to be defined by simultaneous recursion, where for m = 0 I = I*. If $I' = <R_{old}', R_{new}', Q', W', <', \varepsilon', w'>$ is an (m -1)-step extension of the information state I* (with $R'= R_{old}' \cup R_{new}'$), then **each** of the septuples $I = <R_{old}, R_{new}, Q, W, <, \varepsilon, w>$ defined below is an m-step extension of I*.

1. **cursor movement backwards (to a world w")** (identical to D3.1):

$R_{old} = R_{old}'$
$R_{new} = R_{new}'$
$Q = Q'$
$W = W'$
$< = <'$
$\varepsilon = \varepsilon'$
$w = w''$ where w" is an arbitrary world that precedes w'.

2. **introduction of a new referent r"** (identical to D3.3):

$R_{old} = R_{old}'$
$R_{new} = R_{new}' \cup \{r''\}$ where $r'' \in R\backslash R'$
$Q = Q'$
$W = W'$
$< = <'$
$\varepsilon = \varepsilon' \cup \{<r'', u> : w'=u \text{ or } w' <' u\}$
$w = w'$

3. **expansion of a complex relation** (identical to D3.5):

$R_{old} = R_{old}$'

$R_{new} = R_{new}$'

$Q = Q' \cup \{\{q_1, q_2, ..., q_m\}\}$ where $q_1, q_2, ..., q_m \in Q'$

$W = W$'

$< = <$'

$\varepsilon = \varepsilon$'

$w = w$'

4. **introduction of a referent r" for an instance of a relation** (identical to D3.6):

$R_{old} = R_{old}' \cup \{r''\}$ where $r'' = q''$, q'' is an arbitrary element of Q'

$R_{new} = R_{new}$'

$Q = Q$'

$W = W$'

$< = <$'

$\varepsilon = \varepsilon' \cup \{<r'', u> : w' = u \text{ or } w' <' u\}$

$w = w$'

5. **introduction of a new world w"** (almost identical to D3.7; but no change in cursor position):

$R_{old} = R_{old}$'

$R_{new} = R_{new}$'

$Q = Q$'

$W = W' \cup \{w''\}$ where $w'' \in W\backslash W'$, chosen arbitrarily

$< = <' \cup \{<u, w''> : u = w' \text{ or } u <' w''\}$

$\varepsilon = \varepsilon' \cup \{<r, w''> : <r, w'> \in \varepsilon'\}$

$w = w$'

6. **introduction of a set referent r"** (identical to D3.9):

$R_{old} = R_{old}$'

$R_{new} = R_{new}' \cup \{r''\}$

$Q = Q' \cup \{\text{MEMBER}(r, r'')\}$ where MEMBER $\in P_2$ (a distinguished binary lexical relation name) and $r \in R_{old}'$ (arbitrary)

$W = W$'

$< = <$'

$\varepsilon = \varepsilon' \cup \{<r'', u> : w' = u \text{ or } w' <' u\}$

$w = w$'

7. **introduction of a new referent r" to an absorbing world**:

$R_{old} = R_{old}$'

$R_{new} = R_{new}' \cup \{r''\}$ where $r'' \in R\backslash R'$

$Q = Q$'

$W = W$'

$< = <$'

$\varepsilon = \varepsilon' \cup \{<r", u> : u \in W' \backslash W^*\}$

$w = w'$

8. creation of a new relation q" by generalization:

$R_{old} = R_{old}' \cup \{r"\}$ where $r" \in R\backslash R'$

$R_{new} = R_{new}'$

$Q = Q' \cup \{q"\}$ where $q" = q[r, r''']$, r is an available referent of proposition q, r''' is an element of R_{new}' that belongs to a world in $W'\backslash W^*$, $q[r, r''']$ denotes a formula obtained by the substitution of r''' for each occurrence of r in the formula of q, and $r" = q[r, r''']$

$W = W'$

$< = <'$

$\varepsilon = \varepsilon' \cup \{<r", u> : w' = u \text{ or } w' <' u\}$

$w = w'$

9. creation of a new relation q" by specification:

$R_{old} = R_{old}' \cup \{r", r'''\}$ where $r" \in R\backslash R'$, and $r''' \in R_{new}'$ and belongs to w'

$R_{new} = R_{new}' \backslash \{r'''\}$

$Q = Q' \cup \{q"\}$ where $q" = q[r, r''']$, r is an available referent of proposition q and belongs to a world that w' precedes; and $r" = q[r, r''']$,

$W = W'$

$< = <'$

$\varepsilon = \varepsilon' \cup \{<r", u> : w' = u \text{ or } w' <' u\}$

$w = w'$

10. creation of a new relation q" by the application of a conditional relation:

$R_{old} = R_{old}' \cup \{r"\}$ where $r" \in R\backslash R'$

$.R_{new} = R_{new}'$

$Q = Q' \cup \{q"\}$ where the following are satisfied:

there is an old referent (in R_{old}') which belongs (ε') to w' and identical to a conditional relation $<q_1, ..., q_k> \in Q'$;

q" derives from $<q_1, ..., q_k>$ by the following substitutions:

$q" = <\{q_1[r_1',r_1"]...[r_j',r_j"], q_2[r_1',r_1"]...[r_j',r_j"]\},$

$q_3[r_1',r_1"]...[r_j',r_j"], ..., q_k[r_1',r_1"]...[r_j',r_j"]>$

where $r_1', ..., r_j'$ $(j \geq 0)$ are referents contained by the formula of q_1 that do not belong (ε') to w', while $r_1", ..., r_j"$ are referents belonging (ε') to w' that $q_1[r_1',r_1"]...[r_j',r_j"] \in Q'$

$W = W'$

$< = <'$

$\varepsilon = \varepsilon' \cup \{<r", u> : w' = u \text{ or } w' <' u\}$

$w = w'$

Further, if k = 2, then $q_k[r_1',r_1"]...[r_j',r_j"] \in Q$, and every statement it contains but is not contained by a conditional relation is also a member of Q.

3.3.2. The absorbing world

The first remark concerns the first paragraph of the definition above. It is stipulated that the family of worlds is allowed to be increased by exactly one world in spite of the unrestricted later possibilities. The sentence to be embedded will begin to grow in the direction of this "absorbing world".

3.3.3. Cursor movement backwards

The cursor is allowed to move only backwards (D4.1). A fragment of (20) illustrates the positive side of this rule (details there):

(20) (a) If a farmer owns a donkey, HE sells IT to a merchant.

$r^0 = <r^1, r^2>$ where $r^1 = \{$FARMER$(r^1{}_1)$, OWN$(r^1{}_1, r^1{}_2)$, DONKEY$(r^1{}_2)\}$, and

$r^2 = \{$SELL$(r^1{}_1, r^1{}_2, r^2{}_1)$, MERCHANT$(r^2{}_1)\}$

where $w^0 < w^1 < w^2$, $r^i{}_{(j)} \varepsilon w^i$

(c) ... Mary is surprised at THIS STRANGE CUSTOM. $(w^0;$ CUSTOM$(r^0))$

(d) ... Or HE keeps on beating IT. $(w^1;$ HE$(r^1{}_1)$, IT$(r^1{}_2))$

(e) ... Although HE usually gets little money for IT. $(w^2;$ HE$(r^1{}_1)$, IT$(r^1{}_2))$

Remember w^0 is the world that the discourse referent of the entire conditional sentence in (i) belongs to in the hearer's information state. The discourse referent of the premise belongs to world w^1, and that of the conclusion belongs to w^2; and $w^0 < w^1 < w^2$. Suppose further that the cursor in the final state of the embedding of the conditional sentence points to the latest world, w^2. The background of this stipulation is that it is the latest twig in the tree of worlds that serves as a natural point for the following sentence to cling to. (20e) illustrates this case. Nevertheless, the cursor is permitted to return to a preceding world: to that of the premise in (20d) where the continuation provides an alternative to the selling situation with the premise as a common base (*or!*), or to the actual world (w^0) in (20c) where something is asserted of the conditional as a whole.

The introduction of forward cursor movement, which, together with backward cursor movement, would result in the possibility of free jump, requires thorough examination. I conjecture that free jumps are triggered by extralinguistic facts such as speaker change.

3.3.4. New referents, new statements

New referents are allowed to be introduced freely (D4.2). Certain expressions of the new sentence (e.g. 'a pretty girl') will look for them. It is a technical detail that they are assumed to be introduced in advance, instead of assuming that their introduction is triggered by indefinite expressions like the one mentioned. If an asser-

tion is associated with a 'new referent' of an information state, then it will belong to the old referents of the resulting new information state.

D4.3 ensures the possibility of forming complex relations. Notice that complex relations contain no more information than their components, in contrast with simple relation expansion where a new statement appears. The following simple illustration proves that the definition must not lack D4.3 and D4.4, the collection of certain statements and the possibility of referring to them as a whole. The highlighted part may refer to a long story.

(21)   I believe WHAT THAT MAN SAYS.

In connection with D4.3 and D4.4, it is to be mentioned that an infinite number of referents can be produced. Totally chaotic collections, however, are not likely to be referred to. Their exclusion might (partly) be based on the rich world structure: some kind of **complexity** should be defined.

3.3.5. The absorbing world

The rule concerning the introduction of a new world (D4.5) is a slightly modified version of D3.7: the cursor does not enter the world created. The use of this rule is not restricted **here** but an earlier constraint ensures that only a single new world is created relative to the hearer's original information state. I call this world the **absorbing** world because the new sentence will grow in its direction (see later). Thus the task of this rule is restricted to the creation of the absorbing world. D4.7 supplies it with new referents, also without cursor movement.

3.3.6. Generalization

Let us turn to D4.8. It has been mentioned what endless possibilities lie in referring to arbitrary collections of statements about relations. Nevertheless, language has an even greater referential power, as shown below:

(22)   Tegnap   győzött   az A csapatunk   a   spanyolok ellen.
         yesterday won-3sg   the A team-poss-1pl   the   Spanish-pl against
         'Yesterday our team A won a victory over the Spanish team.'
         $r_4 = \{\text{YESTERDAY}(r_3), r_3 = \{\text{DEFEAT}(r_1, r_2), \text{TEAMA} (r_1), \text{TEAMSP}(r_2)\}\}$
   (a) ... Ez csodálatos!
         this   marvelous
         'That is marvelous.'
         $r_4 = \{\text{YESTERDAY}(r_3), r_3 = \{\text{DEFEAT}(r_1, r_2), \text{TEAMA} (r_1), \text{TEAMSP}(r_2)\}\}$

(b) ... Ez   nem   tegnapelőtt   történt?

this   not   the_day_before_yesterday   happened-3sg

'Didn't that happen the day before yesterday?'

$r_3 = \{\text{DEFEAT}(r_1, r_2), \text{TEAMA}(r_1), \text{TEAMSP}(r_s)\}$

(c) ... Bárcsak   sikerülne   EZ (A BRAVÚR)   ma   a   B csapatunknak   is!

I_wish   succeed-cond-3sg   his the success   today   the   B team-poss-1pl   also

'I wish our team B could replicate this victory today.'

$r' = \{\text{DEFEAT}(r_1, r_2), \text{TEAMSP}(r_2)\}[r_1, r^1{}_1] = \{\text{DEFEAT}(r^1{}_1, r_2), \text{TEAMSP}(r_2)\}$

(d) ... Bárcsak sikerülne EZ (A BRAVÚR) ma a B csapatunknak is az angolok ellen!

I_wish succeed-cond-3sg this the success today the B team-poss-1pl also the English-pl against

'I wish our team B could replicate this victory over the English today.'

$r'' = \{\text{DEFEAT}(r_1, r_2)\}[r_1, r^1{}_1][r_2, r^1{}_2] = \{\text{DEFEAT}(r^1{}_1, r^1{}_2)\}$

In the second sentence of each two-sentence discourse there is a referential expression, which refers to different situations in the four different cases. The pronoun in continuation (22a) refers to the content of the entire first sentence (the fourth row of (22a) shows its referent r_4) whereas the pronoun in (22b) only refers to the proposition of the first sentence without the time adverb. Nevertheless, there is no problem because these referents have been at our disposal provided that the first sentence had been embedded.

It is the second two examples (22c–d) that require a special treatment. The highlighted definite expression in continuation (22c) refers to the set of statements denoted by r', whose most interesting feature is not that it consists of only two statements but that a new referent (which does not belong to the active world) has been substituted for the one in the first argument position of DEFEAT. The relevant detail of D.8 is the following:

D4.8: $Q = Q' \cup \{q''\}$ where $q'' = q[r, r''']$, r is an available referent[28] of proposition q, r''' is an element of R_{new}' that belongs to a world in $W' \backslash W^*$, $q[r, r''']$ denotes a formula obtained by the substitution of r''' for each occurrence of r in the formula of q, and $r'' = q[r, r''']$

---

[28] **Availability** of different referents was permanently defined in D3. The relevant point is that a referent that can be found only in a conditional relation is not available so cannot be replaced with the new referent mentioned in D4.8. The reason is that the extension of an information state is to provide no new information. What r' and r'' express in (22c–d) is not new at all. r'', for instance, refers to the situation that a team defeated an (other) team, which is a consequence of the fact that team A defeated the Spanish team. Let us consider, however, that the hearer's information state contains, say, the following conditional relation: "If team A defeats the Spanish team, Mary will be happy." Then it would be undesirable to permit the introduction of the following proposition in the extension of the hearer's information state: "If a team defeats an (other) team, Mary will be happy." That is why referents in conditional relations are not assumed to be available.

What referent r' belongs to is not the situation that team A defeats the Spanish team but the (underspecified?) situation that somebody defeats the Spanish team. The analysis of sentence (22c) contains the identification of this potential winner with team B.

Likewise, in continuation (22d) the pronoun (or definite expression) refers to a one-member set of statement(s), r" above, which can be created in an appropriate extension of the hearer's information state by substituting new referents not belonging to the active world of the hearer's original information state for those in the two argument places of the given single statement (DEFEAT). r" expresses correctly that *EZ A BRAVÚR* 'this success/victory' refers to a situation where somebody (or a team) defeats somebody (or a team). The former entity is to be identified with team B while the latter one with the English team on the basis of sentence (iv).[29]

Notice that $\text{DEFEAT}(r^1{}_1, r^1{}_2)$ essentially corresponds to a λ-expression in a Montagovian logic: $\lambda x.\lambda y.\text{DEFEAT}(x, y)$. $r' = \{\text{DEFEAT}(r^1{}_1, r_2), \text{TEAMSP}(r_2)\}$ also corresponds to a λ-expression: $\lambda x.(\text{DEFEAT}(x, r_2) \& \text{TEAMSP}(r_2))$. Hence, the application of a structured family of worlds provides a generalization of λ-abstraction: referents that do not belong to the active world serve as variables. Thus a **dynamic** treatment of discourses is substituted for λ-abstraction, which is a means threatening with an endless proliferation of types. An advantage of the approach demonstrated in this article that DRSs are constructed from only expressions of the simplest types.

3.3.7. Specification

D4.9 is the opposite of D4.8 in some sense. In D4.8 a new situation is derived from an old situation by deleting some of its specific details. D4.9 ensures the possibility of substituting a new referent belonging to the active world (but still empty) for a referent outside the active world (but not enclosed in a conditional relation):

D4.9: $Q = Q^{\cdot}$ ({q"} where q" = q[r, r"'], r is an available referent of proposition q and belongs to a world that w' precedes; r"' $\in$ R$_{new}$' and belongs to w'; and r" = q[r, r"'])

Thus we make it possible to refer to something whose existence has already been known but seemed to be irrelevant. Let us look at a phenomenon of linguistics where this possibility plays a crucial role.

The following data are intended to illustrate four types of nominal expressions classified along the dimension of predicativity / referentiality:

---

[29] A less simplified treatment should also contain a reference to the predicate 'be a team' but the simplification does not affect the relevant points.

(23) (a) `LEVEL-ET   ⌐kap-tál!   ... Hárm-at is!   Nagyon hosszú-*(ak)
        letter-acc   received-2sg-indef three-acc also   very long-*(pl)
        `You've received a letter. What's more, three letters. They are very long.`
        $r_3 = \{$RECEIVE$(r_1, r`),$ YOU$(r_1),$ LETTER $(r`)\},$ r' does not belong to w

   (a`)`LEVEL-ET   ⌐kap-tál!!   Nagyon hosszú-(*ak)!
        letter-acc   received-2sg-indef   very long-(*pl)
        `You've received at least one letter. They are very long.`
        $r_6 = \{$RECEIVE$(r_1, r`),$ YOU$(r_1),$ LETTER $(r`)\}[r`, r_5], r_5,$ but not r', belongs to w

   (b) `Kap-tál                  EGY `LEVEL-ET!   {*Hárm-at is!} / { Nagyon hosszú!}
        received-2sg-indef   a letter-acc   {*three-acc also} / {very long}
        `You have received a letter. (What's more, three letters.)`
        $r_3 = \{$RECEIVE$(r_1, r_2),$ YOU$(r_1),$ LETTER $(r_2)\}, r_2$ belongs to w, new referent

   (c) `Meg  -  ⌐kap-tál            EGY `LEVEL-ET!
        preverb received-2sg-indef a letter-acc
        `You've received one letter (out of those that you expected).`
        $r_3 - \{$RECEIVE$(r_1, r_2),$ YOU$(r_1),$ LETTER $(r_2)\}, r_2$ belongs to w, new referent but $\cup r_4$

   (d) `Meg  -  ⌐kap-tad            A `LEVEL-ET!
        preverb received-2sg-def the letter-acc
        `You've received the letter.`
        $r_3 = \{$RECEIVE$(r_1, r_2),$ YOU$(r_1),$ LETTER $(r_2)\}, r_2$ belongs to w, old referent

In (23a) above the noun 'letter' is essentially used predicatively, demonstrated by the fact that the first sentence can be followed by a second one where it is claimed that three letters have been received. Thus the bare N has no number feature. The predicative nature of 'letter' can be expressed by giving its argument a referent that does not belong to the active world (23a). Nevertheless, the first sentence of (23a) is allowed to be followed by a second one that contains reference to a single letter. And it is here that rule D4.9 should be applied (just before the embedding of the second sentence). Thus the hearer's task between the first and the second sentence is to produce for the letter a referent that belongs to the active world.[30]

[30] The following sentence illustrates a case where the hearer's task is also to produce for the letter a referent which belongs to the active world: 'Yesterday I was informed that I had been admitted to the university. My mother opened THE LETTER.' After processing the first sentence, the hearer's information state does not contain the relevant letter. What is required here is a means to be discussed later, in **3.3.8**: an extension of the hearer's information state by applying a conditional relation which carries the information that the fact that somebody is informed about something is licensed to be associated with the fact that a letter contains this information. Here is the letter! What is relevant now is that the referent of this letter still does not belong to the active world, likewise the referent that belongs to the bare noun in (23a). This footnote serves the purpose of calling attention to this analogy. The solution is also the same: a new referent of the active world is to be assigned to the potential letter in an appropriate extension of the hearer's information state.

What the first sentence in (23a) illustrates can be called a **non-argument-like** use of a noun. D4.9 ensures the possibility of pretending that the given noun had been used like a (non-specific) argument. The second sentence of (23a'), for instance, forces to produce a state that has been produced immediately by the first sentence of (23b). Here the **non-specific indefinite** use of nouns is exemplified. The sentence contains reference to a new referent of the active world. (23c) illustrates the **specific** use of indefinites. The sentence contains a reference to a new referent of the active world but this referent should stand in a MEMBER_OF relation (Enç 1990) with a set referent of the active world. Finally, the **definite** use of a noun (23d) means that the sentence contains a reference to an old referent of the active world.

3.3.8. How to open conditional relations?

Finally, D4.10 makes it possible to access the internal part of a **conditional relation**. Remember a conditional relation is a sequence $<q_1, ..., q_k>$ of statements about (simple, complex or even conditional) relations. The interpretation of this structure is that terrible dragons (namely, $q_1, ..., q_{k\ 1}$) guard a valuable piece of information (q_k). First of all, dragon q_1 should be defeated, then dragon q_2, and so on, in accordance with the usual fate of dragons. To defeat a dragon amounts to satisfying the corresponding relation q: D4.10 describes its way:

D4.10 $Q = Q' \cup \{q''\}$ where the following are satisfied:
        there is an old referent (in R_{old}') which belongs (ε') to w' and identical to a conditional relation $<q_1, ..., q_k> \in Q'$;
        q'' derives from $<q_1, ..., q_k>$ by the following substitutions:
        $q'' = <\{q_1[r_1',r_1'']...[r_j',r_j''], q_2[r_1',r_1'']...[r_j',r_j'']\},$
        $q_3[r_1',r_1'']...[r_j',r_j''], ..., q_k[r_1',r_1'']...[r_j',r_j'']>$
        where $r_1', ..., r_j'$ ($j \geq 0$) are referents contained by the formula of q_1 that do not belong (ε')
        to w', while $r_1'', ..., r_j''$ are referents belonging (ε') to w' that $q_1[r_1',r_1'']...[r_j',r_j''] \in Q'$

The simplest case is to find q itself freely in the active world ($q \in Q'$) of the hearer's information state. We have also got another weapon: if a referent contained by the formula of q belongs to a world that the active world precedes, it is allowed to try to replace it (its all occurrences) with an old referent that belongs to the active world. This way a new statement q[...] can be produced. If this latter statement is already a member of Q', then a dragon has been defeated. Once q_1 has been satisfied ($q_1[...] \in Q'$), a shorter sequence $<\{q_1[...], q_2\}, q_3 ..., q_k>$ has been derived from $<q_1, ..., q_k>$. Observe that the first two members of the original sequence have

merged yielding a set. Repeating rule D4.10 k–1 times results in a set, i.e. a statement about a complex (but non-conditional) relation. The last sentence of the definition declares that the content of q_k in a specified form is already available:

D4.10:   Further, if k = 2, then $q_k[r_1',r_1'']...[r_j',r_j''] \in Q$, and every statement it contains but is not contained by a conditional relation is also a member of Q.

This procedure plays a central role in verification of conditionals, embedding of sentences and explanation of phenomena like the one below (see also (16)):

(24)   (a) Joe got married yesterday. THE PRIEST spoke very harshly.
   (b) $Q = \{.... \text{GET\_MARRIED}(r_1), \text{JOE}(r_1),$
        $< \text{GET\_MARRIED}(r^1), \{\text{MAKE\_MAN\_AND\_WIFE}(r^2, r^1, r^3), \text{PRIEST}(r^2)\} >, ...\}$
   (c) $\text{GET\_MARRIED}(r^1)[r^1, r_1] \in Q'$
   (d) $\{\text{MAKE\_MAN\_AND\_WIFE}(r^2, r_1, r^3), \text{PRIEST}(r^2)\} \in Q''$
   (e) D6.9: $\{\text{MAKE\_MAN\_AND\_WIFE}(r^2, r_1, r^3), \text{PRIEST}(r^2)\}[r^2, r_2] \in Q'''$
   (f) $\text{PRIEST}(r_2) \in Q'''$ where r_2 belongs to the active world

Suppose the hearer has processed the first sentence in (24a), and is about to process the second one. His set of statements about relations contains the three elements mentioned in (24b). The third one is a conditional referent that must belong to a "bough" of his world structure tree (and hence also to later worlds), being a piece of encyclopedic or cultural knowledge (and not a logical inference, Kálmán 1990). Each referent above with a superindex is regarded as one that does not belong to the active world. Clearly, the referents in the piece of encyclopedic knowledge in (24b) do not belong to the active world.

The information state described in the previous paragraph can be extended by the application of the conditional relation that carries the piece of knowledge concerning weddings. A referent should be found with the properties that it belongs to the active world and 'gets married'. Joe's referent satisfies these requirements so this real fiancé can be substituted for the theoretical fiancé (24c). In this extension of the hearer's information state the second part of the encyclopedic knowledge about a wedding is already available (24d), or rather, its modified version with Joe in the theoretical fiancé's role. The priest is still not available because its referent does not belong to the active world. Rule D4.9 solves the problem, by specifying the priest's referent, i.e. identifying this referent with one that belongs to the active world. Due to the clause of D4.10, the statement PRIEST(r_2) belongs to the set Q''' of statements of a possible extension of the hearer's information state, and hence the hearer can

access this piece of information in the course of the embedding of the second sentence of (24a). He can find the antecedent of the definite expression *the priest*.[31]

3.3.9. Truth-conditional verification

It should be remarked here that **truth-conditional verification** can (and should) be included in the theory (Kamp–Reyle 1993). We can simulate a **total model** on a restricted referent set and vocabulary by considering a hearer who knows everything about this total model. In this approach, the verification of a **conditional**, for instance, means that we try to derive, from a statement <q, q'>, a statement {q[...], q'[...]} by D4.10 in each possible way. If (and only if) each resulting q'[...] is an element of the current set Q (i.e. **true**, since a total model is assumed now), the conditional is **true**.[32]

---

[31] A detailed analysis of the discourse in (17) is available in the Appendix. Another problem is discussed there whose solution is also to be based on the application of conditional relations that contain lexical or encyclopedic information. This problem concerns the interpretation of the connections among elementary statements in a DRS. Is this connection always a simple **conjunction**? Possessive and attributive constructions, among others, suggest a negative answer. *Mary's former suitor*, for instance, cannot be interpreted as a person who is a suitor, who is former, **and** who is Mary's at the same time. Or rather, I argue that a DRS that includes these three pieces of information is not wrong but only **underspecified**. Possessors, attributive adjectives, and several other elements in a sentence are to be assumed to bear a discourse organizing capacity, due to which the hearer's information state is increased by pieces of lexical and encyclopedic knowledge. It can be derived, at least theoretically, that *Mary's former suitor* is a person with the property that *he courted Mary in former times*. See the Appendix.

I argue that the phenomena illustrated in (15b–c) should also be accounted for by accessing referents that (originally) do not belong to the actual world.

(15) (b) It is not true that John does not own a car. IT is red and IT is parked in front of the house.

   (c) Either there is no bathroom here, or IT is in a funny place. In any case, IT is not on the first floor.

In the case of the sentence *John does not own a car* (15b), the referent of the car does not belong to the actual world (similarly to modal contexts), but exists (!), which ensures the possibility for its identification with a new referent of the actual world in an adequate **extension** of the hearer's information state (as a result of an appropriate discourse organizing rule to neutralize double negation). As for the referent of the bathroom in the first clause of the first sentence in (15c), it does not belong to the actual world either, but it exists, too. That is the crucial point again. A disjunctive element is to be assumed to bear the capacity for the creation of "the opposite" of a world in the extension of the hearer's information state, and in this world the bathroom can already be referred to. On this analysis, the second sentence is also to be embedded in this world.

[32] It can be checked that the verification of a conditional like the one in (17), keeping on assuming, naturally, that each definite description must find its antecedent, will provide the correct interpretation. First we should try to satisfy the premise, by specification. In successful cases, we should find the appropriate fiancée, and then comes the moment when the specified conclusion should be checked. The relevant point is that it is not an arbitrary fiancée that has been checked.

3.3.10. The minimal embedding of a new sentence

The next definition is intended to capture the moment when the sentence that the hearer has just heard clings to the absorbing world of an extension of the hearer's information state. The sentence to be embedded takes the shape of a statement about a **conditional relation**, expressing an approach that the new information carried by the sentence is guarded by one or more layers of conditions (Kálmán 1990, Kálmán–Szabó 1990). These conditions are explicit in a conditional sentence but I claim that every sentence consists of a predicative part and such **conditional layers**. Each piece of information that helps in search for referents is to be regarded as (part of) a conditional layer: finding a referent is nothing else but satisfying a condition.

The definition below is an attempt to capture the structural properties of the representation of a sentence that is already clinging to the absorbing world but its pieces of information still manifest themselves as statements about referents belonging to worlds that are preceded by the active world of the hearer's original information state.

Approaching from another perspective, the task is to prepare the statements about relations associated with lexical items for a later embedding. The final purpose is to feed the new information carried by the new sentence to the hearer's information state.

D5   A **(relation) numeration** is defined as a sequence $N = \langle p_1, p_2, ..., p_k \rangle$ $(k \geq 1)$ where $p_i \in$ **P**.[33]
A **minimal embedding** of a relation numeration $N = \langle p_1, p_2, ..., p_k \rangle$ in an extension I' (with an absorbing world denoted by w^+) of a hearer's information state I* is an information state I that can be derived from I' by an application of the procedure described in D that follows the requirements below:

In period 1, first the cursor should be moved to the absorbing world w^- (D.2), and as a last step in this period a referent that belongs to w^- should be assigned to a statement about a relation, denoted by, say, q^1. The whole period is characterized by the constraints that

   a) the cursor must not go through worlds preceding w^-,

   b) referents in R' must not be referred to in any step (hence D3.3, D3.6 and/or D3.8 are to be applied to produce (new) referents)[34]

   c) only lexical relations contained by the numeration can be used in D3.4, and such a step always results in the deletion of (an occurrence of) the given lexical relation from the numeration

---

[33] The order of elements in a **numeration** is assumed to be irrelevant. Nevertheless, a numeration is to be defined as a **sequence** because it is not excluded that a lexical relation name occurs multiply. Multiple membership is not interpreted in the case of a simple **set**.

[34] Remember R' = $R_{old}' \cup R_{new}'$.

d) the formula of q^1 must contain, for every lexical relation deleted from the numeration in the period, a statement about a relation associated with it.

In period t, $T \geq t \geq 2$ for some t and T, provided that tasks in period t 1 has already been executed, first a new world should be created by D3.7 (let w^t denote it), and as a last step in this period a referent that belongs to w^t should be assigned to an instance of a relation, denoted by, say, q^t. The whole period is characterized by the constraints that

a) the cursor must not go through worlds preceding w^+,

b) referents in R` must not be referred to in any step

c) only lexical relations contained by (the remainder of) the numeration can be used in D3.4, and such a step always results in the deletion of (an occurrence of) the given lexical relation from the numeration

d) the formula of q^t must contain every lexical relation (in the form of predicates) deleted from the numeration in the period.

If there is already no relation remained in the numeration, then there comes a final period. The sequence $\cdot q^1, q^2, ..., q^T>$ should be created by means of D3.8 (and D3.1), and be given a referent r" that belongs to w`. If N is a numeration that belongs to a sentence, then let us call r" the **Davidsonian referent** of this sentence. Finally, the cursor is to move to w^t.

The starting-point of the definition above is a sequence of lexical items (called here **numeration**, after Chomsky 1995). The semantic side of a lexical item is a statement that certain entities stand in a certain relation. These statements should be arranged in **layers** (see (26) a bit later). Technically, the numeration should be emptied into the absorbing world and later worlds (c–d above). The absorbing world and the brand new later worlds serve the purpose of still **isolating** the new sentence (a–b). Otherwise, the general rules of information state development are valid (D3). By the last period a series of worlds will have been created, parallel with the creation of situation referents (whose formulas consist of certain members of the numeration) so that each situation referent belongs to a world. The numeration is already empty. The resulting conditional relation is called the **Davidsonian referent** of the sentence (Davidson 1967; Kamp–Rossdeutscher 1994; Parsons 1995). It belongs to w' whereas the referents that the lexical relation names are associated with as predicates belong to later worlds. This construction is intended to capture the moment when the hearer has already accepted that a situation has been referred to by a sentence but the participants are still undecided.

3.3.11. The proper embedding of a new sentence

The extent of embedding a sentence into the hearer's information state depends on the discourse organizing capacity of its words. A definite description, for instance, usually triggers the identification of a temporary referent of the new sentence with

an old referent of an earlier information state of the hearer. Certain conditional sentences, however, remain minimally embedded in the sense that they express an assertion of referents that cannot be found in the actual world.

D6 Suppose I^{min} is a minimal embedding of a relation numeration $N = \langle p_1, p_2, ..., p_k \rangle$ (with a Davidsonian referent q_0) in an extension I' of a hearer's information state I^*. An m-step extension (see D4) of I^{min} (for an arbitrary m > 0) is called a **proper embedding** of N in the extension I' of I^* if q_0 undergoes D4.10 at least once.

A **minimal embedding of a sentence** in an information state I^* of a hearer is a minimal embedding of its numeration in an arbitrary extension I' of I^*. A **proper embedding of a sentence** in an information state I^* of a hearer is a proper embedding of its numeration in an arbitrary extension I' of I^*.

Rule D4.10 plays the central role in the definition above. The essence of my approach is that every sentence clings to the hearer's information state first as a sequence $q_0 = \langle q^1, q^2, ..., q^T \rangle$ of statements about relations. Then the **discourse organizing** elements of the sentence (operators, articles, connectives etc.) determine whether the sentence remains in the state of a minimal embedding, or not. In the latter case statements $q^1, q^2, ...$ should be satisfied, respectively. Let us look at a series of examples:

(25) (a) If a farmer owns a donkey, he beats it.

    $\langle$ FARMER(r'), {DONKEY(r''), OWN(r', r'')}, BEAT(r', r'') $\rangle$

  (b) If the farmer owns a donkey, he beats it.

    {FARMER(r_1), $\langle$ {DONKEY(r''), OWN(r_1, r'')}, BEAT(r_1, r'') $\rangle$ }

  (c) The farmer owns a donkey and beats it.

    {FARMER(r_1), DONKEY(r_2), OWN(r_1, r_2), BEAT(r_1, r_2)}

    where r_1 and r_2 belong to the actual world, and r' and r'' do not belong to it

(25a) illustrates the case of **minimal embedding**. Neither the farmer's referent nor that of the donkey belongs to the actual world. The hearer is not likely to begin to calculate the truth conditions of the sentence; rather, he stores this piece of information in his information state and will apply to a farmer, and then a donkey of this farmer's, if he needs it.

(25b) illustrates **proper embedding**. It is assumed here that the speaker and the hearer have been talking about a particular farmer; so this farmer is assigned an old referent in the hearer's information state just before processing the sentence in (25b). The farmer's temporary referent (in the **isolated** state of sentence processing) could be identified with this old referent.

The difference between (25a) and (25b) lies in the role of the piece of information FARMER(r'). In (25a) it serves as a condition concerning the future: "if you might hear of a farmer..." Whereas in (25b) the same piece of information serves as an instruction (primarily due to the definite article): "look for the farmer in your information state; the speaker assumes that there is a salient farmer referent there..."

(25c) is another example of proper embedding. If the hearer could accept the sentence, then the farmer's (original) referent has been identified with an old referent, and that of the donkey with a new referent (that also belongs to the actual world). As the worlds preceded by the actual world of the hearer's before processing the sentence, which have been created in the course of its embedding, contain no new information any more relative to the actual world after processing the given sentence, they are to be deleted. A case like this can be called a **total embedding** of a sentence into the hearer's information state.

3.3.12. The minimally embedded Hungarian sentence as a conditional relation

This paper serves the purpose of creating the **framework** of a representational dynamic semantics based on the hearer's information state. Further research in this line would require a detailed examination in two main areas. The one area is the construction of the conditional relation to be assigned to a minimally embedded sentence. Here I am going to make a brief remark about the hypothesized connection between this conditional relation and the basic Hungarian operator structure.[35] The other area amounts to the exploration of the discourse organizing capacity of words and grammatical constructions. I conjecture that not only articles and connectives bear such a capacity but perhaps each lexical item.

Now let us consider the standard assumption on the Hungarian operator structure in the early nineties (Kiefer–É. Kiss 1994).[36] The surface form of a sentence may begin with topics. Then quantifiers may come. A focus position may be the last element of the preverbal zone. The postverbal zone usually consists of arguments. (26a) below represents this structure, and (26b) demonstrates the proposed formula of the Davidsonian argument in the minimal embedding:

(26) (a) $T_1 T_2 ... T_i Q_1 Q_2 ... Q_j F V A_1 A_2 ... A_k$

(b) $< T_1, <T_2, ... <T_i, <Q_1, <Q_2, ... <Q_j, <\{A_1, A_2, ... A_k, V\}, <F> >>>>>>$

---

[35] On the assumption (Szabolcsi 1995) that what is overtly expressed in Hungarian is essentially the same as what is expressed covertly in, say, English, the scope of the discussion here promises a possibility for generalization.

[36] The current alternatives (e.g. Brody 1990; É. Kiss 1995; Szabolcsi 1995) contain no element that seems to be irreconcilable with the spirit of the discussion here.

Let us look at a series of examples as an illustration where proper embedding is prohibited by a universal context like this: "The following statement is true for every word of the text written in cipher..." On this reading the sentences in question will remain minimally embedded for a hearer since they express universal statements about occurrences of letters Q and U in a particular text written in cipher.

(27) (a) Egy   'Q-t$_T$ 'követ$_V$ egy 'U$_A$.
         a   Q-acc   follow-3sg   a   U
         'As for a letter Q, it is followed by a letter U.'
         <Q(r'), {U(r''), FOLLOW(r'', r')}>   (UAQU, *QUAQ)
         scheme: <T, {A, V}>

      (b) Egy   'U$_T$ 'követ$_V$ egy 'Q-t$_A$.
         a   U   follow-3sg   a   Q-acc
         'As for a letter U, it follows a letter Q.'
         <U(r''), {Q(r'), FOLLOW(r'', r')}>   (*UAQU, QUAQ)
         scheme: <T, {A, V}>

      (c) Egy   'U$_F$ követ$_V$ egy `Q-t$_A$.
         a   U   follow-3sg   a   Q-acc
         'It is a letter U that follows a letter Q.'
         < {Q(r'), FOLLOW(r'', r')}, U(r'')>   (UAQU, ?QUAQ)
         scheme: <{A, V}, F>

In (27a) above the predicted formula (correctly) says that first an arbitrary Q should be considered, and this Q must be followed by a U. The statement is false if the text contains the word *QUAQ* because the second Q is not followed by a U. (27b) illustrates the opposite case when first an arbitrary U should be considered, and then this U should satisfy the condition of following a Q. The first U of *UAQU* is not compatible with the statement. Finally, the sequence of statements associated with sentence (27c) predicts that first we should consider an arbitrary situation where a Q is followed by something, and then it is claimed that the letter following Q must be a U. According to this analysis, *QUAQ* is compatible with statement (27c) since the first Q is followed by a U, while the second Q does not meet the premise at all. The judgment here is a bit uncertain, due to the fact, I guess, that there is a tendency to regard a conditional statement as being relevant. In this latter case, the interpretation is that first a Q should be find, then it is supposed that this Q is followed by something (since this possibility has been mentioned in the sentence), and finally it should be turn out that the letter following this Q is a U. We can get rid of this disturbing factor by substituting *követhet* 'follow-can' for *követ*, yielding a statement that is in perfect harmony with what has been predicted.

Summary

My main purpose has been to prove that a dynamic semantics based on DRS-like representations (e.g. Kamp 1981; Kamp–Reyle 1993; Heim 1982; 1983) need not necessarily be thought to be non-compositional (Groenendijk–Stokhof 1989). What is required is an appropriate syntax to a representational dynamic semantics, or rather, an **external component** instead of a syntax in the traditional sense. The attractive discourse representation structures themselves inspired me first to raise the possibility of such a system (Alberti 1990). The version demonstrated in this paper is called a Generative Argument Structure Grammar, because of the distinguished role of lexical characterizations. It can be regarded as a special kind of categorial grammar.

In Section 2 it was argued that GASG is a straightforward implementation of such central ideas of current generative linguistics as the condition of lexical inclusiveness, the elimination of phrase structure grammar, and the idea of a morphology-driven grammar (Chomsky 1995). The cornerstone of my reasoning is that these ideas are not necessarily to be associated with the practice of an entire syntactic encoding of morphological and intonational information. Thus, on the one hand, GASG serves as a compositionally adequate formal counterpart of a representational dynamic semantics whereas, on the other hand, it provides a flexible syntactic means, due to the simultaneous accessibility of pieces of external information held usually (but groundless) as being of different nature.

Section 3 was devoted to the demonstration of an attempt to construct a realistic model of the **hearer's information state** and the embedding of the content of discourses in this structure from sentence to sentence. I have argued that it is an indispensable task of dynamic semantics. One reason is that certain features of a sentence cannot be interpreted without considering the hearer's lexical, cultural/encyclopedic, or interpersonal knowledge. Another reason is that it is the cost of returning to Kamp's (1981) original intuition, i.e. DRSs are small partial models, instead of accepting their uncertain theoretical status in present approaches (Kamp–Reyle 1993).

It is worth that cost, however, to construct a level of representation like this because it enables us to capture relations among the basic components of logical systems, defined separately in predicate logic, in so effective a way that sheds new light on stubborn problems of dynamic semantics (e.g. donkey sentences, modal subordination, the introduction of different kinds of implicit information), and provides a uniform treatment of conditionals, canonical scenarios, universal, modal, non-factive and negative contexts (in harmony with the spirit of Karttunen 1976). The hearer's information state, as has been defined here, promises to serve the pur-

pose of reflecting the intricate interactions among referents, propositions, and alternative worlds due to its being defined by simultaneous recursion.

In the last subsection (**3.3**) it was defined how a sentence said to the hearer can be embedded in his/her information state. A new sentence often triggers some reorganization of this information state. Or more precisely, the hearer is assumed to believe in the **coherence** of the discourse performed (Kálmán 1990; Kálmán–Szabó 1990) so he is prepared for reorganizing the structure of his information state to a certain extent unless he could embed the sentence in its original version. In other words, he extends his information state in certain areas. The **extension** of the hearer's information state is defined as an information state to which no new information has been fed (relative to the hearer's original information state before processing the new sentence to be embedded) but which may offer more referents to pronouns, definite descriptions, and other elements of this new sentence, due to three special rules whose task is to produce referents.

Appendix

1. Notes on lexical characterizations in GASG

1. The propositions (or conditions) on words and potential words (to be found) are formulated in a classical **predicate logical language** whose model's domain consists of the set M of morphemes of the sentence (being examined) together with every finite sequence of M (usually denoted by M*). The vocabulary of predicates consists of such that refer to different kinds of external relations, including one-place relations ('is a noun', 'is in Nom.', etc.). The formulas are as usual (if ϕ and ψ are formulas, then $\sim\phi$, $\phi\&\psi$, $\phi\vee\psi$, $\phi\rightarrow\psi$, $\phi\leftrightarrow\psi$, $\exists v.\phi$ and $\forall v.\phi$ are also formulas), except this one: 'if ϕ is a formula, then $rec[\phi]$ and $dom[\phi]$ are also formulas,' whose interpretation is that ϕ is **recessive/dominant**. Informally speaking, a recessive requirement can easily be overridden while a dominant one always must be satisfied.

2. Propositions written in capitals (INT.STRESS) concern inherent lexical properties of words while those written in small capitals (ɪɴᴛ.sᴛʀᴇss) are only due to affixation, intonation, or other extraneous factors. These characterizations are assumed to be arranged in an inheritance network; a stressed inflected word thus inherits properties from its stem, from its kind of stress, and from its affixes.

3. The lexical characterization of a word (in the inheritance network just mentioned) is to be interpreted as its 'category' in the approach according to which GASG is a kind of categorial grammar.

4. There is an asymmetry in lexical characterizations: those of certain words (e.g. finite verbs) contain references to words to be found in an actual sentence ('potential words') besides their own word(s), whereas those of other words (e.g. nouns) contain no such references. It is assumed thus that external relations are asymmetrical: the one word is referred to in the lexical characterization of the other but not vice versa. This asymmetry may be useful, say, in formulation of rules like this: 'out of two recessive external features in conflict, the one described in the superior lexical item will manifest itself.' It is this asymmetry, together with the assumption that the superiority relation between words is transitive (hence, not circular), that enables us to construct phrase structure trees. This observation

does not entail, however, that phrase structure trees are necessarily to be regarded as the appropriate level to represent each piece of linguistic information (cf. **syntactic encoding** in **2.2**). They may be nothing else but attractive representations of superiority relations.

5. Apostrophes before words mark their being stressed. In Hungarian the first syllable of open class words is stressed as a default; the absence of a stress like this may refer to the presence of a focus operator.

6. Finally a brief note on the particular word *nősül*. It is a real intransitive verb which has no transitive version either. It means 'get married' but can be asserted only of men. Or it might be said, too, that it means 'marry a woman' but the woman cannot be expressed in this construction.

II. Analysis of a conditional sentence

(1)   Ha 'Mari-nak 'nősül egy 'régi 'udvarló-ja, akkor 'fel-keres-i a 'menyasszony-t...
        if Mari-dat marryawoman-3sg an old suitor-poss3sg then pref$_{up}$-seek-3sg-defObj the fiancée-acc
        'If a former boyfriend of Mary's is getting married, she visits the fiancée...' (e.g. in order to talk to her about the man's bad habits)

First I am going to make comments on the lexical characterizations of the items that the sentence above consists of. This part of the Appendix serves as a complement to **2.6**. Alberti (1996, 1998) provide even more detailed analyses.

(2)   (a) IF-THEN(v_1, v_2)
          if ... then: $<t^1, t^2; t^1_1, t^1_2, t^2_1, t^2_2>$
row 1   CAT.CONN(t1), SEQ(t^1_1), SYN.PREC(t^1, t^1_1), SYN.NEAR(t^1, t^1_1), MEM(t^1_2, t^1_1), FINITE(t^1_2).
r2   CAT.CONN(t^2), SEQ(t^2_1), SYN.PREC(t^2, t^2_1), SYN.NEAR(t^2, t^2_1), MEM(t^2_2, t^2_1), FINITE(t^2_2)
r3   ~INT.STESS(t^1), ~INT.STRESS(t^1), ~INT.STESS(t^2), ~INT.STRESS(t^2)

Ha 'if' and *akkor* 'then' are assumed to form a unit in spite of the distance between the two own words (t^1 and t^2), which are connectives (row1.1, r2.1). Both immediately precede (r1.3 4, r2.3–4) sequences of words (t^1_1 in r1.2, and t^2_1 in r2.2) that contain (r1.5, r2.5: MEMber) a finite element (t^1_2 in r1.6, and t^2_2 in r2.6). Thus, both connectives precede a finite clause. The connectives must not be stressed in a grammatical sentence (r3.1, r3.3), and they are not stressed in the sentence under examination indeed (r3.2, r3.4). As for semantics, the temporary referents v_1 and v_2 are to be identified with referents that belong to entire DRSs.

(2)   (b) MARI(v_3)
          'Mari-dat: $<t^3>$
          CAT.N.PROPN(t^3), LEG.REF.SPEC.DEF(t^3), MOR.CASE.DAT(t^3),
          INT.STRESS(t^3), MOR.PERS.3(t^3), MOR.NUM.SG(t^3)

The word above is a stressed proper name in the dative case. Details are available in 2.6, with the exception of the last two formulas (r2.2 3) about the person and number of this noun.

(2)   (c) MARRY-A-WOMAN(v_4)

       'marryawoman-3sg: $<t^4; t^4{}_1, t^4{}_2, t^4{}_3, t^4{}_{31}>$

r1   CAT.V.INTR(t4), FINITE(t^4), INT.STRESS(t^4), rec[INT.STRESS(t^4)],

r2 $[\sim(t^4{}_1 = t^4) \Rightarrow$ INT.STRESS($t^4{}_1$) $\Rightarrow$ SYN.PREC($t^4{}_1, t^4{}_2$)]

r3   LEG.PRED($t^4{}_2$), INT.STRESS($t^4{}_2$), rec[$t^4{}_2 = t^4$]

r4 $[\sim(t^4{}_2 = t^4) \Rightarrow$ (SYN.PREC($t^4{}_2, t^4$) & SYN.NEAR($t^4{}_2, t^4$))]

r5   LEG($t^4{}_3$), CAT.N($t^4{}_{31}$), MOR.CASE.NOM($t^4{}_{31}$), rec[INT.STRESS($t^4{}_{31}$)],

r6 $\sim$SYN.PREC($t^4{}_{31}, t^4{}_3$), SYN.NEAR($t^4{}_3, t^4{}_{31}$),

r7   MOR.PER.3($t^4{}_{31}$), MOR.NUM.SG($t^4{}_{31}$)

2.6 provides only a fragmentary analysis of this stressed finite intransitive verb (r1.1–3), which requires stress as a default (r1.4). In addition to the subject ($t^4{}_3, t^4{}_{31}$), the lexical characterization contains references to two potential elements of the sentence. The one ($t^4{}_1$ in r2) is the topic of the sentence, which usually bears a normal stress (r2), and the other ($t^4{}_2$) is the element called the **verb carrier** in Kálmán   Nádasdy (1994). In focused sentences, for instance, it is the focused element that serves as a verb carrier. The given verb bears the property that its own word plays the role of the verb carrier in a neutral sentence (r3.3). The verb carrier is always predicative in a certain sense (r3.1), and stressed (r3.2). The topic precedes the verb carrier (r2); if there is no normal topic in a sentence, the verb itself is regarded as the topic. If the verb carrier differs from the verb, i.e. the sentence is focused, then the former immediately precedes the latter (r4). Thus the logical/rhetorical structure of Hungarian sentences is described in the lexicon as the potential environment of finite predicators.

    Rows 5 7 characterize the potential subject, which is to contain a determiner-like element ($t^4{}_3$), and a nominal element ($t^4{}_{31}$). As is mentioned in **2.6**, a proper name can play the roles of these two elements at the same time. A detail ignored in **2.6** is expressed by the formula in r5.4: the nominal element is stressed as a default (e.g. in a neutral sentence).

(2)   (d) AN(v_5)

       a(n) : $<t^5; t^5{}_1>$

       CAT.DET.ART.IND(t^5), LEG.REF.NON-SPEC(t^5),

       $\sim$INT.STRESS(t^5), $\sim$INT.STRESS(t^5),

       CAT.N($t^5{}_1$), SYN.PREC($t^5, t^5{}_1$), SYN.NEAR($t^5, t^5{}_1$), MOR.NUM.SG($t^5{}_1$)

The characterization of the indefinite article is complete in **2.6**.

(2)   (e) FORMER(v_6)

       'former: $<t^6; t^6{}_1>$

r1   CAT.A(t^6),

r2   CAT.N($t^6{}_1$), SYN.PREC($t^6, t^6{}_1$), dom[SYN.NEAR($t^6, t^6{}_1$)],

r3   rec[INT.STRESS(t^6) $\leftrightarrow$ INT.STRESS($t^6{}_1$)]

As for the adjective, r3 above provides an additional piece of information: attributive adjectives are stressed or unstressed parallel with the noun that belong to them (as a default).

(2)   (f) BOYFRIEND$_{poss}$(v_7, v_8)

       'boyfriend-poss3sg: $<t^7; t^7{}_1, t^7{}_{11}>$

       CAT.N(t^7), MOR.CASE.NOM(t^7),

$LEG(t^7{}_1)$, $CAT.N(t^7{}_{11})$, $\sim SYN.PREC(t^7{}_{11}, t^7{}_1)$, $NEAR(t^7{}_{11}, t^7{}_1)$,
$MOR.PERS.3(t^7)$, $MOR.NUM.SG(t^7)$, $MOR.PERS.3(t^7{}_{11})$,
$MOR.CASE.NOM(t^7{}_{11}) \vee MOR.CASE.DAT(t^7{}_{11})$
$dom[MOR.CASE.NOM(t^7{}_{11}) \Rightarrow (SYN.PREC(t^7{}_{11}, t^7) \& SYN.NEAR(t^7{}_{11}, t^7))]$,
$rec[MOR.CASE.DAT(t^7{}_{11}) \Rightarrow (SYN.PREC(t^7{}_{11}, t^7) \& SYN.NEAR(t^7{}_{11}, t^7))]$,
$rec[INT.STRESS(t^7) \leftrightarrow INT.STRESS(t^7{}_{11})]$, $INT.STRESS(t^7)$

2.6 provides a thorough analysis of the noun above, too. It lacks only the last subsection here, which says that the nominal head of the possessor is stressed or unstressed parallel with the possession as a default. The current occurrence of the possession is stressed here.

(2)　(g)　$VISIT(v_9, v_{10})$, $HE/SHE/IT(v_9)$
　　　　　　$`pref_{up}$-seek-3sg-defObj: $<t^9, t^8; t^9{}_1, t^9{}_2, t^9{}_3, t^9{}_{31}>$

r1　　　$CAT.V.TR(t^9)$, $FINITE(t^9)$, $rec[\sim INT.STRESS(t^9)]$, $\sim INT.STRESS(t^9)$,

r2　　　$CAT.VPREF(t^8)$, $rec[LEG(t^8)]$, $\sim LEG.REF(t^8)$,

r3　　　$INT.STRESS(t8)$, $(SYN.PREC(t^8, t^9))$, $SYN.ADJAC(t^8, t^9)$

r4　　　$[\sim(t^9{}1 = t^9)]$ $((SYN.PREC(t^9{}1, t^9{}_2) \& INT.STRESS(t^9{}1))]$

r5　　　$LEG.PRED(t^9{}_2)$, $INT.STRESS(t^9{}_2)$, $rec[\sim(t^9 = t^9{}_2)]$,

r6　　　$[\sim(t^9 = t^9{}_2)]$ $((SYN.PREC(t^9{}_2, t^9) \& SYN.NEAR(t^9{}_2, t^9))]$

r7　　　$LEG(t^9{}_3)$, $CAT.N(t^9{}_{31})$, $MOR.CASE.ACC(t^9{}_{31})$, $rec[INT.STRESS(t^9{}_{31})]$,

r8　　　$\sim SYN.PREC(t^9{}_{31}, t^9{}_3)$, $SYN.NEAR(t^9{}_3, t^9{}_{31})$,

r9　　　$MOR.PER.3(t^9{}_{31})$, $LEG.REF.SPEC.DEF(t^9{}_3)$

Let us turn to the second clause, which has not been discussed in **2.6**. The first word is a prefixed finite (r1.2) transitive verb (r1.1) with two semantic argument slots for the visitor and the person visited (v_9 and v_{10} above). It consists of two own words: the verb stem (t^9) and a separable verbal prefix (t^8, r2.1). The verb stem is unstressed as a default (r1.3), and its current occurrence is unstressed indeed (r1.4). The verbal prefix is usually legitimate (r2.2) but its legitimacy cannot come from its referentiality (r2.3) but only from its serving as a verb carrier (Alberti 1996). It is stressed now (r3.1) and immediately precedes the verb stem (r3.2 3; ADJACent) so it is likely to be the verb carrier.

$t^9{}_1$ represents the potential topic of the sentence (r4), which is usually stressed and precedes the verb carrier. It will be turned out, however, that the clause in question contains no explicit topic so $t^9{}_1$ will be identified with the verb. Nevertheless, the semantic topic is assumed to be the referent of a nominative pronoun whose absence is permitted by the current version of the verb: the formula *HE/SHE/IT(v_9)* indicates this assumption (gender plays no role in Hungarian grammar). That is why there is no reference to a subject (but only to an object) in the morpholexical characterization of this version.

The potential verb carrier, which belongs to the lexical characterization of every finite predicator, is marked here with $t^9{}_2$. It is always predicative (r5.1) and stressed (r5.2), and immediately precedes the verb stem unless they coincide (r6). These three properties of verb carriers are inherited by every finite verb in the morpholexical inheritance network, whereas formula r5.3, which expresses the recessive condition that the verb carrier does not coincide with the verb stem, is inherited only by a subclass of verbs. The finite verb of the first clause ('marry a woman'), for instance, does not belong to this subclass because it definitely requires the main stress in a neutral sentence.

$t^9{}_{31}$ and $t^9{}_3$ in r7 9 represent the two relevant parts of the object: the nominal (r7.2) element in the accusative case (r7.3), which is stressed as a default (r7.4), and the determiner-like (r7.1) element, which immediately precedes the former (r8) unless they coincide. The conjugation of the verb deter-

mines the person feature (r9.1) of the nominal element of the potential object and the definiteness (r9.2) of its determiner-like element.

(2)   (h) THE(v_{11})
            the: $<t^{10}; t^{10}{}_1>$
r1   CAT.DET.ART.DEF(t^{10}), LEG.REF.SPEC.DEF(t^{10}),
r2   ~INT.STESS(t^{10}), ~INT.STRESS(t^{10}),
r3   CAT.N($t^{10}{}_1$), SYN.PREC($t^{10}, t^{10}{}_1$), SYN.NEAR($t^{10}, t^{10}{}_1$)

The definite article (r1.1 above) is characterized as an obligatorily unstressed element (r2.1), which can legitimize (r1.2: LEGitimate, REFerential, SPECific, DEFinite) a nominal expression (r3.1) that it immediately precedes (r3.2-3). Fortunately, the current occurrence of this definite article is unstressed indeed (r2.2).

(2)   (i) FIANCÉE(v_{12})
            'fiancée-acc: $<t^{11}>$
            CAT.N(t^{11}), MOR.CASE.ACC(t^{11}), INT.STRESS(t^{11})

The last word is a stressed noun in accusative.

Our second task is to verify that the morpholexical requirements concerning potential environments in a sentence can be satisfied. Thus we should point out that the sequence of stressed and inflected words given above does meet these requirements indeed. The verification amounts to an identification of each potential element referred to in the lexical characterizations with own words or other potential elements. Parallel with the establishment of this system of equations among morpholexical elements, another system of equations among semantic arguments can be established. Thus, a successful satisfaction of lexical conditions has two results: it will have been proved, on the one hand, that a grammatical sentence can be constructed from a numeration of stressed and inflected words, and a DRS will have been assigned to this sentence, on the other hand, where instances of co-predication have already been determined.

Let us review the parallel equation systems:

(3) r1 $t^1{}_1 = <t^3, t^4, t^5, t^6, t^7>$ $v_1 = \{$MARI(v_3), MARRY-A-WOMAN(v_4), AN(v_5),
                                                          FORMER(v_6), BOYFRIEND$_{poss}$(v_7, v_8)$\}$

r2 $t^1{}_2 = t^4$ gramm.
r3 $t^2{}_1 = <t^8, t^9, t^{10}, t^{11}>$ $v_2 = \{$VISIT(v_9, v_{10}), THE(v_{11}), FIANCÉE(v_{12})$\}$
r4 $t^2{}_2 = t^9$ gramm.
r5 $t^4{}_1 = t^3$ MARI(v3) is in the topic of the sentence whose finite verb
                                                          is MARRY-A-WOMAN(v_4)
r6 $t^4{}_2 = t^4$ gramm., no focus
r7 $t^4{}_3 = t^5$ $v_4 = v_5$
r8 $t^4{}_{31} = t^7$ $v_4 = v_7$
r9 $t^5{}_1 = t^7$ $v_5 = v_7$
r10 $t^6{}_1 = t^7$ $v_6 = v_7$
r11 $t^7{}_1 = t^3$ $v_8 = v_3$
r12 $t^7{}_{11} = t^3$ $v_8 = v_3$
r13 $t^9{}_1 = t^9$ gramm., no new topic ($v_3 = v_9$?)

r14	$t^9_2 = t^8$	gramm., no focus
r15	$t^9_3 = t^{10}$	$v_{10} = v_{11}$
r16	$t^9_{31} = t^{11}$	$v_{10} = v_{12}$
r17	$t^{10}_1 = t^{11}$	$v_{11} = v_{12}$

The IF-THEN operator is assumed here to be similar to a verb with two argument slots. Its specialty amounts to the fact that these arguments are referents that belong to entire discourses and are to be expressed as finite clauses. These finite clauses are characterized as sequences of words with a finite element. The sequence <Marinak, nősül, egy, régi, udvarlója> serves as an adequate realization of t^1_1 (r1) with *nősül* 'marry a woman' as a finite element (r2). The potential element t^2_1 can be identified with the sequence <fel, keresi, a, menyasszonyt> (r3) where *(fel-)keresi* 'visit' will play the role of the finite element required (r4). As for semantics, the sequences of words correspond to sets of semantic items, which can be regarded as (still underspecified) DRSs (r1, r3). The demonstration of finite elements (r2, r4) only amounts to the satisfaction of a requirement concerning grammaticality.

The potential topic (t^4_1) in the lexical characterization of the first finite verb can be identified with the stressed word *Marinak* 'Mary-dat' (r5). It will turn out soon that this word precedes the verb carrier indeed. Here no semantic consequence is attributed to the identification of an element with the topic (which is undoubtedly a simplification). It will be relied on, however, that the implicit topic of a clause equals to the topic of the previous clause as a default.

The verb carrier t^4_2 can be identified with the stressed verb (r6), which has been mentioned to be the default case. This fact is to be regarded as an indication of the absence of a focus operator.

The subsequent six equation pairs (r7–12) have already been commented in 2.6. The pieces of information mentioned only here corroborate the equations. The final result is that the boyfriend is the same as the person who is going to marry a woman (r7–8), Mary is the possessor of the boyfriend (r11–12), and the boyfriend is a "non-specific former boyfriend" (r9–10). These latter two statements will be interpreted a bit later.

Let us turn to the second clause and begin with the verb carrier (r14). The verbal prefix satisfies every requirement (it is stressed and immediately precedes the verb stem) so it serves as a verb carrier in the clause in question. This fact is an indication of the absence of a semantic focus operator.

What can play the role of the potential topic (r13)? The following simplified answer can be provided here. As nothing precedes the verb carrier in the clause, t^9_1 can be identified with the verb. As was mentioned, the incorporated subject of this clause is to be identified with the topic of the previous clause so Mary will be the topic of the second clause, too, and the implicit subject as well.

The definite determiner-like element is obviously the definite article (r15) so the definite entity (v_{11}) is the visited person (v_{10}). Fortunately, this element immediately precedes a noun in accusative, which can be identified with the nominal element of the potential object referred to in the lexical characterization of 'visit' (r16). The semantic result is that the visited person is to be identified with the fiancée ($v_{10} = v_{12}$).

Finally, the nominal element that belongs to the definite article (t^{10}_1) can be unified with the subsequent noun (t^{11}) so the definite entity is nothing else but the fiancée.

The equations among semantic arguments have been summarized below:

(4) $v_1 = \{...\}$:   the premise
     $v_2 = \{...\}$:   the conclusion
     $v_3 = v_8 = v_9$:   Mary, who will visit somebody
     $v_4 = v_5 = v_6 = v_7$: Mary's former boyfriend, who is going to marry a woman
     $v_{10} = v_{11} = v_{12}$: the fiancée

Now we are in a position to specify the two DRSs in the argument slots of the IF-THEN operator so we can list the set of semantic items corresponding to the numeration of lexical items:

(5)　{IF-THEN({MARI(v_3), MARRY-A-WOMAN(v_4), AN(v_4), FORMER(v_4), BOYFRIEND$_{poss}$(v_4, v_3)},
　　{VISIT(v_3, v_{10}), THE(v_{10}), FIANCÉE(v_{10})}), MARI(v_3), MARRY-A-WOMAN(v_4), AN(v_4), FORMER($v4$),
　　BOYFRIEND$_{poss}$(v_4, v_3), VISIT(v_3, v_{10}), THE(v_{10}), FIANCÉE(v_{10})}

The formula above serves as an input for the embedding of the sentence in the hearer's information state (3.3). Then the connection between the simple DRSs that constitute the input is interpreted as a conditional relation (3.3.10), primarily due to the discourse organizing capacity of the topic operator (3.3.12) and the IF-THEN operator (3.2.6). This conditional relation below is to be regarded as the result of the minimal embedding of the sentence:

(6)　　< MARI(v_3)?, { MARRY-A-WOMAN(v_4), FORMER(v_4)?, BOYFRIEND$_{poss}$(v_4, v_3)?},
　　　{VISIT(v_3, v_{10}), FIANCÉE(v_{10})?}>

The four question marks in the formula above are intended to indicate other elements' capacity for discourse organization, which will result in the process of embedding going on.

I. Proper names, likewise (other) definite expressions, require their temporary referent to be identified with an old referent of (an extension of) the hearer's information state before processing the new sentence. Thus the speaker assumes that the hearer's actual information state contains only one person named Mari.

II. The attributive construction *régi udvarló* 'former boyfriend' cannot be interpreted in a conjunctive way like this: "x is former and x is a boyfriend."

III. The possessive construction *Mari udvarlója* 'Mary's boyfriend' cannot be interpreted in a conjunctive way either. The conjunction "x is Mary's and x is a boyfriend" is at least underspecified because this formula does not express the fact that x belongs to Mary as a boyfriend (Alberti 1995, 1997b).

IV. The definite expression *a menyasszony* 'the fiancée' requires its temporary referent (v_{10}) to be identified with an old referent.

The following conditional relations (or similar ones), expressing lexical/cultural/encyclopedic knowledge, are to be assumed to be contained by the hearer's information state (3.2.6, 3.3.8):

(7)　(a) <BOYFRIEND$_{poss}$(x, y), COURT(x, y)>
　　(b) <COURT(x, y), COURT$_{time}$(x, y, t)>
　　(c) <{FORMER(x), COURT$_{time}$(x, y, t)}, {NOT-COURT$_{time}$(x, y, t_0), $t<t_0$}>
　　(d) <MARRY-A-WOMAN(x), {MARRY(x, y), FIANCÉE(y)}
　　(e) <MARRY(x, y), MARRY$_{time}$(x, y, t)>
　　(f) <VISIT(x, y), VISIT$_{time}$(x, y, t)>

The first conditional relation declares that the possessor of the boyfriend is a person that he courts. Row 4 demonstrates a similar piece of knowledge: the elementary DRS that belongs to *nősül* 'marry-a-woman' is associated with the DRS expressing that the person who gets married marries a person that can be called a fiancée. Rows 2, 5, 6 provide specialized versions of the general assumption that actions can be supplied with a time referent. Finally, the formula in r3 also expresses a special instance of a general rule, which concerns the use of 'former'. It says that 'former' is an assertion of a time referent. Precisely, if x is asserted to be 'former' and x courts y at a point of time t, then t precedes

the present time of the actual world (marked with t_0), and x bears a NOT-COURT relation to y at the present time (I have simplified here an idea in Partee 1984).

A question arises as to whether these pieces of information can be assumed to be at the hearer's disposal. Obviously, no lexicon or encyclopedia can be assumed to contain such a gigantic amount of knowledge. The hearer's information state, however, is not a lexicon but a store of earlier discourses. There must be a mechanism (of generalization) that turns pairs (or groups) of concrete DRSs associated in real discourses into abstract conditional relations. The hearer permanently attempts to interpret different discourses by means of arbitrarily combining (generalized versions (**3.3.6**) of) earlier DRSs, and saves the combinations that have proved successful in the course of an interpretation.

Suppose now the conditional relations in (7) are at the hearer's disposal. (S)he should open them in the way discussed in **3.3.8**. Here I show the final result, and then I am going to sketch the steps.

(8) $w^{\cdot} < w^{+} < w^{++}$
        r_M belongs to $w^{\cdot}$,
        r_S, r_F, t_0, t_1, t_2 belong to w^{+},
        t_3 belongs to w^{++}

(9) $<\{\text{COURT}_{\text{time}}(r_S, r_M, t_1), \text{NOT-COURT}_{\text{time}}(r_S, r_M, t_0), \text{MARRY}_{\text{time}}(r_S, r_F, t_2), \text{FIANCÉE}(r_F), t_1 < t_0 < t_2\},$
        $\{\text{VISIT}_{\text{time}}(r_M, r_F, t_3), t_0 < t_3\}>$

World $w^{\cdot}$ in (8) is the active world of the hearer's information state before processing the sentence. The Davidsonian referent of the sentence (**3.3.10**) belongs to this world. w^{+} denotes the absorbing world, and w^{++} denotes an even later world.

Mary's temporary referent (v_3 in (6)) is to be identified with an old referent that belongs to $w^{\cdot}$, as was mentioned. The speaker assumes that this referent, marked with r_M in (8), exists in the hearer's current information state. If this is true, the first conditional layer of the conditional relation in (6) can be satisfied and then can be deleted (see 9).

The temporary referent v_4 is to be identified with a new referent that can be assumed to belong to the absorbing world w^{+}. This new referent is marked with r_S, to remind of the word *boyfriend*. Now let us notice that the referents of the conditional relations in (7) belong to the absorbing world (since if a referent belongs to a certain world, it belongs to every later worlds). The (referent of the) DRS $\{\text{BOYFRIEND}_{\text{poss}}(x, y)[r_S, r_M]\}$ also belongs to w^{+}, so $\{\text{COURT}(x,y)[r_S, r_M]\} = \{\text{COURT}(r_S, r_M)\}$ belongs to the absorbing world as well. The content of (7.2 3) can be feed to the absorbing world likewise. The appearance of the referent of the fiancée is due to (7.4): referent y in (7.4) is to be identified with a new referent of w^{+} (3.3.6). t_0 in (9) denotes the current tense of the absorbing world, while t_2 denotes the time of the marriage. The computation of their relation is based on a thorough analysis of tense, aspect and verb type (not discussed here).

In the DRS $\{\text{VISIT}(v_3, v_{10}), \text{FIANCÉE}(v_{10})?\}$ in (6), the definite expression 'the fiancée' requires explanation. Where is the old referent for v_{10} to be identified with? There is one that belongs to the absorbing world but does not belong to $w^{\cdot}$: it is marked with r_F. The time of the visit can also be introduced; its referent t_3 will belong to (only) w^{++}. It obviously marks a point of time that follows the present time that belongs to w^{+} (the reasons are not discussed here).

I argue that (9) represents the final state of the embedding of sentence (1) in the hearer's information state because the sentence has a conditional form and a non-specific boyfriend is referred to in it. The interesting point is that the **definiteness** of the fiancée in the second clause is to be interpreted according to the world (that belongs to the DRS) of the first clause. So the hearer is not

assumed to have a salient fiancée in his information state before processing the sentence. No problem arises, fortunately, since the abstract conditional relations can be applied to the absorbing world as well.

Thus the sentence in (1) has been embedded in the hearer's information state as the conditional relation in (9), which provides the following information: If the hearer hears about a person with the properties that (1) he does not court Mary (a definite woman known by the hearer) that time, (2) but he courted her some time before, (3) and he intends to marry somebody a bit later (who, hence, can be referred to as a fiancée), then the hearer will know that (according to the speaker) Marry will visit this fiancée (also a bit later relative to the time of hearing about the fiancé).

The truth-conditional verification of the sentence (in the traditional sense) might be based on assuming an **ideal hearer**, who bears a total knowledge about the possible relations among the referents that belong to the active world of his current information state. The conditional sentence in (1) is true relative to w if Situation A below always (for any selection of the referents mentioned) entails Situation B (where r_M is a fixed referent that belongs to a definite person named Mari):

(10) Situation A: there are referents $r_S{}^\cdot$, $r_F{}^\cdot$, $t_0{}^\cdot$, $t_1{}^\cdot$, $t_2{}^\cdot$ that all belong to w, and
$\{$COURTtime$(r_S{}^\cdot, r_M, t_1{}^\cdot)$, NOT-COURTtime$(r_S{}^\cdot, r_M, t_0{}^\cdot)$, MARRYtime$(r_S{}^\cdot, r_F{}^\cdot, t_2{}^\cdot)$, FIANCÉE$(r_F{}^\cdot)$, $t_1{}^\cdot < t_0{}^\cdot < t_2{}^\cdot\}$ also belongs to an extension of w.
Situation B: there is a referent $t_3{}^\cdot$ that $\{$VISITtime$(r_M, r_F{}^\cdot, t_3{}^\cdot)$, $t_0{}^\cdot < t_3{}^\cdot\}$ belongs to an extension of the former extension.

Nevertheless, I think that everyday reasoning is based on other kinds of strategies.

References

Alberti, G. 1990. Quantifiers and the chronology of unification. Paper read at the 3rd Symposium on Logic and Language, Révfülöp, Hungary.

Alberti, G. 1995. Role assignment in Hungarian possessive constructions. In: Kenesei, I. (ed.): Approaches to Hungarian 5: Levels and structures, 11 28. JATE, Szeged.

Alberti, G. 1996. Perfect syntax and syntactic encoding. Paper read at the Research Institute for Linguistics, Hungarian Academy of Sciences. June 4.

Alberti, G. 1997a. Restrictions on the degree of referentiality of arguments in Hungarian sentences. In: Acta Linguistica Hungarica 44: 341 62.

Alberti, G. 1997b. Argument selection. Peter Lang, Frankfurt am Main.

Alberti, G. 1998. GASG: The Grammar of Total Lexicalism. To appear in: Working Papers in the Theory of Grammar. Theoretical Linguistics Programme, Budapest University (ELTE) and Research Institute for Linguistics, Hungarian Academy of Sciences.

Babarczy, A. 1996. The transition to the functional stage in Hungarian language acquisition. Paper read at the 3rd International Conference on the Structure of Hungarian. Published in 1998 in: Kenesei, I. de Groot, C. (eds): Approaches to Hungarian 6. JATE, Szeged.

Bartsch, R. 1987. Frame representations and discourse representations. In: ITLI Working Papers, Amsterdam.

Bresnan, J. (ed.) 1982. The mental representation of grammatical relations. The MIT Press, Cambridge MA.

Bresnan, J. — Kanerva, J. 1989. Locative inversion in Chichewa: a case study of factorization in grammar. In: Linguistic Inquiry 20: 1–50.

Brody, M. 1990. Remarks on the order of elements in the Hungarian focus field. In: Kenesei, I. (ed.): Approaches to Hungarian 3: Structures and arguments, 95–122. JATE, Szeged.

Chomsky, N. 1981. Lectures on government and binding. Foris, Dordrecht, Holland.

Chomsky, N. 1995. The Minimalist Program. The MIT Press, Cambridge MA.

Davidson, D. 1967. The logical form of action sentences. In: Rescher, N. (ed.): The logic of decision and action. 81–95. University of Pittsburgh Press, Pittsburgh PA.

Dowty, D.R. — Wall, R.E. — Peters, S. 1981. Introduction to Montague semantics. Dordrecht, Reidel.

É. Kiss, K. 1995. Definiteness Effect revisited. In: Kenesei, I. (ed.): Approaches to Hungarian 5: Levels and structures, 63–88. JATE, Szeged.

Enç, M. 1990. The semantics of specificity. In: Linguistic Inquiry 22: 1–25.

Groenendijk, J. — Stokhof, M. 1989. Dynamic Predicate Logic: towards a compositional, non-representational semantics of discourse. ITLI, Amsterdam.

Groenendijk, J. — Stokhof, M. 1990. Dynamic Montague Grammar. In: Kálmán, L. — Pólos, L. (eds): Papers from the 2nd Symposium on Logic and Language, 3–48. Akadémiai, Budapest.

Heim, I. 1982. The semantics of definite and indefinite noun phrases. Ph.D. dissertation. University of Massachusetts, Amherst MA.

Heim, I. 1983. File change semantics and the familiarity theory of definiteness. In: Bäuerle, R. — Schwarze, C. — Stechow, A. von (eds.): Meaning, use and interpretation of language, 164–90. De Gruyter, Berlin.

Heim, I. 1990. E-type pronouns and donkey anaphora. In: Linguistics and Philosophy 13: 137–77.

Kálmán, L. 1990. Deferred information: the semantics of commitment. In: Kálmán, L. — Pólos, L. (eds): Papers from the 2nd Symposium on Logic and Language, 125–57. Akadémiai, Budapest.

Kálmán, L. — Nádasdy, Á. 1994. A hangsúly [Stress]. In: Kiefer, F. (ed.): Strukturális magyar nyelvtan. 2. Fonológia [A structural Hungarian grammar. 2. Phonology], 393–467. Akadémiai, Budapest.

Kálmán, L. — Pólos, L. — Szabó, Z. 1989. Can representations solve problems? Technical report, Applied Logic Laboratory Inc., Budapest.

Kálmán, L. — Rádai, G. 1996. Discontinuous constituents in Hungarian from a constructionist perspective. Published in 1998 in: Kenesei, I. — de Groot, C. (eds.): Approaches to Hungarian 6. JATE, Szeged.

Kálmán, L. — Szabó, Z. 1990. D.I.R.T.: an overview. In: Stokhof, M. — Torenvliet, L. (eds): Proceedings of the Seventh Amsterdam Colloquium, 253–275. ITLI, University of Amsterdam.

Kamp, H. 1981. A theory of truth and semantic representation. In: Groenendijk, J. — Janssen, T. — Stokhof, M. (eds): Formal methods in the study of language, 277–343. Mathematical Centre, Amsterdam.

Kamp, H. — Reyle, U. 1993. From discourse to logic. Vol. I. Kluwer Academic Publishers, Dordrecht.

Kamp, H. — Rossdeutscher, A. 1994. Remarks on lexical structure and DRS construction. In: Theoretical Linguistics 20: 97–165.

Karttunen, L. 1976. Discourse referents. In: McCawley, J.D. (ed.): Syntax and semantics 7, 363–85. New York, Academic Press.

Kiefer, F. — É. Kiss, K. (eds) 1994. The syntactic structure of Hungarian. Syntax and semantics 27. New York, Academic Press.

Landman, F. 1986. Towards a theory of truth and information: the status of partial object in semantics. Foris, Dordrecht.

Oehrle, R.T. — Bach, E. — Wheeler, D. (eds) 1988. Categorial grammars and natural language structures. Reidel, Dordrecht.

Parsons, T. 1995 Thematic relations and arguments. In: Linguistic Inquiry 26: 635–62.

Partee, B.H. 1984. Nominal and temporal anaphora. In: Linguistics and Philosophy 7: 243–86.

Partee, B.H. – Meulen, G.B. ter – Wall, R.P. 1990. Mathematical methods in linguistics. Kluwer Academic Publishers, Dordrecht.

Roberts, C. 1989. Modal subordination and pronominal anaphora in discourse. In: Linguistics and Philosophy 12: 683–721.

Szabolcsi, A. 1995. Strategies for scope taking. In: Working Papers in the Theory of Grammar 2/1. Theoretical Linguistics Programme, ELTE, Budapest.

Szépe, Gy. 1964. Az egyeztetés kérdéséhez [Comments on agreement]. In: Tanulmányok a magyar nyelv életrajza köréből [Studies on Hungarian], 355–9. Budapest.

Wescoat, M.T. 1987. Practical instructions for working with the formalism of Lexical Functional Grammar. Stanford University, Stanford CA.

Williams, E. 1995. Theta Theory. In: Webelhuth, G. (ed.): Government and Binding Theory and the Minimalist Program. Principles and parameters in syntactic theory, 99–124. Basil Blackwell, Cambridge MA.

Address of the author: Gábor Alberti
                            Janus Pannonius University
                            Department of Linguistics
                            H–7624 Pécs, Ifjúság útja 6.
                            Hungary
                            E-mail: albi@btk.jpte.hu

THE OUTLINES OF A METRICAL SYNTAX OF HUNGARIAN

LÁSZLÓ HUNYADI

Abstract

The paper addresses the question of how and to what extent syntax, prosody, communicative and logical functions are related. It is assumed that prosody plays a more general role in indicating scope relations than surface configuration.

The basic assumptions of a metrical syntax of Hungarian are outlined according to which only those sentences are grammatical which are assigned a proper prosodic structure. The prosodic rules are based on rules of stress reduction between adjacent prosodic components. It is shown that stress reduction does not automatically follow rhythmic rules, it is rather constrained by a hierarchy of categories available for reduction.

The relation of important communicative functions of a sentence to its prosody, such as topicalization and focusing, including multiple foci is discussed demonstrating that the prosody of a sentence is both determined by its logical and communicative functions.

Introduction

It was observed in Hunyadi (1981a) that, in Hungarian, there is a certain relation between linear order and scope interpretation. A wide scope operator precedes its scope in such sentences as (1), where the universal quantifier *mindenki* has wide scope over the focused *Péterrel*.

(1) [$_Q$ Mindenki [$_F$ Péterrel [beszélgetett.]]]
     everyone Peter-with talked
    'Everyone talked to Peter'

It was also noticed that there is a certain relation between stress and the function of scope assignment. Namely, a wide scope operator 'takes over' the stress of the element in its immediate scope. This can be seen in the comparison of the following pair of sentences (capital letters indicate the main stress-bearing word):

(2) (a) [F PÉTERREL [beszélgettem.]]

        Peter-to   I talked

        'It was Peter who I talked to'

    (b) [F NEM   Péterrel [beszélgettem.]]

        not   Peter-to I talked

        'It was not Peter who I talked to'

Apparently, the focused *Péterrel* in (2a) has main stress, whereas, when it is in the scope of the negative *nem*, it is unstressed (as in (2b)) with *nem* 'taking over' the stress of *Péterrel*.

It was also pointed out that the indication of the scope of an operator by stressing the operator appears to be more general than linear scope assignment. Regardless of linear order, the universal quantifier has wide scope in (3a–b) and narrow scope in (4a–b):

(3) (a) PÉTER evett meg MINDENT.

        Peter   ate   CONV everything-acc

        'For every x, it was Peter who ate x'

    (b) MINDENT   Péter evett meg.

        everything-acc Peter ate   CONV

        'For every x, it was Peter who ate x'

(4) (a) PÉTER evett meg mindent.

        Peter   ate   CONV everything-acc

        'It was Peter who ate everything (others may have eaten less)'

    (b) Mindent   PÉTER evett meg.

        everything-acc Peter   ate   CONV

        'It was Peter who ate everything (others may have eaten less)'

As can be seen in (3) and (4), it is the stressed/unstressed character of the quantifier rather than its linear position that suggests its wide/narrow scope, respectively.

These and similar examples indicate that there is a relation between syntactic structure and scope assignment, on the one hand, and prosodic structure and scope assignment, on the other.

Apparently prosody (including stress and intonation) does not only take part in the assignment of scope, it plays yet another, probably even more general role, i.e. the denotation of certain communicative functions, specifically, focusing and topicalization. Since, in Hungarian, in addition to their prosodic difference, the communicative functions of **focus** and **topic** are built into syntax with separate positions for each of them (cf. É. Kiss 1978), we can assume a certain interdependence between the syntactic structure, prosodic structure, communicative and logical functions of a Hungarian sentence. With both syntactic and prosodic structure having formal properties the question may arise to what extent each of them participates in the denotation of the communicative and logical functions of a sentence. The question may also arise whether syntax and prosody only meet in their 'joint venture' of assigning the communicative and logical functions of a sentence or whether there is also a direct relation between syntax and prosody, i.e. whether at least some aspects of either of them are directly determined by the other. Finally, we may also ask whether the communicative and logical functions of a sentence only meet via the mediation of syntax and prosody or they can also be directly related; cf. (5):

(5)
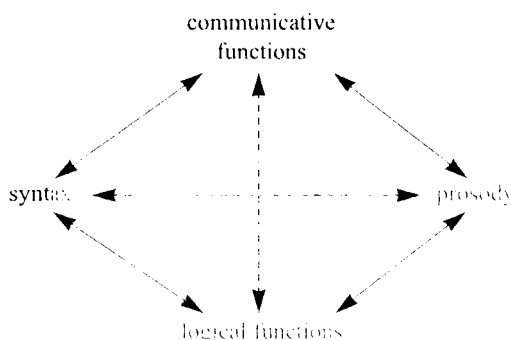
Out of the potential binary relations above, É. Kiss (1987) studies the relation of syntax and its communicative and logical functions. In her account, syntax directly serves the above functions, i.e. the communicative functions of **focus** and **topic** are expressed in the corresponding syntactic positions F and T, respectively. As for the logical functions of a sentence, her **precedence principle** supports the view that it is linear order based on c-command that determines scope (either directly or via the traces of moved operators). In the case of the wide scope postverbal operator (as in (3a) *PÉTER evett meg MINDENT*) she assumes a stylistic movement in Phonetic Form. In her recent work (É. Kiss 1998a; 1998b) she assumes that in sentences with multiple foci all movement takes place at the level of S-structure. She claims that the preverbal field consists of a multiple of [FP-QP-TopP] projections

and these positions are filled in recursively. Maintaining the precedence principle, her example (6a) correctly yields the desired interpretation. It does not, however, seem to work on (6b), in which the postverbal quantifier phrase *minden ételből* has wide scope over the preceding focus phrase *csak János*:

(6) (a) Csak JÁNOS vett minden ételből   csak KÉTSZER.
         only John   took every   meal-from only twice
         'It was only John who had twice from every meal'
     (b) Csak JÁNOS vett   MINDEN .ételből.
         only John   took every   meal-from
         'From every meal, it was only John who had'

The comparison of (6a) and (6b) suggests that the syntactic status of an element (in these cases the fact that *minden ételből*, a quantified NP, is in a certain syntactic position (here the Spec of QP) is not sufficient to associate this position with a certain absolute (wide) scope. On the other hand, the difference in the prosodic realization of the phrase *Csak János vett minden ételből* in (6a) and (6b), i.e. the fact that the condition for the quantified NP *minden ételből* to have wide scope over the focused phrase *csak János* is to have main stress, indicates that prosody must play a role in the assignment of scope.

In another approach, Kornai and Kálmán (1988) study the prosodic properties of Hungarian sentences presenting a model of the Hungarian intonational system within the framework of autosegmental phonology. In this work, describing accent rules to yield the proper prosody of sentences, they assume that there is a direct relation between syntactic structure and sentence prosody: Kornai and Kálmán assign a diacritic f (focus, or in their reading: Foykes) marking to certain syllables. The formation of intonational patterns is the result of the Eradication rule, a rule which deletes the accent and word boundaries of the subsequent words up to the next diacritic f marking. As a result of the Eradication rule segments from one f to the next or to the end of the phrase form one phonological word with one accented syllable.

Kornai and Kálmán do not make a distinction between quantified and focused constructions, and, as a consequence, they cannot generate the otherwise possible prosody of (3b) [$_Q$*MINDENT* [$_F$*Péter* [*evett meg*]]]. The Eradication rule only deletes the accent to the next diacritic f marking, and if both *mindent* and *Péter* are assigned the diacritic f, the second element with the diacritic f (*Péter*) cannot be unaccented (i.e. this rule excludes the deaccenting of an element with the diacritic f in general, a property characteristic of narrow focus).

The relation of syntactic and prosodic structure is addressed in Vogel–Kenesei (1987; 1990), presenting a metrical phonological account of how syntax, prosody,

and the expression of logical scope are related to one another. They maintain that intonational phrases, derived from phonological phrases, can be identified at the level of S-structure. Certain elements are marked as [+SC] for wide scope quantifiers and others as [+OS] (for 'operator status'). Phonological phrases are grouped into intonational phrases from left to right, starting from the element with the widest scope, until another element with a logical function (either marked [+SC] or [+OS]) or the end of the sentence is reached.

Since, in their account, relative scope follows the left-to-right order, and so wide scope is always denoted by an operator on the left, from their hypothesis it follows that if two sentences consist of the same IPs but the relative order of the IPs is different, they will also have different scope relations:

(7) (a) [IP MINDEN nyúl] [IP JÚLIÁT szereti a   legjobban]
            every   rabbit   Julia-acc loves   the most
        'For every rabbit, it is Julia that it likes best'

     (b) [IP JÚLIÁT szereti][IP MINDEN nyúl a legjobban]
        'It is Julia that every rabbit likes best'   (Vogel–Kenesei 1990, 360)

In fact, (7b) is an example of Quantifier postposing. The scope of such a stressed quantifier has been described as having wide scope over the focused element (cf. Hunyadi 1981b; É. Kiss 1987), thus the proper gloss of (7b) should rather be equivalent to that of (7a), regardless of surface word order differences. This fact indicates again that the rule of left-to-right scope assignment has its restrictions.

Both the Kálmán–Kornai and the Vogel–Kenesei phonological approaches are based on the assumption that certain elements have an absolute operator status (determined by their syntactic position and/or their lexical–semantic properties). Since cases of unstressed universal quantifiers are not treated in these models, the unstressed instances of the quantifier *mindent* in (4a) *PÉTER evett meg mindent* and (4b) *Mindent PÉTER evett meg* with narrow-scope reading cannot be accounted for either.

Our data suggest that there is indeed a certain relation between syntactic structure and scope assignment on the one hand, and syntactic structure and prosody on the other. The data, however, also suggest that neither scope relations nor prosodic patterns can exclusively be predicted from the proposed syntactic analyses. The relation of the formal and semantic aspects of a sentence appears to be more complex: it may well be the case that both the communicative and logical aspects of a sentence are expressed on more than one interacting level, including the levels of the lexicon, syntax, and prosody. In what follows, we are going to explore some of the properties of this interaction.

1. Stress and metrical phonology

We are going to apply to Hungarian syntax a framework of metrical phonology developed in Liberman–Prince (1977) and in subsequent work. According to this theory, stress is not considered to be a phonological feature, it is rather captured in a rhythmic structure. This structure is hierarchically organized and the various patterns are the result of certain phonological rules.

Following Halle–Vergnaud (1987), a sequence of syllables *1, 2, 3, 4* will receive a metrical constituent structure as in (8):

```
(8)          .          x
       (x     .     x    .)
       (x    x)   (x    x)
        1     2     3    4
```

First syllables *1* and *2*, on the one hand, and syllables *3* and *4*, on the other, are concatenated to form the substrings 1^2 and 3^4, respectively. The complete string $1^2 3^4$ is the result of the concatenation of the two previous substrings.

This way of representation reflects an important property of the given phonological operations: the order of their application is not arbitrary, it is rather constrained by the existence of a constituent structure at the phonological level of linguistic representation.

In metrical stress theory a sequence of stresses is represented in a metrical grid intended to denote temporal (rhythmic) structure, such as in (9):

```
(9)  x
     x                              x
     x                 x            x                       x
     x       x         x        x   x       x       x       x
     x   x   x   x      x   x    x   x   x   x   x   x   x   x   x
     1       2          3        4       5       6       7       8
```

The above sequence of syllables is grouped in a sequence of beats represented by columns. The height of each column corresponds to a certain degree of strength which is relative rather than absolute (the height of a column depends on how many levels can be created which in turn is directly related to the number of syllables in the sequence). The rows are as relevant as the columns: on each level each stress marking stands for a beat unit determined by the overall beat structure of the given row, thus creating a rhythmic structure of the row (or *layer*, following Halle and Vergnaud).

Generalizing characteristics of rhythm, Hayes (1995) notes the following: (a) rhythmic structure is **hierarchical**, with sequences of beats having multiple levels of strength; (b) it has a tendency to **even spacing** at all intervals of repetitions; (c) it respects **downward implication**. The latter means that any beat on a high layer must also serve as a beat on all lower layers. In practice this implies that if a syllable is stressed on a high level then it must have been stressed on the underlying layer as well.

Although significant work has been done on the syllable (word) level (e.g., Selkirk 1980) and there are investigations into the phrasal level as well (cf. Selkirk 1984, Halle–Vergnaud 1987, or most recently Hayes 1995), these studies have not been specifically directed to the description of how the syntactic and prosodic properties of a sentence are related to its primary logical representation. Since, in Hungarian, there is an apparent relation between these aspects of the language, and since stress appears to have an important share in all of them, a metrical model of syntax may capture their relation in one single approach. In what follows, we are going to describe the main properties of such a metrical model of Hungarian syntax.

2. Proposal: a metrical model of Hungarian syntax

Since many important aspects of the description of Hungarian syntax are associated with reference to prosody (e.g. elements in the **focus** position are characterized by heavy stress, those in the **topic** position by a characteristic intonation; cf. É. Kiss 1987 and scope assignment is also related to prosody; cf. Hunyadi 1981b, 1996; for a detailed analysis of Hungarian intonation cf. Varga 1996), it may appear reasonable to study the relation between syntactic and prosodic structure in a metrical model.

Although metrical phonology is essentially concerned with properties of the syllable structure and although metrical structure on the syllable level basically follows rhythmic rules (such as even spacing at all intervals of repetition), this does not always seem to be the case on the syntactic level, cf. (10a) and (10b):

(10) (a) x x   x   (AZT aKARtam MONdani)
   x x x x x x x
   azt akartam mondani
   that wanted-I to say
   'I wanted to say'

```
(b)*x              x       x           *(AZT János MONDta MINdig)
    x     x        x       x
    azt   János    mondta  mindig
    that  John     said    always
    'John always said'
```

But the fact that a Hungarian word has a single (lexical) stress makes it possible for us to assign each word one column in the metrical grid (as if it consisted of just one syllable); cf. (10c):

```
(10) (c) x                x              (AZT akartam MONdani)
         x    x           x
         azt  akartam     mondani
```

Apparently, the metrical representation of sentence prosody, in addition to general phonological rules, obeys its own rules, however; cf. (10d):

```
(10) (d)*x             x                *(AZT János MONDta mindig)
         x    x        x       x
         azt  János    mondta  mindig
```

The aim of the present paper is to describe those phonological rules of metrical representation by the help of which the possible prosodic representations of Hungarian sentences can be derived.

The model to be presented operates upon the output of a propositional component which has already taken care of the propositional composition of the sentence, including the formation of the NPs. The order of the arguments is left undefined. The task is to identify all and only the grammatical permutations of the arguments in the given sentence. It is assumed that the grammaticality of a permutation is constrained by prosodic rules, namely, by whether a phonological structure can be assigned to it. Only those with a proper phonological structure can be grammatical.

Below sentences are presented in a metrical grid format. Each word in a sentence will be considered to have one stress only (i.e. every phonological word has already undergone metrical reduction on the level of syllable structure). Since Hungarian is a left prominent language (word stress falls on the initial syllable and the first element of compounds carries main stress), column formation will proceed from right to left.

Each layer is intended to represent a valid prosodic structure of the given sentence. Although the height of columns can be associated with the indication of relative strength of stress, throughout this paper we will only be concerned with the horizontal reading of each layer. The lack of a stress mark in a grid will indicate that the stress of the given word, present at the previous layer, has been reduced.

Let us consider the following examples:

```
(11) (a) * x      x                    layer 2
           x    x          x           layer 1
         Látta János a    kiállítást.
         saw   John  the exhibition
         'John saw the exhibtion'
     (b)        x                       layer 3
         (Bár)  x              x        layer 2 (bár = 'although')
                x    x         x        layer 1
              Látta János a kiállítást.
```

(*Bár* 'however' in this and further examples is intended to indicate that such sentences are only grammatical as an embedded sentence of the kind *Bár látta János a kiállítást, egyéb dolog nem történt* 'Although John saw the exhibition, nothing else happened'.)

```
(12) (a) * x         x                 layer 2
           x         x        x        layer 1
         Látta a   kiállítást  mindenki.
         saw   the exhibition everyone
         'Everyone saw the exhibition'
     (b)        x                       layer 3
         (Bár)  x              x        layer 2
                x    x         x        layer 1
              Látta a kiállítást mindenki.
```

The grammaticality/agrammaticality of (11) and (12) demonstrates that the column formation between adjacent stress marks of the same layer does not automatically follow the rhythm rule on the basis of which the metrical constituent structures in (8) and (9) could be derived. The reduction of the stress on the right of two adjacent phonological words was possible in the (b) sentences whereas it was ruled out in the (a) variants.

(13) (a) * x   layer 3
         x   x   layer 2
         x   x   x   layer 1
      János elment a   kiállításra.
      John   went   the exhibition-to
      'John went to the exhibition'
    (b) * x   x   layer 2
         x   x   x   layer 1
      János elment a   kiállításra.

(14) (a) x   layer 3
         x   x   layer 2
         x   x   x   layer 1
      Látta mindenki a   kiállítást.
      saw   everyone the exhibition
      'Everyone saw the exhibition'
    (b)   x   layer 3
      (Bár)   x   x   layer 2
         x   x   x   layer 1
           Látta   mindenki a   kiállítást.

It is reasonable to assume that in determining the rules of prosodic derivation properties other than phonological ones should also be considered. On the basis of the above examples we propose the following poperties of the derivation of sentence prosody:

(15) Properties of derivation of sentence prosody
    (a) In the underlying structure every word has its own stress and represents a single phonological word. Rules of generation are reduction rules describing how the stress of a given word is reduced.
    (b) In the course of stress reduction phonological words are combined into intonational phrases and on higher levels reduction takes place between intonational phrases and phonological words or further intonational phrases.
    (c) Stress reduction only takes place between adjacent phonological components so that, in case of two adjacent phonological components C_1 and C_2 (C_1 preceding C_2), C_2 undergoes stress reduction.
    (d) The combination of C_1 and C_2 into one intonational phrase (the reduction of the stress of C_2) depends on the **hierarchy of the categories** C_1 and C_2.

The reduction of the stress of phonological words depends on their operator status. Operators are organized into a hierarchy which fundamentally determines the order and direction of stress reduction; cf. (16):

(16) Hierarchy of categories for stress reduction (in increasing order of ranking):
verb > non-quantified NP > quantified NP > sentential operator.

The condition of stress reduction is formalized in (17):

(17) Conditions of hierarchical stress reduction:
(a) In a sequence of phonological components C_1 and C_2, C_1 is the trigger and C_2 is the domain of stress reduction.
(b) C_2 undergoes stress reduction if it is lower on the hierarchy than C_1.
(c) Stress reduction of phonological words is obligatory in cases of cliticization and in other cases it is optional.

As the above data suggest, stress reduction does not directly follow from the Nuclear Stress Rule (Chomsky–Halle 1968). The rules of metrical grid formation are also constrained by the above hierarchical rule of stress reduction.

Thus, comparing (11a) and (11b), (11a) has an ungrammatical derivation on layer 2 because in the sequence [*Látta*] *János a kiállítást* the reduction of the stress of *a kiállítást* in *[*Látta*] *JÁNOS a kiállítást* is ruled out by the fact that this phonological word is not lower on the hierarchy than the preceding *János*. The reduction in the phrase *LÁTTA János* [*a kiállítást*] in (11b) is grammatical, because the verb, the carrier of the sentential operator 'assertion, identification' is higher on the hierarchy than the non-quantified NP *János*.

The ungrammaticality of the second layer derivation in (12a) *LÁTTA a KIÁLLÍTÁST mindenki* is again accounted for by the violation of the hierarchical constraint in the sequence *a kiállítást mindenki* (the stress of the universal quantifier cannot be reduced to the preceding non-quantified NP). In (12b), on the other hand, the hierarchy of stress reduction is respected both on layer 2 and layer 3 derivation. First the stress of the non-quantified NP is reduced to the verb (the carrier of the sentential operator) and the stress of the postverbal universal quantifier *mindenki* is unchanged (i.e. remains stressed; hence the perception of a double stress construction). On layer 3 the stress of the universal quantifier is reduced to the verb (the sentential operator) which again corresponds to the hierarchical constraint.

The hierarchical position of the surface sequence of all phonological words *Látta mindenki a kiállítást* both in (14a) and (14b) allows for a proper derivation. This example demonstrates that, although the direction of stress reduction is from

right to left, it is not obligatory to start reduction from the very end of the sentence
(cf. layer 2 of (11a)). This condition makes it possible to derive prosody with an
additional stress at the end of the sentence, too. (The derivation of prosody with
final stress only is ruled out by the right-to-left direction of reduction based on left
prominence in Hungarian.)

The ungrammaticality of the top layer of both (13a) *JÁNOS elment a kiál-
lításra and (13b) *JÁNOS elment a KIÁLLÍTÁSRA indicates that no stress reduc-
tion can take place between two adjacent phonological components with the same
hierarchical value: both János and el are non-quantified NPs and occupy the same
hierarchical position.

If C_1 and C_2 are on the same hierarchical level, no reduction can take place
even between two universal quantifiers; cf. (18a) and (18b):

(18) (a) * x   layer 3
          x   x   layer 2
          x   x   x   layer 1
        Mindenki mindent   megnézett.
        everyone everything-acc saw
        'Everyone saw everything'
    (b) * x   x   layer 2
          x   x   x   layer 1
        Mindenki mindent megnézett.

The conjunction is 'also' is a sentential operator; cf. (19):

(19)   x   layer 3
          x   x   layer 2
          x   x   x   layer 1
        Péter is   mindent   megnézett.
        Peter too   everything-acc saw
        'Peter, too, saw everything'

Since in a sequence with the universal quantifier preceded by an is-expression
stress reduction can take place, we have to conclude that is is ranked higher in the
hierarchy than the universal quantifier. That is is thus listed among sentential oper-
ators can be accounted for by the fact that with the conjunction is two propositions
are conjoined, thus representing a sentential operation.

In the case of negative sentences the negative nem 'no' or ne 'do not [imp.]' is
normally stressed and as such appears to be higher on the hierarchy than the non-

quantified or quantified NPs; cf. (20) and (21), suggesting that *nem* represents sentential operation.

(20) NEM Péter ment el a kiállításra.
     'It was not Peter who went to the exhibition'
(21) NEM mindenki ment el a kiállításra.
     'Not everybody went to the exhibition'

In (22), however, the absence of stress on *nem* suggests that, having undergone stress reduction, it represents narrow scope negation.

(22) PÉTER nem ment el a kiállításra.
     'It was Peter who did not go to the exhibition'

On the basis of these examples we assume that there are two kinds of negation: sentential and predicate negation. Sentential negation is an operation with the highest position in the hierachy of stress reduction (hence the grammaticality of (20) and (21)). Predicate negation is lower than the non-quantified NPs but higher than the verb in the hierarchy (hence the grammaticality of (22).

The fact that (23) is ungrammatical, however, follows a possibly universal constraint: most universal quantifiers of the type *minden-* must have narrow scope in relation to negation.

(23) *MINDENKI nem ment el   a   kiállításra.
      everyone   not went CONV the exhibition-to

As a consequence of stress reduction, a stressed word forms one prosodic phrase (phonological word) with the subsequent unstressed word. Whereas non-quantified NPs and universally quantified NPs can be the heads of such phrases, existentially quantified NPs cannot; cf. (24a–d):

(24) (a) JÁNOS látta a   kiállítást.
         John   saw the exhibition
         'It was John who saw the exhibition'
     (b) MINDENKI látta a   kiállítást.
         everyone   saw the exhibition
         'Everyone saw the exhibition'
     (c) *VALAKI látta a   kiállítást.
          someone saw the exhibition

(d) Valaki LÁTTA a kiállítást.
   'Someone saw the exhibition'

But this restriction only refers to constraints of the existential quantifier as trigger of stress reduction, its stress can be reduced, cf. (25a):

(25) (a)   x   layer 3
          x   x   layer 2
          x   x   x   layer 1
       Mindenki látott valakit.
       everyone saw   someone
       'Everyone saw someone'

On layer 2 the stress of the verb is reduced (the existential quantifier is higher on the hierarchy than the verb, so its stress cannot be reduced here). On layer 3, however, the stress of the existential *valakit* can already be reduced.

Since both *mindenki* and *valaki* are quantifiers and are thus on the same hierarchical level, the quantifier *mindenki* cannot be the trigger of stress reduction on layer 3 of (25) by itself; stress reduction only takes place due to the fact that in sentence-initial position *mindenki* serves also as the carrier of the sentential operation of *assertion*. Since, as it was shown above, an existential quantifier cannot receive main stress, it cannot be the carrier of the operator of assertion in sentence-initial position, either. Hence the ungrammaticality of (24c) above and of (25b):

(25) (b) *VALAKIT látott mindenki.

The above properties of the existential quantifier are further demonstrated in focused sentences like (26a) where the reduction in the sequence *valakit látott* cannot take place because the existential does not trigger reduction; and (26b) where reduction in the phrase *János valakit* cannot take place because *János*, the 'trigger' of reduction is lower on the hierarchy.

(26) (a) * x   x   layer 2
          x   x   x   layer 1
       János   valakit látott.
     (b) * x   x   layer 2
          x   x   x   layer 1
       János   valakit látott.

3. Communicative functions vs. stress reduction rules

Consider the following sentences:

(27) (a) * x   x   layer 2
         x   x   x   layer 1
        János olvasta a   könyvet.
        John   read   the book-acc
        'John read the book'

    (b) * x   x   layer 2
         x   x   x   layer 1
        János olvasta a   könyvet.
        John   read   the book-acc
        'John read the book'

    (c) x   layer 3
         ?   ?   ?   layer 2
         x   x   x   layer 1
        János olvasta a   könyvet.
        John   read   the book-acc
        'John read the book'

The question is: if both layer 2 of (27a) and layer 2 of (27b) is ruled out by the hierarchical constraint, what is the derivation structure of sentence *JÁNOS olvasta a könyvet* (layer 3 of (27c)).

If we consider the communicative function of an unstressed element in sentence final position (like *a könyvet* in (27c)), we notice that these elements always express a piece of **known information,** as part of the **theme** of the sentence. Since the denotation of known information is part of the communicative function of the sentence, it is reasonable to assume that this function is not subject to any categorial or hierarchical constraint. This part of the sentence may consist of an unlimited number of lexical items with equal stresses, representing the **neutral component** of the sentence. This neutral component can be prosodically considered a complex intonational phrase which undergoes stress reduction in such a way that each of its constituents undergoes stress reduction. Since the determining factor in the realization of stress reduction here is the communicative function of the phrase rather than the hierarchical characteristics of each of the constitutents, we need not expect hierarchical constraints of prosodic phrasing to apply. (This operation resembles, e.g., assigning the proper intonation to a yes-no question, in which each of the intonational phrases of the sentence is subject to the same prosodic operation.)

In order to allow the unconstrained derivation of prosodic phrases for the expression of communicative functions we will introduce a rule of **neutralization** for the derivation of the **neutral component** as part of the known information:

(28) Neutralization: derivation of the neutral component:
Form a continuous segment of sentence-final phonological words expressing known (redundant) information.
Apply unconstrained stress reduction to each of the constituents of the neutral component.

By separating this neutral phrase from the rest of the structure it is assured that the hierarchical rules of stress reduction can still be maintained for the focus part of the sentence.
Thus (27c) *JÁNOS olvasta a könyvet* can be derived as (27d):

(27) (d) x   layer 3
          x   x   layer 2
         (x   x   ) (x   )   layer 1
         János olvasta a   könyvet.

Since neutralization is not hierarchically constrained, it also allows for the stress reduction of a longer sequence of (even hierarchically equal) elements:

(29)   (JÁNOS olvasta) (a   könyvet   tegnap   délután.)
          John   read   the book-acc yesterday afternoon
          'It was John who read the book yesterday afternoon'

The above rules of stress reduction based on a hierarchy of elements and right-to-left order properly generate prosodic phrases in which the initial element of the phrase is stressed (a universal quantifier, a non-quantified NP or a verb). On the other hand, the same rules do not properly generate the prosodic structure of elements with contrastive topic function (**Left Dislocation** in the sense of É. Kiss 1987). There are two reasons for this: first, the minor accent (or quasi-unaccentedness) characteristic of a topicalized element can only be achieved by a stress reduction from the opposite (left-to-right) direction (since this topicalized element is the leftmost element of the sentence), second, it is obvious that even if this opposite direction of stress reduction is allowed, the hierarchical rules will not work:

(30) (a) János MINDENKIT látott.
     (b) Mindenkit JÁNOS látott.
     (c) Mindenkit NEM látott János.

Whereas in (30a), regardless of the opposite direction of reduction, the hierarchical constraint allows for the reduction of the stress of a non-quantified NP by a quantifier, this constraint would be violated in (30b) and (30c).

The specific nature of contrastive topicalization can be shown by the fact that different categories (both quantified and non-quantified NPs) can serve as topic and their stress is reduced regardless of their predefined hierarchical relation to the following adjacent (stressed) element.

Similarly to constituents in the neutral part of the sentence, the topic phrase can also consist of an unlimited number of constituents, which make up the complex topic-phrase. By default, each topic-constituent has the specific topic-intonation with some possible minor, secondary stress. This intonation contour, however, does not make up a complete intonation phrase: it never arrives at L (low) tone, the indicator of the right boundary of a (non-interrogative) intonational phrase. In order to be completed, a topic phrase should always be followed by a focus phrase. Although it is a general prosodic property of topic constituents not to carry main stress, even to be fairly unstressed, what makes this phrase prosodically distinct is its intonation contour and (partial) stress reduction is just a consequence of the formation of the proper intonation. This is why the hierarchical constraints of stress reduction do not have to apply.

Pre-focus element(s) can represent either contrastive or non-contrastive topic. In (31a) *tegnap* has not undergone contrastive topicalization and maintains its default stress and default neutral intonation. (31b), on the other hand, is an example of topicalization:

```
(31) (a)   x          x
           x          x          x
           x          x    x     x
         'Tegnap    'tüz  ütött  ki.
          yesterday  fire  broke  out
         'A fire broke out yesterday'
     (b)              x
                      x          x
                      x    x     x
         ( x    )   ( x    x     x)
         Tegnap    'tüz  ütött  ki  (, ma  'vihar  volt.)
         yesterday  fire  broke  out  today  storm  was
         'Yesterday, a fire broke out (today it was stormy)'
```

Thus the underlying prosodic-communicative structure of the sentence consists of three main parts: the **topic**, denoting the logical subject of the sentence, the **focus**,

the informational center of the sentence, and the unmarked **neutral part**. According-
ly, the Hungarian sentence consists of the following basic segments determined
prosodically and interpreted communicatively:

(32) The underlying prosodic-communicative structure of the Hungarian sentence:
      [*topic*]* [*focus*]* [*neutral part*]

Topicalization takes place recursively, i.e. more than one constituent can receive the
sentence initial topic function. The topicalized components are phonologically
coordinated, which means that each component has its own intonational phrase and
the intonational phrases are of the same type. The **topic** is followed by a complex
stucture of **focus**. Phonologically it is a subordinate phrase with the (leftmost) trig-
ger of stress reduction as head and the subordinated prosodic components with
reduced stress. Focus is followed by the communicatively **neutral** part. Whereas
both the **topic** and the **focus** are prosodically marked (the focus has its own stressed
head and initiates an intonational contour, the topic, too, has a special intonational
contour), the **neutral part** is prosodically unmarked: it has no stress or specific
intonation (its contour is flat).

From a categorial point of view, **focus** can equally be expressed by a universal
quantifier, a non-quantified NP and a verb. Whether these elements do express the
above function is determined by the hierarchical constraint of stress reduction.

As for topicalization, its understanding can offer an account for the descriptive
observation according to which if a sentence contains a topic, then it is obligatori-
ly followed by a focus (a quantified or non-quantified NP or a stressed verb). The
reason is obvious: since topic is prosodically denoted by stress reduction, from its
sentence intitial position it follows that the only way to reduce its stress is to reduce
it onto the following element on the right. The element receiving this reduced stress
will in turn be focused.

4. The semantic property of 'in situ' focus.
An instance of sentential vs. cross-sentential operations

As it was shown in section **2**, multiple focus sentences can simply be derived by
applying stress reduction to the preverbal focus and leaving the default stress of the
'in situ' focus unreduced. The two foci have different semantic functions: the pre-
verbal (operator) focus expresses exhaustiveness and the *in situ* (information or
presentational) focus does not express exhaustiveness (cf. É. Kiss 1998a; 1998b).

Considering the contexts in which sentences with the two kinds of focus appear, we assume that both kinds of focus express some sort of contrast. The kind of exhaustiveness they express, however, largely depends on the status of the focus operator they carry. Whereas the preverbal focus expresses contrast within the given proposition, the 'in situ' focus is an instance of cross-sentential operation: the 'in situ' focus contrasts the whole proposition with all other relevant propositions rather than representing an additional 'exclusion' within that proposition. Thus the *in situ* focus is double-faced: it is non-contrastive within the proposition its carrier is a constituent of, and it is contrastive and exhaustive with respect to the relevant set of propositions:

(33) (Mi történt ezután? Kati felhívta a kórházat?   Nem,) [What happened afterwards? Did Kate call the hospital?   No.]

  (a) JÁNOS ment el   a   NAGYSZÜLÖKHÖZ.
      John   went CONV the grandparents-to
      'All that happened was that John went to the grandparents'
  (b) *JÁNOS hívta fel   a   KÓRHÁZAT.
      John   called CONV the hospital-acc
  (c) *KATI ment el   a   NAGYSZÜLÖKHÖZ.
      Kate went CONV the grandparents-to

Accordingly, the relevant set of propositions the multiple focus sentence must be contrasted with should include propositions differing both in their predicates and arguments. A partial difference, such as in (33b) or (33c), is not sufficient. Thus multiple focus expresses full exhaustiveness: it exhaustively identifies the proposition out of a set of possible propositions.

Topic can also function as sentential and cross-sentential topic:

(34) (a) Katit   NEM láttam.
         Kate-acc not   saw-I
         'Kate, I didn't see her'
  (b) Azt,   hogy KATI zárta-e   be   az ajtót,   NEM láttam.
      that-acc that Kate locked-whether CONV the door-acc not   saw-I
      'Whether it was Kate who locked the door, I did not see'

In (34b), *azt* is topic, denoted by its position and prosody. This element is specified by the embedded *hogy KATI zárta-e be az ajtót*, which also functions as topic. Its topic-function is such that the whole proposition it is included in is 'topic-wise' related to other relevant propositions.

The fact that this embedded sentence is a cross-sentential topic allows the given sentence to have its own sentential information structure. Accordingly, in (34b) *Kati* is, in fact, the focus of the embedded sentence.

Prosodically, there appears to be a significant difference between sentential and cross-sentential operations: whereas sentential operations are expressed on the left periphery of the clause, cross-sentential operations appear to be expressed on the right periphery.

5. The interpretation of scope

The scope interpretation of phonologically grammatical sentences directly follows from the given prosodic structure:

(35) (a) The stress-bearing head of each intonational phrase has wide scope over the rest of the operators in the same intonational phrase;

   (b) The relative scope of intonational phrases is determined by the hierarchy of stress reduction; i.e. the higher on the hierarchy a stressed operator of a phonological phrase, the wider its scope over an intonational phrase with its main operator on a lower level in the hierarchy;

   (c) The relative scope of intonational phrases with equal ranking wide scope operators is ambiguous.

Thus, for illustration:

In (4a) *PÉTER evett meg mindent* the focus has wide scope over the universal quantifier: since this sentence consists of one intonational phrase with *Péter* having main stress, the focus has wide scope over the universal quantifier.

In (3a) *PÉTER evett meg MINDENT* there are two intonational phrases (*PÉTER evett meg* and *MINDENT*). It is the hierarchical position of the category of the wide scope operators of each phrase that determines their relative scope. Since *Péter* with its focus function is lower in the hierarchy than the universal quantifier *mindent*, the quantifier has wide scope over the focus.

The reliance of scope assignment on hierarchical stress reduction can also account for the logical equivalence of (3a) *PÉTER evett meg MINDENT* and (3b) *MINDENT Péter evett meg*. Namely, (3b) consists of one intonational phrase, and it is the stressed universal quantifier that must have wide scope over the rest of the sentence, including the non-quantified *Péter* with its focus status.

Scope relations between a universal and an existential quantifier can also be accounted for in a similar fashion:

(36)   x   x   layer 2
        x   x   x   layer 1
        Valamikor mindenki hazament.
        sometime   everyone went-home
        'Everyone went home at some time'

The sentence is ambiguous due to the fact that the quantifiers belong to two differ-
ent intonational phrases and they are on the same hierarchical level.

(37)   x   layer 3
        x   x   layer 2
        x   x   x   layer 1
        Mindenki   hazament   valamikor.
        'Everyone went home at some time'

The second layer derivation (*MINDENKI hazament VALAMIKOR*) is ambiguous
for the same reason as the previous (36). The third level derivation, however (*MIN-
DENKI hazament valamikor*) is unambiguous: with its single main stress it forms
one intonational phrase and, following the scope rule, the main stress bearing uni-
versal quantifier has wide scope over the unstressed existential quantifier.

    Condition (35a), according to which the stress-bearing head of an intonational
phrase has wide scope over the rest of the operators in the same intonational phrase,
seems not to apply in (38a):

(38) (a) PÉTER táncol MINDIG Katival.
         Peter dances always Kate-with
         'It is always the case that it is Peter who dances with Kate'

It might appear that there are two intonational phrases in (38a), namely *PÉTER tán-
col* and *MINDIG Katival* with the prosodic phrasing and—applying condition (35b)
of scope assignment—logical interpretation as shown in (38b). Instead, the proper
logical interpretation of (38a) is (38c), due to the fact that the proper prosodic
phrasing of (38a) is (38d), where the neutral part with its flat prosody is indepen-
dent from the intonational phrase focus$_2$. Its independence is based on its separate
derivation: it is derived by destressing rather than stress reduction, i.e. its stress is
not taken over by the head of the focus$_2$ phrase. Since wide scope is denoted in a
construction with stress reduction, the neutral part is outside the immediate scope
of the preceding intonational phrase.

(38) (b) (PÉTER táncol) (MINDIG Katival)
            'It is always Kate that PETER dances with'
      (c) 'It is always Peter who dances with Kate'
      (d) (PÉTER táncol) (MINDIG) (Katival)
            [focus₁   ] [focus₂   ] [neutral part]

Although topicalization is similar to neutralization in that it does not follow the hierarchy of stress reduction, it differs from the latter in that it forms part of the adjacent intonational phrase. Consequently, condition (35a) of scope assignment applies; cf. (39):

(39)   (Mindenkit NEM látott) (JÁNOS)
            [topic   ] [focus   ] [neutral part]

The obvious relation between prosodic structure and scope interpretation can further be demonstrated in sentences having more than one scope expression in the postverbal field. Szabolcsi's various interpretations of (40) and (41) can be supported by their distinct prosodic structure:

(40)   Egy keddi   napon harapta meg hatnál   több kutya Katit   és   Marit.
        a   Tuesday day-on bit   CONV six-than more dog   Kati-acc and Mari-acc
        'It was on a Tuesday that more than six dogs bit Kati and Mari'
    (i) (a Tuesday >) more than six dogs > Kati and Mari
    (ii) (a Tuesday >) Kati and Mari > more than six dogs

The above interpretations have the following corresponding prosodic structures:

    (i) 1. (Egy KEDDI napon harapta meg) (HATNÁL több kutya) (Katit és Marit)
            [focus₁   ] [focus₂   ] [neutral part   ]
         2. (Egy KEDDI napon harapta meg) (HATNÁL több kutya) (KATIT és MARIT)
            [focus₁   ] [focus₂   ] [focus₃   ]
    (ii) (Egy KEDDI napon harapta meg hatnál több kutya) (KATIT és MARIT)
         [focus₁   ] [focus₂   ]

(41)   Egy keddi   napon harapott meg minden kutya kevés fiút.
        a   Tuesday day-on bit   CONV every dog   few   boys-acc
        'It was on a Tuesday that every dog bit few boys'
    (i) (a Tuesday >) every dog > few boys
    (ii) * (a Tuesday >) few boys > every dog

The following prosodic structures correspond to the above interpretations:

(i) (Egy KEDDI napon harapott meg) (MINDEN kutya) (kevés fiút)
    [focus$_1$] [focus$_2$] [neutral part]

(ii)*(Egy KEDDI napon harapott meg minden kutya) (KEVÉS fiút)
    [focus$_1$] [focus$_2$]

If the latter interpretation were possible, it should have the above (ii) prosodic structure. The derivation of this prosody is, however, blocked for the following reason: in contrast to (i), (ii) consists of only two intonational phrases and the quantifier phrase *minden kutya*, belonging to the first intonational phrase is in the scope of its head, the focus-operator *egy keddi napon*. The phrase *kevés fiút* has the same hierarchical ranking as the head of the first intonational phrase, *egy keddi napon*, and it means that neither of the two operators has wide scope over the other. Their relation is rather that of **coordination**. Coordination would require that the two operators will include the same elements in their scopes. Despite of the fact that the phrase *egy keddi napon* is a non-quantified NP and as such has a hierarchical ranking lower than the quantifier, it does include the quantifier in its scope, since, in sentence intitial position, it is the carrier of a sentential operator. Since scope is marked by stress reduction, the wide scope of *kevés fiút* over the quantified phrase *minden kutya* cannot be denoted since the would-be wide scope operator, *kevés fiút* has lower ranking (it cannot be the carrier of a sentential operator either).

Similarly, (42a–c) also demonstrate that in the sequence C_1 and C_2, C_2 can only have scope over C_1 if (a) C_2 is higher on the hierarchy or (b) (having equal ranking) C_1 does not include a narrow scope operator whose hierarchical ranking is higher than that of the head of C_2:

(42) (a) (JÁNOS beszél) (KATIVAL)

    (b) (JÁNOS beszél) (MINDIG) (Katival)
        [focus$_1$] [focus$_2$] [neutr. p.]

    (c)*(JÁNOS beszél mindig) (KATIVAL)
        [focus$_1$] [focus$_2$]
        János   talks   always   Kati-with

6. Conclusions

In this paper we outlined the basic assumptions of a metrical syntax of Hungarian. We assumed that, from the point of view of prosody, the grammaticality of a sentence depends on whether a proper prosodic structure can be assigned to it. The

prosodic rules are based on rules of stress reduction between adjacent prosodic components. Stress reduction is constrained by a hierarchy of categories available for reduction.

The paper also addressed the question of how and to what extent syntax, prosody, communicative and logical functions are related. It was found that there is a significant relation between the basic syntactic structure (as described in É. Kiss's model) and prosodic structure of Hungarian sentences: the ordering of the universal quantifier, the focus-element and the verb in the syntactic structure also corresponds to their hierarchy determining stress reduction. It was also found that the study of the relation of the prosody of sentences and their communicative functions enables us to differentiate three, prosodically and communicatively distinct parts: the topic field, the focus field and the postverbal neutral part. From the point of view of prosody and communicative functions, both the (universally) quantified NPs and the non-quantified focused NPs appear to constitute material within the same focus field. Relying on the prosodic features of the communicatively distinct parts of the sentence the given communicative structure also enables us to relate communicative structure to logical interpretation: the intonational phrases which also determine the communicative structure of a sentence are the basis for the determination of relative scope interpretations.

References

Chomsky, N. – Halle, M. 1968. The sound pattern of English. Harper and Row, New York.

É. Kiss, K. 1978. A magyar mondatok egy szintaktikai modellje [A syntactic model for Hungarian sentences]. In: Nyelvtudományi Közlemények 80: 261–86.

É. Kiss, K. 1987. Configurationality in Hungarian. Akadémiai Kiadó, Budapest.

É. Kiss, K. 1994. Sentence structure and word order. In: Kiefer, F. – É. Kiss, K. (eds): The syntactic structure of Hungarian. Syntax and semantics 27: 1–90. Academic Press, San Diego/New York.

É. Kiss, K. 1998a. Identificational focus vs. information focus. In: Language 74: 245–73.

É. Kiss, K. 1998b. Multiple topic, one focus? In: Acta Linguistica Hungarica 45: 3–29.

Halle, M. – Vergnaud, J-R. 1987. An essay on stress. The MIT Press, Cambridge MA.

Hayes, B. 1995. Metrical stress theory. The University of Chicago Press, Chicago.

Hunyadi, L. 1981a. A nyelvi polaritás kifejezése a magyarban [The expression of linguistic polarity in Hungarian]. PhD dissertation, ms, Debrecen.

Hunyadi, L. 1981b. Remarks on the syntax and semantics of topic and focus in Hungarian. In: Acta Linguistica Hungarica 31: 107–36.

Hunyadi, L. 1996. Hungarian syntactic structure and metrical prosody. In: Language Sciences 18: 139–52.

Kornai, A. – Kálmán, L. 1988. Hungarian sentence intonation. In: Hulst, H.v.d. – Smith, N. (eds): Autosegmental studies on pitch accent, 183–95. Foris, Dordrecht.

Liberman, M. – Prince, A. 1977. On stress and linguistics rhythm. In: Linguistic Inquiry 8: 249–336.

Selkirk, E.O. 1980. The role of prosodic categories in English word stress. In: Linguistic Inquiry 11: 563 605.

Selkirk, E.O. 1984. Phonology and syntax: The relation between sound and structure. The MIT Press, Cambridge MA.

Szabolcsi, A. 1995. Strategies for scope taking. HAS Linguistics Institute, Budapest.

Varga, L. 1996. Hungarian intonation contours. In: The Even Yearbook 2. ELTE SEAS Working Papers in Linguistics, 111 44. ELTE, Budapest.

Vogel, I. Kenesei, I. 1987. The interface between phonology and other components of grammar: The case of Hungarian. In: Phonology 4: 243 -63.

Vogel, I. Kenesei, I. 1990. Syntax and semantics in phonology. In: Inkelas, S. Zec, D. (eds): The phonology syntax connection, 339 63. The University of Chicago Press, Chicago.

Address of the author: László Hunyadi
Department of General and Applied Linguistics
Kossuth Lajos University
H-4010 Debrecen, Hungary
e-mail: hunyadi@llab2.arts.klte.hu

WEAK SUBJECTS IN FIXED SPACE

MÁRTA MALECZKI

Abstract

The paper investigates how sentential predicates influence whether an indefinite (weak) subject has weak or strong interpretation. It is argued that these interpretational possibilities are determined by the specifying or non-specifying character of the predicate, which in turn depends on other predicate properties. The results of examining telic/atelic, bounded/unbounded, stage-level/individual-level distinctions is that it is telicity and locatedness that can make a predicate specifying. The main claim is that the whole story is the result of a general specifying criterion. English, Hungarian and French data are considered during the argumentation.

Introduction[1]

It is well known from Milsark's frequently cited dissertation (1974) that noun phrases can be divided into two groups according to whether they occur freely in existential constructions (*there*-sentences) or not. The former group is called weak whereas the latter strong. Barwise and Cooper (1981) explicated the difference between the two kinds of noun phrases on the basis of their formal properties. By their definition, a noun phrase (a generalized quantifier, denoting a family of sets) is weak iff it is not strong. An NP is positively strong iff the set the common noun part of the NP denotes always belongs to the family of sets the whole NP denotes, and negatively strong iff the denotation of the common noun can never belong to the NP-denotation. That is, a noun phrase is weak iff the relation between the denotation of the common noun and the whole NP is affected by the contingent properties of the given model. Thus noun phrases with indefinite determiners (*a, some, few, many,* etc.) are considered weak, both by the test Milsark proposed and by the definition given in Barwise–Cooper (1981).

As de Hoop (1995) has argued recently, weakness and strength as character-ized above are based exclusively upon the semantic properties of determiners, con-trary to the possible weak or strong readings of NPs, which depend on the syntac-tic context. Accepting the first part of this claim I argue that it is not only the syn-tactic context which is relevant to the emergence of weak or strong interpretations.

Phenomena related to the weak or strong interpretation of noun phrases can be partly identified with the problem set connected with the specific/non-specific dis-tinction.[2] Specificity has been investigated from several points of view. Different approaches to specificity show that its sources are various in nature. From these one major issue is picked out in the present paper: effects originating from the senten-tial predicate.[3] I argue that the aspectual and some other properties of sentences are a factor of utmost importance in determining whether a (subject) noun phrase can or must have weak or strong (non-specific or specific) reading. Consider (1)–(3):

(1)   A student walks willingly.
(2)   A student is walking in the garden.
(3)   A student has already walked (in the garden).

The most natural interpretation for (1) is the generic (strong) reading (it is charac-teristic of students that they walk willingly); for (2) the non-specific (weak) read-ing (there is a (non-identified) student walking in the garden); and for (3) the spe-cific (strong) reading (one element of a previously introduced set of students has already walked). Our concern in (1)–(3) is that whereas the indefinite subjects have different readings, the only overt opposition is between the aspectual properties of the (same) verb phrase (and the concomitant adverbs corresponding with the dif-ferent aspects). Being so, aspectual features of verbal predicates surely can influ-ence whether the subject noun phrase gets a weak or strong interpretation. It is unclear, however, which are the relevant properties in this respect and how they are to be defined in order to show their impact on the possible readings of the subjects. This is the topic of the present article.

The paper is organized in the following way:

The first section presents what will be understood by weak and strong inter-pretations of indefinite noun phrases.

---

[2] As de Hoop (1995) notes in footnote 10, the difference is that the notion of strong reading also covers generic and partitive readings beside the specific ones. Notice, however, that the specificity definition given in Enç (1991) clearly extends to partitive readings as well.

[3] As will be emphasized below, the term sentential predicate cannot be equated to the verb alone. See section 1.1 for more details.

In the next section some predicate properties are examined. Section **2.1** deals with the stage-level/individual level and generic/non-generic distinctions, and comes to the conclusion that "timeless" sentences develop a strong interpretation of their indefinite subjects. Section **2.2** is an intermezzo about thetic and categorical modes of judgements, which will contribute to restricting the set of data to be considered in such a way that predicate properties be kept separated from effects originating from syntax as clearly as it is possible. Section **2.3** presents the specifying/non-specifying distinction as developed to account for the present problem by Bosveld-de Smet (1993). Section **2.4** examines whether telicity and boundedness can influence the interpretational possibilities of indefinite subjects.

In section **3** it is shown that in Hungarian the specifying property can be derived from the telic or place-bounded character of the predicate. The final section argues that there is a general specifying criterion on statements that can be made responsible for the observed linguistic facts.

1. Weak noun phrases with weak or strong reading

While I admit that strong (definite) NPs (as defined by Barwise–Cooper 1981) can both have strong (specific) or weak (non-specific) readings in some sense (see Groenendijk–Stokhof 1980; de Hoop 1995), indefinite NPs exhibit these different interpretational possibilities in a far more striking way. Notice that if in (1)–(3) we replace the indefinite subjects with definite ones, the observed differences in their meaning disappear.

(1') The student/every student walks willingly.
(2') The student/every student is walking in the garden.
(3') The student/every student has already walked.

On the other hand, if we replace the indefinite article *a* in (1)–(3) with some other weak determiner, weak or strong interpretational possibilities show up again:

(1") Some students/many students walk willingly.
(2") Some students/many students are walking in the garden.
(3") Some students/many students have already walked.

Since definite NPs do not exhibit the weak/strong opposition as clearly as the indefinite ones, they will be ignored in the present paper.

In examples (1")–(3"), due to the individual meaning of the determiners, the difference is not quite the same as in (1)–(3). However, strong and weak readings are present here exactly as expected: indefinite noun phrases in (1") and (3") are strong in the sense that the sentential predicates are claimed of a subset of students (that is, the implicatures arise that there are some students who do not walk willingly/have not walked yet); whereas (2") is not necessarily interpreted in that way. (1") and (3") have a proportional or partitive reading, whereas a prominent reading of (2") is weak, that is, existential or non-specific.

In order to ascertain whether a noun phrase can be interpreted weakly I use the definition of specificity given in Enç (1991): the reading of indefinite NPs is considered weak iff they introduce discourse-referents that are new in the sense that they are unrelated to previously established discourse referents. Another well-known test can be attached to this notion of weakness: the entities the weak NP refers to may possibly constitute the whole denotation of the common noun part of the NP. That is, *some students* in (2") does not necessarily presuppose any previously mentioned set of students, and the sentence can be true in a model where there are no other students than those walking in the garden.

Following de Hoop (1992; 1995), a non-contrastively stressed NP is considered here strong in three cases (cf. footnote 2). First, it is strong if it is specific in the sense that it does not have the discourse referent introducing potential, but refers to a known (or so assumed) discourse referent. Second, it is strong if it is proportional or partitive, that is, the sentence containing it cannot be true if there are no other individuals belonging to the extension of the common noun beside the ones the NP refers to. Third, the generic reading of an indefinite noun phrase is also considered strong.

Whether an indefinite noun phrase can or must have a strong or a weak reading depends on several, seemingly very different factors. Ladusaw (1994) argues that there are two basic types of judgements, thetic and categorical, and thetic judgements allow only weak NPs (in both senses of weakness). De Hoop (1992) argues for a syntactic solution: in her opinion there are two types of Case-assignment according to different syntactic configurations, and we get weak readings under the weak structural (or inherent) case assignment, whereas strong readings arise when the NP gets its case via strong structural case assignment. Others stress the contextual, pragmatic factors (e.g. Groenendijk–Stokhof 1980; Ludlow–Neale 1991). Some authors notice the important role the meaning of the main predicate can play in getting weak or strong readings (Bosveld-de Smet 1993; de Hoop 1995; É. Kiss 1994).

The topic of the present paper is to examine how the semantic properties of sentential predicates can influence whether an indefinite subject NP can or must

have weak or strong reading. Although I am fully aware of the fact that other factors mentioned above (syntactic position, context, etc.) can also be significant, they will be put aside for the present.

What is referred to by the expression "semantic properties of sentential predicates" is quite a complex issue, and is determined not only by the lexical properties of the verb alone. Features connected with tense and aspect will be especially important for our present concern, and these can be treated more appropriately as sentential properties. Thus what is called here simply a predicate property, is influenced by the lexical meaning of the verb, its aspectual form, its tense, its complements, and also by some kinds of adjuncts. With these caveats in mind, firstly some properties that have recently been used for classification of verbal predicates will be examined.

2. Classifications of predicates

2.1. Stage-level vs. individual-level predicates

Since Carlson's famous dissertation (1977) much attention has been paid to the stage-level and individual-level properties. The basis of this differentiation is the observation that predicates behave differently according to whether they express permanent features or accidental properties. In order to explain that non-uniform behaviour, Carlson distinguished two different possible levels of entities in the model: predicates expressing permanent qualities apply to individual level arguments, whereas those expressing transient properties take some spatiotemporal realization of individuals (stage-level arguments).

Kratzer (1995) argues that stage-level/individual-level predicate properties can be derived better from a systematic difference in the argument-structure of predicates: stage-level predicates have an extra (event) argument slot for spatiotemporal specification. Individual-level predicates normally express properties that are independent of place and time, so they cannot be modified by place or time adverbials, contrary to stage-level predicates. Kratzer (1995) illustrates this with the following examples:

(4)   Manon is dancing on the lawn.
(5)   Manon is dancing this morning.
(6)   Manon is a dancer.

The stage-level predicate *is dancing* can easily be modified by place or time adverbials, whereas the individual-level predicate *is a dancer* "cannot be modified by locatives. If it can, it has turned into a stage-level predicate" (Kratzer 1995, 128).

The inverse of this reasoning also holds: if a sentence is not located either in time or in space, it expresses a permanent or habitual property of the subject, even if the predicate is not inherently individual level (generic sentences). As Dahl (1995) argues, genericity is cross-linguistically expressed with the least marked form of the verb with respect to the tense/aspect. This verb form is the Simple Present in English; it may indicate that the spatiotemporal argument slot is not filled (for the present argumentation the reportive and futurate uses are irrelevant).[4] Thus in order to get a non-generic interpretation, it may be enough to use any tensed or aspectually marked form of a predicate that is not characteristically individual-level. This can indicate by itself that the predicate has the extra argument slot making it stage level.

Consider again our examples (1)–(3):

(1)   A student walks willingly.
(2)   A student is walking in the garden.
(3)   A student has already walked (in the garden).

The only overt difference between (1)–(3) is in their time/aspect. Now it is clear that the "timeless" (1) is a generic (more exactly, a habitual) sentence; it is about students as a kind. The existence of that kind ("student-kind") is not claimed but presupposed. Since the indefinite NP does not introduce new referents into the discourse, the indefinite subject is strong.

In fact, concerning their timeless character, there is no difference between individual level predicates and the class of stative verbs. As Bach (1981) argues, the most distinctive feature of statives as opposed to dynamic (process or event) predicates is that they are timeless in the sense that establishing the truth-conditions for them does not necessarily involve more than one moment of time. Thus it is not surprising that statives that are neither inherently stage-level nor made specific in some way or other,[5] trigger the strong reading of their indefinite subjects.

---

[4] It is well known that there are some predicates which cannot be interpreted for stages; e.g. *extinct* (in the sense 'having died out'), *widespread*. Of course these always will be interpreted as individual-level predicates, irrespective of the time/aspect of the sentence. On the other hand, there are inherently stage-level predicates as well (*drunk, awake* (adj)). The point is that if a predicate is neither inherently individual nor stage-level, then it is the Present Tense that can trigger a generic or habitual reading most easily.

[5] Objects, for instance, can influence the interpretation of the subject as well: cf. (i) *A girl knew poems* with (ii) *A girl knew that poem*. While sentence (i) with the non-specific bare plural object can be interpreted generically ('a girl (in the old times) used to know poems'), the demonstrative pronoun makes the object specific, so in (ii) the predicate becomes stage-level. Thus the subject in (ii) cannot be (easily) interpreted as referring to an individual-level entity.

The next problem is that although both (2) and (3) are non-generic sentences containing stage-level non-stative predicates, still there is a difference in the strength of their indefinite subjects. That puzzle will be examined in the following sections.

2.2. An intermezzo: thetic and categorical modes of judgements

As mentioned in section 1, Ladusaw (1994) argues that the weakness or strength of the subject is influenced by the judgement type it occurs in. Thetic statements, as opposed to categorical ones, are predications as a whole, with no prominent or pre-supposed arguments. They do not express judgements about some well-delineated subject, but they are **presentations** of an object (an individual or an eventuality).[6] Since the indefinite subjects of thetic statements themselves are presentations (descriptions), they are non-specific. Typical instances of this judgement-type are the existential constructions (*there*-sentences): they express a presentation of some individuals by their very nature, and this explains why definite or specific noun phrases are normally excluded from these sentences.

Categorical judgements, on the other hand, have a clearly articulated subject-predicate structure: they are statements about a presupposed subject. Typical examples of this mode of judgement are generic sentences: these attribute properties to subjects as characteristic of them, independently of their spatiotemporal location. A generic property can be predicated either of an individual or of a group of individuals, but not of presentations (descriptions). Thus subjects of generic sentences have to be strong.

Between the two extremes, there are the sentences with indefinite subjects that ought to be ambiguous in principle: we should be able to interpret them as both thetic and categorical judgements. In a range of well-known examples that is the case, indeed (Milsark 1974, 199):

(7) Some unicorns entered the garden.
(8) Many people were at the party.

The subjects in (7), (8) can have both weak (existential) and strong (proportional) readings. However, as (2)–(3) show, there are factors favouring the weak or the strong reading, also in sentences which are neither existential constructions nor generic statements. In the following sections our attention will be directed to sentences whose form is neither utterly thetic (existential) nor utterly categorical (generic).

[6] Ladusaw equates the notion of a presentation of an object for a description of an object, where "a description is something which can be satisfied by an object", and objects range over both individuals and eventualities (Ladusaw 1994, 223 4).

2.3. The specifying/non-specifying distinction

Relative clauses in French, like in English, have two basic types: "relatives spéci-fiantes" (defining or restrictive relatives) and "relatives non spécifiantes" (non-defin-ing or non-restrictive relatives). Examining these, Kleiber (1981) observes that the propositions the former type of relatives expresses cannot have a generic or habitual reading, while the latter type can be interpreted as attributing an "inner property" to the noun phrase the relative clause modifies. Kleiber argues that the occurrence of these interpretational possibilities depends solely on the properties of the predicate the clause contains. Relative pronouns in French play no role at all in this respect; this fact is illustrated by minimal pairs with specifying/non-specifying relatives that differ only in the predicates of the relative clauses (Kleiber 1981, 216).

From these observations Kleiber concludes that predicates themselves are specifying or not. The distinctive property of these predicate classes is that speci-fying predicates, as opposed to non-specifying ones, are somehow anchored in time and space, and that is why their subjects are to be interpreted non-generically. Kleiber observes that in order to be able to anchor individuals, specifying predi-cates have to express either some event (verbes d'action), or have to contain some explicitly expressed location. Non-specifying predicates, on the other hand, result in ambiguity: the clauses they occur in can be interpreted as specifying or non-spec-ifying relatives.

Bosveld-de Smet (1993) extends Kleiber's analysis, and examines how the specifying or non-specifying property of a predicate influences whether its indefi-nite subject can be interpreted weakly or strongly. She argues that specifying pred-icates allow or favour the weak reading of their subjects, while non-specifying ones require strong subjects. Trying to find an explanation, she combines Kleiber's observations with Kratzer's distinction between individual-level and stage-level predicates, and defines a predicate as specifying iff there is some information avail-able that fills the spatiotemporal argument slot of the (stage-level) predicate. This distinction does not result in the same classes as the individual-level/stage-level opposition, since a predicate is non-specifying irrespective of whether it does not have a spatiotemporal location slot altogether (individual level predicates), or it has one but it is unfilled (non-specifying stage-level predicates). Of course, individual-level predicates are non-specifying by default, but stage-level predicates can be either specifying or not. The non-specifying versus specifying distinction manifests itself in a well-formedness criterion put on the use of French plural indefinite arti-cle *des*. This determiner can only be interpreted weakly, and it is good with speci-fying predicates (*sont venus* in (9)), but it is unacceptable with non-specifying pred-icates (*sont remplis* in (10), see Bosveld-de Smet 1993, 32–4):

(9) Des amis sont venus.
    '(Sm) friends have come'

(10) *Des verres sont remplis.
     '(Sm) glasses are full'

Notice that in (9), (10) both predicates are stage-level. I will assume in what follows that the specifying/non-specifying distinction as Bosveld-de Smet defines it is the relevant predicate property to allowing or prohibiting weak or strong readings of indefinite subjects. Then the next question is what kinds of linguistic means can indicate that the spatiotemporal slot of a predicate is filled or not; that is, what can make a predicate specifying (in a construction which is not typically existential). Different interpretations of the subjects of (2) and (3) show that aspectual properties of predicates can be candidates.

2.4. Telicity and boundedness

The Vendler-classification and the aspectual properties of predicates are usually considered at least partially overlapping. (Vendler-classes are sometimes called aspectual classes.) The need for defining and separating the underlying properties these distinctions are based upon has emerged from time to time. Although it is a matter of debate what features are the distinctive ones in defining the Vendler-classes, it is widely accepted that the telic/atelic opposition is relevant. Recently Depraetere (1995) has argued that a clear distinction has to be made between the atelic/telic character of a situation on the one hand, and its boundedness on the other. The usual aspectual distinctions (e.g. perfective/imperfective) can be then given with these more basic notions. Depraetere claims that it is the former distinction which is relevant to the Vendler-classification, and it is based on whether the situation described contains an inherent endpoint or not. Specifically, a sentence is telic iff some intended or inherent (natural) endpoint is reached in the described situation when the event is completed, independently of whether the endpoint is actually reached in the given context or not. That is, telicity is given inherently in the meaning of the (possibly complex) predicate; (11), (12) are telic, while (13), (14) are atelic (examples (11)–(14) are selected from those of Depraetere (1995), with different numbering):

(11) Sheila collapsed.
(12) John was opening the parcel.
(13) Sheila is working in the garden.
(14) Julian lived in Paris from February 1989 until May 1989.

The bounded/unbounded property, on the other hand, expresses some limitation in time. This means that a telic predicate becomes bounded when the inherent end-point is actually reached (see (11)); whereas an atelic predicate is bounded when some limitation in time is given (usually by an adverbial phrase, see (14)). As can be seen from the above examples, (un)boundedness cross-classifies (a)telicity: both atelic and telic sentences can be either bounded or unbounded. (11) describes a telic bounded, (12) a telic unbounded, (13) an atelic unbounded, and (14) an atelic bounded situation.

Now let us return to our examples (1)–(3):

(1)   A student walks willingly.
(2)   A student is walking in the garden.
(3)   A student has already walked (in the garden).

These all are atelic sentences, but (1) and (2) are unbounded whereas (3) is bounded. Putting aside the generic sentence (1), comparing (2) and (3) may lead us to the tentative generalization that boundedness favours the strong reading of the subject, while unboundedness develops the weak reading. Let us put this assumption to the test and replace the definite subjects of (11)–(14) with indefinite ones:

(15)   A woman collapsed.
(16)   A boy was opening the parcel.
(17)   A man is working in the garden.
(18)   A student lived in Paris from February 1989 until May 1989.

(15) is a telic bounded sentence, and its subject can be easily interpreted as weak. The prominent reading of the subjects of the telic unbounded (16) and atelic unbounded (17) is also weak. The atelic bounded (18), on the other hand, seems to allude to a specific student, or else it is hopelessly uninformative.

Thus it seems that telic and atelic verbs behave differently in the relevant respect. Boundedness alone does not favour either reading (see the bounded (15) and (18) with subjects different in strength), but when the verb is atelic, boundedness seems to trigger the strong reading (see (3) with an activity verb and (18) with a stative verb). Let us see some more examples, changing now the tense of the verb and the type of boundedness:

(19)   A woman has collapsed.
(20)   A boy is opening the parcel.
(21)   A man was working in the garden.

(22) A man was working in the garden from eight to ten.

(23) A man has already worked in the garden.

(24) A student has already lived in Paris.

The reader can check with (19)–(24) and similar examples that telicity in itself guarantees that the subject may have a weak reading (see (19), (20)). Boundedness plays a role in this respect only if the verb is atelic, but not each kind of boundedness: (22) is bounded, and the weak reading of the subject is available without any difficulty; but the Present Perfect makes the weak reading much less accessible.

In sum, our first guess seems incorrect: (un)boundedness in general does not influence the possible readings of the subjects. Unboundedness alone could not even emerge as a candidate in determining whether a predicate is specifying or not: it is neither necessary nor sufficient for triggering either the weak or the strong interpretation of an indefinite subject (see unbounded statives with strong subjects, unbounded activities with strong (in the Present Tense) or weak (in the Continuous) subjects, and bounded telic sentences with both weak and strong subject interpretations). Boundedness alone does not seem to develop the strong (or weak) reading of the subject, either; but it is remarkable that the perfective aspect in atelic sentences favours the strong reading of the subject.

Telicity, on the other hand, seems to be definitive: if a predicate is telic, the weak subject interpretation is always available (irrespective of whether the sentence is bounded or not, perfective or not). That is, telic predicates are specifying (recall that a predicate is specifying if its subject can have weak interpretation, see **2.3**), but atelic predicates behave non-uniformly: activities in the Present Perfect and stative predicates tend to be non-specifying, while non-perfective and non-generic activities seem to be specifying.

Now if we try to match these observations with Bosveld-de Smet's theory, it seems that telicity involves or indicates that the spatiotemporal slot of the predicate is filled (that is, telic predicates are spatiotemporally located events). The question arises, why and how telicity makes a predicate specifying. A tentative answer might be given along the following line of reasoning. Per definitionem, a telic predicate inherently alludes to the endpoint of the eventuality it describes. The notion of endpoint involves necessarily some change (independently of whether the change actually happened or not). The notion of change in turn is based on comparing states of affairs in two different temporal intervals. But the reverse holds as well: without any change in the states of affairs, time cannot be measured, defined, perceived—that is, with no alteration in the world the notion of time would lose its sense and content. From the mutual dependency of time and change it follows that in telic predicates, where an endpoint (that is, some change) is inherently present in the

meaning of the predicate, reference to time is involved as well (or else no change whatsoever can be alluded to, that is, the predicate cannot be telic). Thus the time-dependency of the eventuality described is inherently present in telic predicates.

Statives are not located, thus they are non-specifying, and so are activities in the Simple Present. So far these results are not surprising. What seems to be a really interesting question is what can anchor activities spatiotemporally, and why they behave like non-located eventualities in the Perfect.

3. Hungarian

We can arrive at interesting generalizations if we consider a language having a less clearly articulated aspectual system than English. The next sections examine the emergence of weak and strong subject interpretations in Hungarian.

3.1. Thetic and categorical judgements in Hungarian

In Hungarian the specificity of the grammatical subject heavily depends on the sentence type it occurs in, thus some remarks on Hungarian sentence structure are necessary here.

Hungarian has two basic sentence-types: neutral and non-neutral sentences. The latter type contains a constituent in a fixed position (immediately before the verb) with a special stress and focus interpretation. That sentence-type will be ignored in the present paper; attention will be directed to neutral sentences alone.

Kálmán (1985) claims that (putting aside the special imperfective and identifying sentences) there are two basic types of neutral sentences in Hungarian: the one contains some constituent before the verbal part (V'), the other does not. The logical subject (the topic) of a sentence is necessarily before the V'; thus sentences with no constituent in the pre-verbal part cannot have the Aristotelian subject–predicate structure, they "tend to express simple events rather than n-ary relations" (Kálmán 1985, 15; this observation is attributed to A. Tich). This semantic difference between Hungarian verbal and non-verbal neutral sentences is essentially the same as the thetic/categorical distinction (see **2.2**), so we can risk the generalization that topicless Hungarian neutral sentences are used to express thetic judgements. As we have seen in section **2.2**, grammatical subjects of thetic sentences are descriptive expressions semantically, or as Kálmán expresses the same idea for Hungarian, constituents behind the verb or before it but under the V' are not arguments but verbal modifiers. The thetic judgement type is unproblematic from our present point of view, since the indefinite NP subject in these sentences cannot be strong normally. This can be nicely exemplified with Hungarian data as well: stative predi-

cates, which require their subject to be strong, do not tolerate the verbal sentence form. (25) is not a well-formed Hungarian sentence, as opposed to (the normative generic) (26):

(25) *Utál   egy   béka   minden   gólyát.
      hates   a   frog   every   stork-acc

(26) Egy   béka   utál   minden   gólyát.
      a   frog   hates   every   stork-acc
     'A frog hates every stork'

In sum, non-verbal neutral sentences (more exactly, neutral sentences with the subject in the topic) will be examined in the following sections. This sentence-type has a so-called level prosody: each phrasal constituent bears a uniformly slight stress. In some examples the stress will be indicated with a ' mark before the stressed word, but only if it seems necessary for the sake of clarity.

3.2. Telicity and boundedness in Hungarian

In this section the relevant predicate properties and their combinations will be examined in a systematic way with Hungarian data. Examples will be given in the Past Tense whenever it is possible, in order to avoid generic readings, which are expressed with the (only) Present Tense, and might blur the picture.

3.2.1. Atelic predicates

As we have seen in section **2.4**, a predicate is atelic iff there is not any endpoint reached when the eventuality the predicate expresses is carried out. Thus stative and activity verbs will equally belong to this class. If there is no adverb or other explicit marker in the sentence making it bounded, atelic predicates by themselves are always unbounded. (Recall that boundedness means some limitation in time.)

On the basis of Hungarian data it seems that **unbounded** atelic predicates allow only strong subjects in non-verbal sentences. That is, their subject can be either a definite, or—less acceptably—a specific (contrastive) indefinite noun phrase. This is exemplified by (27)–(29):

(27) Hugó/a kutya/minden kutya aludt/futott/bűzlött.
     Hugh/the dog/every   dog   slept/ran /stank
     'Hugh/the dog/every dog was sleeping/running/stinking'

(28) ?Egy 'kutya 'aludt/'futott/'büzlött.
    a   dog   slept ran   stank
    'A dog was sleeping/running/stinking'

(29) 'Néhány 'kutya 'aludt/'futott/'büzlött.
    some   dog   slept ran   stank
    'Some dogs were sleeping/running/stinking'

Examples in (27) are wholly acceptable, unambiguous sentences. (28) is unaccept-
able with the indicated level-prosody; but it becomes acceptable if we lay empha-
sis on the determiner instead of the common noun. However, this contrastive stress
makes the indefinite subject specific in the sense of Enç (1991): a previously intro-
duced set of dogs is presupposed, and the indefinite NP refers to one of them. That
is, sentences in (28) must have some marked intonation pattern to express the
strength of their subject, indicating that a context with some opposition is given,
like e.g. in (30), (31)):

(30) 'Egy kutya 'büzlött, a 'többinek 'nem volt szaga.
    'One dog was stinking, but the others did not smell'

(31) 'Egy kutya 'aludt, egy 'másik 'játszott, egy 'harmadik 'csontot rágcsált.
    'One dog was sleeping, another was playing, a third one was chewing a bone'

The preferred versions of sentences in (28) with specific (but not necessarily con-
trastive) indefinite subjects would be the wholly acceptable (32)–(34), which do not
trigger the contextual requirements mentioned above, and they are good with a non-
contrastive neutral intonation as well:

(32) Egy 'kutya 'megbüdösödött.
    a   dog   pfx-stank
    'A dog became stinky'

(33) Egy 'kutya 'elaludt.
    a   dog   pfx-slept
    'A dog fell asleep'

(34) Egy 'kutya 'elfutott.
    a   dog   pfx-ran
    'A dog ran away'

That is, when the subject is specific, some prefixed version of the verb is preferred. The prefixes in (32)–(34) make the originally atelic verbs telic, and that is the only difference between sentences in (28) and their prefixed counterparts.[7]

Almost the same can be said about (29): the indefinite NP has to have a partitive reading. But due to the meaning of the determiner *néhány* 'some', the partitive reading can easily arise, so the contextual and intonational requirements are not so strong here as in the case of (28). The prefixed, telic counterparts of the atelic verbs fit here as well.

On the basis of these observations, it seems that atelic verbs (statives and activities) by themselves are non-specifying in Hungarian: their subject has to be strong, weak reading cannot arise. However, there might remain some doubt whether telicity or boundedness is relevant here: since our examples are in the Past Tense, telic examples are bounded as well. That telicity is the critical factor can be supported with atelic bounded examples:[8]

(35) Egy / 'néhány 'kutya 'tegnap   estig   'büzlött / 'aludt.
     a   / some   dog   yesterday evening-till stank   / slept
     'A/some dog(s) was/were stinking/sleeping until yesterday evening'

(36) Egy / 'néhány 'fiú 'reggelig   'sétált.
     a   / some   boy morning-till walked
     'A/some boy(s) was/were walking until morning'

In (35) and (36) the utterly specific reading of the subject shows that it does not matter whether atelic verbs are bounded or not.

Of course there are atelic Hungarian sentences with weak subject interpretation; but what makes that interpretation available without any doubt is not boundedness in time but in space. That is, if we localize the atelic non-stative[9] sentences

---

[7] This does not hold for Hungarian prefixes in general; lexically telic verbs (e.g. achievements) also can have prefixes. Moreover, prefixes do not have a fixed meaning that can be attached compositionally to each verb.

[8] Of course it could be tested by telic unbounded examples as well. But telic unbounded sentences have to be either in the Present Tense, or in the Future (in Hungarian). This would be disturbing for several reasons. As mentioned above, Present Tense is used to express genericity as well, thus we could not set minimal pairs as clearly as in the Past. Moreover, Present Tense does not express a real present in the case of telic verbs: it refers rather to some future eventuality. The future also has its own problems: it has either the flavour of some prediction, or it is a promise, and both interpretations might affect the specificity of the subject.

[9] Stative verbs, because of their very nature, cannot be easily localized. See section **2.2** for more details.

in (28) and (29) with some place adverbial, then weak readings are allowed in all the examples above:

(37) Egy kutya aludt/futott itt/ a   kertben.
      a   dog   slept/ran   here/the garden-in
      'A dog was sleeping/running here/in the garden'

(38) Néhány kutya aludt/futott itt   /a   kertben.
      some   dog   slept/ran   here/ the garden-in
      'Some dogs were sleeping/running here/in the garden'

(39) Egy kutya tegnap estig aludt/futott itt/a kertben.
      'A dog was sleeping/running here/in the garden until yesterday evening'

Thus we might conclude that on the basis of Hungarian data there is a kind of boundedness that has to be distinguished from the boundedness defined with respect to time in Depraetere (1995). Locatedness in space is of great importance to allowing the weak interpretation of the subject; it definitely makes a difference in Hungarian whether the Kratzerian spatiotemporal argument slot is filled with information about place or time. In the former case I will speak about place-boundedness, separating it from time-boundedness.
     The conclusion based on the notion of place-boundedness is formulated in (40):

(40) In Hungarian atelic predicates (in non-verbal sentences) can be specifying (allow weak subjects) only if they are place-bounded.

Recall that specifying predicates do not require their subject to be weak but allow weakly interpreted subjects. In accordance with this, (40) does not involve that subjects of place-bounded predicates cannot have strong interpretation.[10]

3.2.2. Telic predicates
With telic predicates both strong and weak readings of the subject seem possible, and place adverbials do not favour either interpretation.

---

     [10] It is interesting that the preference of strong or weak interpretation depends on where the place adverbial occurs: if it is placed BEFORE the verb, strong interpretation more easily arises. This is in a remarkable correlation with Dutch, where scrambling causes strength (see de Hoop 1992; 1995).

Telic, unbounded predicates:

(41) Néhány remete egy csónakot fest (a folyóparton).
some hermit a boat-acc paints the river-bank-on
'Some hermits are painting a boat (on the river-bank)'

(42) Egy kislány észreveszi majd ezt a falfirkát (az utcán).
a little-girl notices later this the graffiti the street-on
'A little girl will notice this graffiti (in the street)'

(41) contains an accomplishment, (42) an achievement predicate. In both sentences, subjects can have both weak and strong interpretations.

Telic, bounded predicates:

A telic situation becomes bounded when the endpoint inherently involved is actually reached. As we have seen in section **2.4**, boundedness can be indicated by using a perfect form, and this turns the predicate non-specifying. In Hungarian there are no separate inflectional forms to express the different aspectual properties, but there are verbal prefixes which can be used to express perfectivity.

However, prefixed telic predicates do not behave uniformly with respect to their specifying property. In general, prefixes do not seem to affect the interpretational possibilities of indefinite subjects: in (41') and (42') the subjects continue to have both weak and strong readings:

(41') Néhány remete befestett egy csónakot (a folyóparton).
some hermit pfx-painted a boat-acc the river-bank-on
'Some hermits painted a boat (on the river-bank)'

(42') Egy kislány észrevette ezt a falfirkát (az utcán).
a little-girl noticed this the graffiti-acc the street-on
'A little girl noticed this graffiti (in the street)'

Here are some more examples supporting this observation:

(43) Egy fiú elment a városba.
a boy away-went the city-in
'A boy has gone to the city'

(44) Néhány vándor felkerekedett.
some wanderer up-arose
'Some wanderers set off'

Thus the generalization about telicity based on English data seems to hold in Hungarian as well:

(45) Telicity puts no restriction on the interpretation of subjects: telic predicates can give rise to weak and strong readings alike.

However, there are also some apparent counterexamples to (45) in Hungarian. The subject in the English sentence (46) can have both strong and weak interpretations, due to the telic character of the verb; but this sentence has two, unambiguous counterparts in Hungarian. Although both (47) and (48) are telic bounded sentences, in (48) the subject is utterly specific, while in (47) it can only be interpreted weakly.

(46) A chimney-sweep arrived.

(47) Egy kéményseprő   érkezett.
     a   chimney-sweep arrived

(48) Egy kéményseprő   megérkezett.
     a   chimney-sweep pfx-arrived

(47) and (48) indicate that the specificity of the subject can be marked in Hungarian by a prefixed verb-form of the same verbal stem, at least in some cases.[11] However, the semantic difference between these prefixless and prefixed predicates cannot be grasped with either the bounded/unbounded or the telic/atelic opposition, both (47) and (48) being telic and bounded.

The restrictions on the interpretation of the subjects in (47) and (48) can be attributed to a special property of the verb *érkezik* 'arrive'. This verb in Hungarian exibits the Definiteness Effect (henceforth DE), that is, it rules out subjects with strong determiners:

(49) *A/minden kéményseprő   érkezett.
      the/every chimney-sweep arrived

Now recall that the reasoning that the DE is due to the thetic character of judgements (see Ladusaw 1994) has been accepted here (see section **2.2**). Thus sentence

---

[11] It is far from being true that prefixes always make a (telic) predicate non-specifying in non-verbal sentences. For instance there are prefixes with some adverbial function, and these do not influence the strength of the indefinite subject (see later in this section).

(47) is a thetic, non-categorical statement, licencing only weak subjects, in both senses of weakness. This is also supported by the syntactic fact that the subject is not in the topic position, but is incorporated under the V' (see Szabolcsi 1986). That is, (47) is a verbal sentence according to the terminology used in Kálmán (1985).

Turning now to the opposite restriction in (48), notice that it constitutes a minimal pair with (47), so the only element that can be blamed for the constraint is the prefix *meg*.[12] While it is true that prefixed verbs never show the DE, they usually do not force strong reading upon their subjects either (see (41'), (43), (44)). So a general restriction ruling out non-specific arguments does not exist as an independent constraint in Hungarian (contra É. Kiss 1995).[13] Let us examine then, what differentiates between prefixed verbs in this respect, and why.

The prefixless verb *fest* 'paint' differs from the prefixless *érkezik* 'arrive' in that it does not restrict its subject's determiner in any way. On the other hand, prefixed versions of *fest* do not force strong reading upon the subject, contrary to the prefixed *megérkezik*. Comparing some more prefixless–prefixed verb pairs, it can be observed that if the prefixless verb does not force its subject to be weak, its prefixed versions do not require a strongly interpreted subject, either. Consequently, the fact that the prefixed verb *megérkezik* forces the strong interpretation upon its subject may be due to the fact that its prefixless counterpart is an intransitive DE-verb.

It is interesting that versions of *érkezik* with prefixes other than *meg* allow either the strong or the weak reading of the subject:

(50) Egy turista felérkezett a   csúcsra.
    a   tourist up-arrived   the top-to
    'A tourist arrived at the top'

[12] The translation of *meg* is missing here because its meaning cannot be given with any English word. It usually expresses perfectivity, but it can contribute to the meaning of a verb in quite different ways. Because of the non-compositional behaviour of verbal prefixes, Hungarian prefixed verbs are best treated semantically as lexical units.

[13] É. Kiss (1995) argues that there are some verbs in Hungarian exhibiting a Specificity Effect, that is, requiring the subject (and also the object when the verb is transitive) to be specific. This claim is based on data containing stative verbs and some prefixed verbs. However, the behaviour of statives, which are inherently individual-level predicates, can be explained with their argument structure (see Kratzer 1995), whereas it is simply not true that all prefixed verbs require strong arguments. As we have seen, the obligatorily strong interpretation of subjects of individual-level predicates can be attributed to the non-specifying (atelic and not place-bounded) character of these verbs, and this is supported by the fact that the same principle is valid for (non-place-bounded) activities as well. On the other hand, since prefixed verbs do not require specific arguments in general, an independently functioning Specificity Effect cannot be motivated by them, either.

(51) Egy lány elérkezett   a   tóhoz.
     a   girl   away-arrived the lake-to
     'A girl arrived at the lake'

Maybe the possibility of weak interpretation of subjects in (50), (51) follows from the directional nature of the prefixes, which requires place adverbials as complements. Then the effect of place-boundedness that licences weakly interpreted subjects extends beyond atelic verbs.

Thus it seems that DE-verbs have a counterpart with some non-directional prefix that prescribes strong interpretation for the argument which is obligatorily weak with the DE-verb. This observation is borne out by transitive DE-verbs as well: in (52) the object is obligatorily weak (in both senses of weakness), in (53) the indefinite object must be interpreted strongly, while in (54) the interpretation of the object can be either strong or weak.

(52) Hugó írt   egy levelet.
     Hugh wrote a   letter-acc
     'Hugh wrote a (non-specific) letter'

(53) Hugó megírt   egy levelet.
     Hugh pfx-wrote a   letter-acc
     'Hugh has written a (specific) letter'

(54) Hugó aláírt   egy levelet.
     Hugh under-wrote a   letter-acc
     'Hugh signed a letter'

Thus we can conclude that (45) is valid, that is, telicity in general makes the predicate specifying (licencing weakly interpreted subjects). Exceptions can be attributed to a disturbed symmetry: since the DE-verbs exclude strong noun phrases from their specified argument-positions (transitives from the object, intransitives from the subject position), there are prefixed counterparts of DE-verbs exhibiting an (almost) opposite restriction that excludes the weak interpretation[14] of the same arguments. This is presumably due to some tendency to have a lexically balanced chance of appearing weak and strong readings.

---

[14] Notice that these verbs exclude only the weak interpretation of their closest argument, but they tolerate weak determiners.

Summarizing the results of this section, a predicate can be specifying in Hungarian non-verbal neutral sentences if it is telic or place-bounded. Telicity is enough to allow weak subjects in most cases, place-boundedness seems to have no exceptions.

4. Conclusions

It is time to explain or at least arrange somehow the linguistic facts observed above.

Assume that there is a General Specifying Criterion put on every statement. This criterion reflects the trivial fact that sentences are about something:

(55) General Specifying Criterion (GSC): a statement must have at least one Specifying Feature.

The GSC can be met by the subject, by the predicate, or both. Thus Specifying Features satisfying the GSC belong either to the subject or to the predicate.

Features that count as Specifying Features with respect to the GSC are the following:

(56) Specifying Features of Subjects:
    1. strong determiners
    2. strong interpretation (of weak subjects)
  Specifying Features of Predicates:
    1. telicity[15]
    2. locatedness

Locatedness can be expressed differently in different languages. Here are some possibilities:

1. There are certain **syntactic constructions** that indicate or imply locatedness by themselves: *there*-sentences in English (notice that the expletive indicates some (possibly abstract) place); verbal sentences in Hungarian.

2. There are certain **verb forms** that are able to implicate locatedness. For instance, in English, where there are two different verb forms for the simple and continuous tenses, the latter form can indicate that the Kratzerian spatiotemporal slot is filled (notice that this form often goes hand in hand with some adverbial phrase(s) locating the eventuality spatially or temporally).

---

[15] See the reasoning on the necessary connection between telicity and locatedness at the end of section **2.4**.

3. If neither of the previous two types of markedness occur, there has to be present some other marker of locatedness in order to be able to interpret the subject weakly. For instance, in Hungarian, where there is only one Present and one Past Tense, there has to be some place-adverbial in the sentence to allow the weak reading of the subject.

If none of the Specifying Features of Predicates appear, it is the subject that must satisfy the GSC; that is, an SF must be present on it. This means that the subject has to be either a strong NP or a strongly interpreted weak NP.

In some cases strong interpretation of the subject is required even if there seems to be an SF of predicates present. An SF of this type is the Perfect in English, and in Hungarian there are prefixes with the same effect when attached to verbs that are DE-verbs in the prefixless form.

I close this paper with pointing at a remarkable consequence of the GSC. On the basis of (55), (56) it is possible to explain why individual-level predicates usually trigger a generic (strong) reading: a prototypical individual level predicate is stative (that is, not telic), and cannot be place-bounded (not having the argument slot Kratzer assumes). Thus, no predicate-SF being present, the subject has to be a strong or a strongly interpreted weak NP. But recall that there are generic sentences that can have an existentially interpreted (weak) subject as well:

(57) Typhoons arise in this part of the Pacific.

The predicate in (57) is explicitly place-bounded, and since locatedness is an SF, it is possible to interpret the subject weakly as well.

Finally, I would like to emphasize repeatedly that there are several other factors influencing the specificity of subjects. Syntactic arrangement, for instance, has been only mentioned very tangently. But I suppose that effects arising from other factors can be more clearly separated and more appropriately studied if constraints originating from semantic properties and principles have been clarified. The present paper aims at contributing to that end.

References

Bach, E. 1981. On time, tense, and aspect: an essay in English metaphysics. In: Cole, P. (ed.): Radical pragmatics, 63–81. Academic Press, New York.
Barwise, J. Cooper, R. 1981. Generalized quantifiers and natural language. In: Linguistics and Philosophy 4: 159–219.

Bosveld -de Smet, L. 1993. Indefinite subjects in French and stage-level versus individual-level predicates. In: Boer, A. de Jong, J. de Landeweerd, R. (eds): Language and cognition 3, 29 38. University of Groningen.

Carlson, G. 1977. Reference to kinds in English. Ph.D. dissertation. University of California, Irvine.

Dahl, Ö. 1995. The marking of the episodic/generic distinction in tense-aspect systems. In: Carlson, G.N. Pelletier, F.J. (eds): The generic book, 412 26. The University of Chicago Press, Chicago.

Depraetere, I. 1995. On the necessity of distinguishing between (un)boundedness and (a)telicity. In: Linguistics and Philosopy 18: 1 19.

É. Kiss, K. 1994. Generic and existential bare plurals and the classification of predicates. Working Papers in the Theory of Grammar, Vol. 1, No. 2. Research Institute of Linguistics, Hungarian Academy of Sciences, Budapest.

É. Kiss, K. 1995. Definiteness Effect revisited. In: Kenesei, I. (ed.): Levels and structures (Approaches to Hungarian, Vol. 5), 63 88. JATE, Szeged.

Enç, M. 1991. The semantics of specificity. In: Linguistic Inquiry 22: 1 25.

Groenendijk, J. Stokhof, M. 1980. A pragmatic analysis of specificity. In: Heny, F. (ed.): Ambiguities in intensional contexts, 153 90. Reidel, Dordrecht.

Hoop, H. de 1992. Case configuration and noun phrase interpretation. Ph.D. dissertation. University of Groningen.

Hoop, H. de 1995. On the characterization of the weak-strong distinction. In: Bach, E. Jelinek, E. Kratzer, A. Partee, B.H. (eds): Quantification in natural languages, 421 50. Kluwer Academic Publishers, Dordrecht.

Kálmán, L. 1985. Word order in neutral sentences. In: Kenesei, I. (ed.): Approaches to Hungarian 1, 13 23. JATE, Szeged.

Kleiber, G. 1981. Relatives spécifiantes et relatives non-spécifiantes. In: Le français moderne 49: 216 33.

Kratzer, A. 1995. Stage-level and individual-level predicates. In: Carlson, G.N. Pelletier, F.J. (eds): The generic book, 125 75. The University of Chicago Press, Chicago.

Ladusaw, W.A. 1994. Thetic and categorical, stage and individual, weak and strong. In: Harvey, M. Santelmann, L. (eds): Proceedings from Semantics and Linguistic Theory IV, 220 29. Cornell University, Ithaca.

Ludlow, P. Neale, S. 1991. Indefinite descriptions: in defense of Russell. In: Linguistics and Philosophy 14: 171 -202.

Milsark, G. 1974. Existential sentences in English. Ph.D. dissertation. MIT, Cambridge MA.

Szabolcsi, A. 1986. From the Definiteness Effect to lexical integrity. In: Abraham, W. de Meij, S. (eds): Topic, focus and configurationality, 332 60. John Benjamins, Amsterdam.

Address of the author: Márta Maleczki
                           Department of General Linguistics
                           József Attila University
                           Egyetem u. 2.
                           H-6722 Szeged, Hungary
                           e-mail: maleczki@hung.u-szeged.hu

NEGATIVE POLARITY ITEM LICENSING IN HUNGARIAN

ILDIKÓ TÓTH

Abstract

The present paper is concerned with licensing Negative Polarity Items (NPI) in Hungarian, both locally and across clause boundaries. Two types of NPIs are described and their distributional properties are examined. After considering two possible analyses of NPI-licensing, one based on Generalized Binding and one capitalizing on the properties of NPIs as indefinites, I argue that Hungarian NPIs are better captured within the latter framework. Data are drawn from different constructions like wh-questions with rhetorical readings, long distance lincensing of negatives, factive islands, and multiple negation within a single clause. I conclude that in Hungarian the two different NPIs should be distinguished and that they involve different licensing mechanisms, both crucially depending on their indefiniteness.

Introduction

Polarity Sensitive Items (PSIs) are elements whose distribution and interpretation are sensitive to negative vs. affirmative contexts. PSIs include Negative Polarity Items (NPIs) and Positive Polarity Items (PPIs). NPIs cannot freely occur in any sentence, they have to be licensed by a negative polarity environment. Polarity environments are understood to involve overt negation, conditionals, yes/no questions, adversative predicates etc. The distribution of these items has been a long studied phenomenon in generative linguistics and several theories have been developed to account for the data that rely on the semantic features of polarity environments (see e.g. Ladusaw 1980; 1992; 1994; Linebarger 1981; Giannakidou 1994; Giannakidou–Quer 1995a; 1995b). An alternative analysis has been proposed by Progovac (1988; 1991; 1992a; 1992b), who claims that NPI-licensing can be more successfully analysed within the syntactic framework of the Generalized Binding Theory.

The purpose of this paper is to discuss the distributional properties of NPIs in Hungarian and consider the differences between them and NPIs in other languages. In order to give an account for NPI-licensing in Hungarian, I will consider and compare two different theories of NPI-licensing: Progovac's Generalized Binding Approach (in particular Progovac 1992a) and an approach based on Ladusaw's

(1992; 1994) analysis of negative indefinites (NIs) which was further developed in Giannakidou–Quer (1995a; 1995b). I will argue that in spite of the fact that Hungarian data prima facie appear to support the Generalized Binding Approach, such an analysis faces both theoretical and empirical problems. Showing that an analysis in Progovac's framework is neither the only possible one nor the most successful one and that treating NPIs as negative indefinites has certain advantages, I will finally adopt the analysis of NIs proposed in Ladusaw (1994) and in Giannakidou–Quer (1995a; 1995b).

1. Two analyses

1.1. Generalized Binding Approach

On the basis of Serbian/Croatian and English data, Progovac (1988; 1991; 1992a; 1992b) argues that since NPIs depend on negation, they are anaphoric. They are subject to Principle A of the Binding Theory: they have to be bound in their governing category.[1] She takes the negative particle to be in an A'-position and she adopts Aoun's Generalized Binding Framework (Aoun 1985; 1986), which deals both with A- and A'-binding. To give a unified account of NPIs licensed in negative environments without overt negation and by matrix negation, she makes use of the relativized notion of a governing category (Aoun 1985; Chomsky 1986): it is the first potential antecedent that creates the governing category. Potential antecedents for NPIs are negation in Infl, a null polarity operator Op in [Spec,CP], and matrix negation. Thus she posits the following structure with all the potential antecedents for the VP-internal NPI in an embedded clause:

(1)   [$_{CP}$ Op [$_{C'}$ [$_{IP}$ [$_{I'}$ Neg [$_{VP}$ [$_{CP}$ Op [$_{C'}$ [$_{IP}$ [$_{I'}$Neg [$_{VP}$ ... *any* ...]]]]]]]]]]]]

The governing category for a VP-internal NPI will be the Infl of its own clause, since it contains the NPI itself, the first potential antecedent (negation) and an accessible subject.

Progovac claims that being subject to Principle A of the Binding Theory is a universal characteristic of NPIs, and that the wide crosslinguistic variation can be accounted for with the help of parametric variations:
   – subjecthood to different Binding Principles
   – raising possibilities at LF

---

[1] Y is the governing category for X if and only if Y is the smallest maximal projection containing X, a governor of X, and a SUBJECT accessible to X.

The first parameter can have three possible values. Some NPIs are A'-anaphors, subject only to Principle A of the Generalized Binding framework (e.g. Serbian/Croatian NI-NPIs, Chinese *congali* and English NPIs). These NPIs will have to be bound in their governing category at LF and, consequently, can be licensed only by clausemate negation. Other NPIs are anaphoric pronominals subject to Principles A and B (e.g. Serbian/Croatian I-NPIs). In this case, NPIs have to satisfy two conditions. They have to be free in their governing category at S-Structure but bound in their governing category at LF. Consequently, these elements cannot appear with clausemate negation. If, however, they can raise at LF (I-NPIs for example can), they will be licensed by superordinate negation and in non-negative contexts, since these licensers fall outside their governing category. Finally, some NPIs are subject to Principles A and C (e.g. Italian and French NPIs).

The raising parameter at LF has four possible values. Some NPIs do not raise (e.g. NI-NPIs, English strict NPIs like *until,* and Chinese *congali*). Other NPIs can raise at LF and thus can be bound long distance. NPIs that can raise both by IP-adjoining and by moving through the Spec of Comp will be licensed by superordinate negation and in non-negative contexts (e.g. Serbian/Croatian I-NPIs).

Some NPIs can only IP-adjoin at LF (e.g. Catalan NPIs like *ningu*), so their first potential antecedent will be the polarity operator in the Spec of Comp.

Finally, those NPIs that can only move through the Spec of Comp at LF (e.g. Turkish NPIs) extend their governing category to the matrix IP and are licensed only by matrix negation.

The role of the null polarity operator Op is to account for the licensing of NPIs in non-negative contexts like conditionals, questions, complement clauses of adversative predicates, etc. With the help of this operator, Progovac (1992a) proposes a semantico-syntactic analysis which is claimed to resolve several problems that a purely semantic or purely syntactic analysis inevitably faces. Her model combines a modified version of Ladusaw's (1980) downward entailment (DE) theory and Progovac's (1988) binding approach. The former accounts for licensing conditions while the latter for locality conditions.

The modification of Ladusaw's theory proposed by Progovac (1992a) is a shift from DE to non-upward entailment (UE). This step is claimed to be necessary since certain environments allow NPIs but they are not DE.[2]

To solve this problem Progovac argues that there is a polarity operator (Op) in [spec, CP] of any non-UE clause and it is this operator that licenses NPIs. The role

---

[2] Progovac examines two such contexts: yes/no questions and the determiner *only*. For more details see Progovac (1992a; 1992b).

played by DE does not disappear completely but becomes indirect by virtue of the following filter:

(2)   UE filter: *Polarity operator in an Upward-Entailing (UE) clause.

Adopting the UE hypothesis Progovac correctly predicts that yes/no questions will be NPI-licensers since they are neither UE nor DE. There are, however, some problems with this analysis. Firstly, the operator *exactly* is neither UE nor DE and yet, it does not license NPIs.

(3)   *Exactly two boys read anything.

Progovac makes a remark about this problem in a footnote and suggests that one additional assumption is needed to account for the ungrammaticality of (3). Her tentative suggestion is that it is ungrammatical since the operator *exactly* commits us to the positive truth value of the sentence and it is incompatible with the appearance of NPIs. Although this is true, this in itself sounds like a restatement of the facts.

Secondly, there are contexts which cannot be judged by the criteria of DE and UE. In another footnote Progovac (1992a) mentions one such case: belief contexts. There are, however, other examples, which belong to the class of non-negative polarity contexts and license NPIs. In embedded clauses in Hungarian, the most frequent complementizer *hogy* can combine with other complementizers and/or heads to form complex conjunctions. Kenesei (1992; to appear) gives a list of the large array of these conjunctions from among which *nehogy* and *mintsem hogy* license the NPIs *valamit is, valakit is*.

(4)   Elhallgattam,   nehogy   valamit is   meghalljanak.
       shut up -1sg   not-that   anything-acc   hear-3pl-subj
       'I shup up lest they hear anything'
(5)   Inkább   elmegyek,   mintsem   hogy   valakit is   megbántsak.
       rather   away-go-1sg   than-not   that   anybody-acc   hurt-1sg-subj
       'I rather go away than hurt anybody'

Of course it is possible to say that these contexts are neither UE nor DE, since the criteria for entailments cannot be applied. Consequently, Op can be generated in [spec, CP]. But such an argument would miss an important characteristic of these examples: they have a purposive interpretation and the embedded clauses do not have their truth value fixed.

Thirdly, Progovac offers a solution to the problem of wh-questions containing NPIs, which faces several difficulties. Her argument is based on two claims: wh-questions are UE and the presence of NPIs always triggers rhetorical readings. The first of these claims is simply not true and the second is only partially true. I will return to this problem in section **3.1.**

Finally, the postulation of Op is problematic in itself. Progovac (1992a) is rather vague about the nature of this negative operator. Syntactically she needs it in order to be able to establish locality restrictions that NPIs obey. In a very short section titled "The semantics of operator" she remarks that "Op has bearing on truth conditions of its clause" and assumes that Op represents a switch with a +/− choice, the minus value being responsible for NPI-licensing. But the semantic content of the operator is not stated clearly, thus it is possible to make use of Op whenever the data seems to require it.

1.2. Weak and strong licensing of negative indefinites

Negative Indefinites (NIs) are a subclass of NPIs that are either NPs or adverbs (like *anything, anywhere, ever,* etc.). A different approach to NPI-licensing stems from Ladusaw's (1992) analysis of NIs in negative concord (NC) languagues.

Ladusaw (1992) claims that NIs are Heimian indefinites without any quantificational force of their own. He proposes that licensing of NIs is not a uniform process and assumes that even in negative contexts two different mechanisms are at play. This accounts for the interpretational ambiguity displayed by NIs. The strong construal results in a universal negative interpretation, while the weak construal gives rise to existential reading. Furthermore, Ladusaw (1994) proposes that a direct mapping between syntax and semantics becomes possible if we consider the tripartite structure Op [Restrictor] [Matrix] for the interpretation of quantificational phrases and identify the syntactic specifier with the restriction of an operator and the scope of the operator with the matrix. In this way, strong licensing becomes an instance of A'-movement to a specifier position at least at LF establishing a spec-head relation between the NI and negation, thus satisfying the Neg-criterion (in the spirit of Haegeman 1992; Haegeman–Zanuttini 1991). Syntactic complements are mapped onto the matrix of the licensing operator and thereby weak licensing is conceived of as in situ licensing of NIs via existential closure. In this way the universal negative and the existential (NPI) readings of NIs are results of two distinct syntactic mechanisms. Giannakidou and Quer (1995a; 1995b) follow Ladusaw (1992; 1994) in their analysis of NPI-licensing in Greek and Catalan and go even further. They extend Ladusaw's theory to NPI-licensing in general and claim that long distance licensing of NPIs can be analysed in the same way. This analysis has the advantage of unifying negative concord and NPI-licensing in general. Complement clauses are mapped onto the matrix of the licensing operator and therefore weak

licensing of NPIs is available. Strong licensing is possible only if the embedded domain does not block movement of the NI across a clause boundary. Giannakidou and Quer (1995a; 1995b) claim that such transparency is determined by the type of the matrix predicate and that tense dependencies play a crucial role.

Considering NPI-licensing in general also requires a definition of what counts as an "appropriately negative operator" that can license NPIs. Giannakidou and Quer (1995a; 1995b) propose that NPIs are grammatical if and only if they are in the scope of non-veridical operators. An operator Op is non-veridical if and only if Op p does not entail p. As noted before, under this analysis NC becomes just a sub-case of NPI-licensing in general. To capture the different licensing conditions of negative concord terms and NPIs like *anybody*, Giannakidou and Quer (1995b) pro-pose that strong licensing (which in effect corresponds to NC) can take place only under averidical operators in which case Op p entails $\neg p$.

2. The distribution of NPIs in Hungarian

Let us now consider the Hungarian data. There are two types of NPIs in Hungarian. I will call one group SE-NPIs (since they involve words beginning with the prefix *se*, like e.g. *senki* 'nobody', *sehol* 'nowhere', *semmit* 'nothing'),[3] and the other group VALA-NPIs (they involve words like *valaki is* 'anybody', *valami is* 'anything', *valahol is* 'anywhere'.[4]

A few words on the role of *is* are necessary here. The particle *is* in Hungarian has several different meanings. To begin with, we must differentiate between emphatic and quantificational *is*. Emphatic *is* is similar to English 'indeed' and need not con-cern us here. Quantificational *is* means 'also' and it most typically modifies NPs cre-ating a quantifier phrase out of an NP (see Hunyadi 1981; Piñón 1992).

The *is* occurring in VALA-NPIs should be considered as a third instance of this particle: it turns PPIs into NPIs. While Hungarian PPIs like *valaki* 'somebody', *vala-mi* 'something', *valahol* 'somewhere' need no licenser and have the same distribution as their English counterparts, the presence of *is* changes this situation. *Valaki is, valami is* etc. can only occur in polarity contexts. To say that *is* has quantificational force in this case would be incorrect since, as I will argue later, NPIs are best analysed as Heimian indefinites. At this point, all I can say about *is* in VALA-NPIs is that it sig-

---

[3] Throughout the paper, SE-NPIs will be glossed as *nobody, nothing, nowhere*, etc., because they have morphological negation. Note, however, that their distribution is different from that of the English negative words.

[4] VALA-NPIs will be glossed as *anybody, anything, anywhere*, etc., although their distribution is not identical with that of English *any*-phrases.

nals the lack of existential entailment that according to Haspelmath (1993) character-izes both the licensing contexts and the polarity items themselves. Unfortunately, I can offer no satisfactory explanation why it is exactly *is* that plays this role.

Returning to the distributional properties of Hungarian NPIs, we can say that while SE-NPIs occur only with clausemate negation, VALA-NPIs are licensed by superordinate negation and in non-negative polarity contexts, which do not contain the negative particle *nem*. Thus the two NPI types in Hungarian are in complemen-tary distribution:

SE-NPIs

(6) Pál   nem   látott   senkit.
    Paul   not   saw   nobody-acc
    'Paul did not see anybody'

If there is no negative particle, or negation occurs in the superordinate clause, the sentence is ungrammatical:

(7) *Pál   látott   senkit.
    Paul   saw   nobody-acc
    'Paul saw nobody'

(8) *Mária   nem   mondta   hogy   Pál   látott   senkit.
    Mary   not   said   that   Paul   saw   nobody-acc
    'Mary did not say that Paul saw anybody'

Furthermore, SE-NPIs cannot appear in non-negative polarity contexts like yes/no questions, conditionals and complement clauses of adversative predicates:

(9) *Látott   Mária   semmit?
    Saw   Mary   nothing-acc
    'Did Mary see anything?'

(10) *Ha   Mária   hallott   volna   semmit,   megsértődött   volna.
     If   Mary   heard   would   nothing-acc hurt   would
     'If Mary had heard anything, she would have been hurt'

(11) *Péter   kétli,   hogy   Mária   látott   semmit.
     Peter   doubts that   Mary   saw   nothing-acc
     'Peter doubts that Mary has seen anything'

VALA-NPIs

VALA-NPIs are allowed in the complement clause of a negated matrix sentence but are illicit with clausemate negation:

(12) Pál   nem   mondta,   hogy   Mária   valakit is   látott.
     Paul   not   said   that   Mary   anybody-acc   saw
     'Paul did not say that Mary saw anybody'

(13) *Pál   nem   mondott   valamit is.
     Paul   not   said   anything-acc
     'Paul did not say anything'

They can also appear in non-negative polarity contexts:

(14) Tanultál   valaha is   oroszul?
     studied-2sg   ever   Russian
     'Have you ever studied Russian?'

(15) Kétlem,   hogy   ezt   valaki is   megértette.
     doubt-1sg that   this-acc   anybody   understood
     'I doubt that anybody has understood this'

(16) Ha Mária   valamit is   elfelejtene,   ismételd el   neki.
     if Mary   anything-acc   forgets-cond   repeat-imp-2sg her
     'If Mary forgets anything, repeat it to her'

So we can see that while SE-NPIs are licensed by clausemate negation, VALA-NPIs are licensed only by superordinate negation and in non-negative polarity contexts. In positive contexts, however, neither SE-NPIs nor VALA-NPIs can appear:

(17) *Pál   látott   senkit.
     Paul   saw   nobody-acc
     'Paul saw nobody'

(18) *Pál   látott   valakit is.
     Paul   saw   anybody-acc

3. Analysing the Hungarian data

Let us first have a look at the evidence in favour of the Binding Analysis. At first sight the Hungarian data above suggest that SE-NPIs behave like NI-NPIs in Serbian/Croatian.[5] They are pure A'-anaphors and must be bound in their governing category, the local clause. (6) is grammatical as *senkit* is bound in its governing category by the negative particle *nem*. In (7)–(11), on the other hand, there are no binders for SE-NPIs inside their governing categories and if we suppose that SE-NPIs cannot raise at LF, ungrammaticality will follow.

Let us now consider Hungarian VALA-NPIs. As we have already seen, they are licensed by matrix negation and in non-negative contexts but cannot occur with clausemate negation. This suggests that they behave as anaphoric pronominals: they must be free in their governing category at S-structure and they have to be bound in their governing category at LF. The claim that LF movement is indeed involved in the case of these anaphoric pronominals seems to be supported by examples where licensing of VALA-NPIs takes place across two (or more) clause boundaries:

(19) Kétlem,   hogy   Mari   mondta, hogy   valakit is   meghívott.
     doubt-1sg that   Mary   said   that   anybody-acc invited
     'I doubt that Mary said that she invited anybody'

It has also been claimed (Baker 1970; Ross 1967; Progovac 1988) that NPI raising at LF exhibits some island effects, e.g., Complex NP islands seem to block NPI-licensing in English:[6]

(20) *We did not know the fact that anyone had arrived.
(21) *Do you believe the claim that anybody was looking for anything?

The same blocking effects arise in the case of Hungarian:

(22) *Nem   tudtam   azt   a   tényt, hogy   valaki is   megérkezett.
      not   knew-1sg that   the   fact   that   anybody   arrived-3sg
      *'I did not know the fact that anybody had arrived'

---

[5] In fact, this similarity is even more striking once we also compare I-NPIs with VALA-NPIs. As Szabolcsi (1994) remarks, Serbian/Croatian and Hungarian seem rather unique in the distributional properties of their NPIs: a perfect complementary distribution. SE-NPIs and VALA-NPIs (just like NI-NPIs and I-NPIs) together cover the wide range of negative polarity contexts without any overlap. But differences do exist, some of which will be discussed later.

[6] Examples (20) (21) are from Ross (1967).

Considering these data it is tempting to say that Hungarian NPIs are well account-
ed for in the Binding Analysis: the complementary distribution follows from the
assumption that SE-NPIs are anaphors while VALA-NPIs are anaphoric pronominals.

Yet, a closer look at some other phenomena suggests that the Binding Approach
cannot offer a satisfactory account of the facts. Among the several possible objec-
tions I will concentrate on the following four:

(i) The problem of wh-questions containing NPIs and giving rise to rhetorical
readings.

(ii) SE-NPIs in the focus field in Hungarian and double instances of the nega-
tive particle within one clause.

(iii) Long distance licensing of SE-NPIs in Hungarian and related phenomena
from other languages.

(iv) Blocking effects of factives in long distance licensing.

Let us take these points one by one and see how they support one analysis or
the other.

3.1. Wh-questions and NPIs

Adopting Progovac's analysis would imply that we also accept the postulation of her
negative operator Op in Comp in order to be able to account for licensing in non-neg-
ative contexts. Apart from the problems mentioned in section **1.1** concerning the exact
nature of this operator, some futher problems arise in connection with Op.

One of the arguments in favour of Op is that it is supposed to explain why wh-
questions are rhetorical questions when they contain NPIs. As Progovac points out,
the assumption that NPI-licensing involves an operator in CP raises the question of
(in)compatibility with other operators occupying the same position. Since wh-
words are analysed as occupying the same position at LF (and also at S-structure in
case of overt wh-movement), we would expect Op and wh-words to be mutually
exclusive. Contrary to this expectation, NPIs do appear in wh-questions:

(23) Who did Mary ever kiss on the first date?

(24) Mikor hivtál   meg   valakit is   a   születésnapodra?
      when   invited-2sg VM   anybody-acc   the   birthday-2sg-on
      'When did you invite anybody to your birthday?'

What should be noticed in these wh-questions containing NPIs is that the answers
do not range over a domain of different possibilities. Rather, the question is rhetor-
ical and a negative answer is already implied in the question itself. Progovac's argu-
mentation goes as follows:

Let us suppose that Op is freely generated in the Spec of CP in all wh-ques-
tions and that the operator switch is set negatively. In order to avoid a doubly filled
Spec of CP, the wh-word and the negatively set Op merge. This merger produces a
negated NPI. The rhetorical interpretation of wh-questions with NPIs follows and
NPIs are licensed by the negated NPI in Comp. An implicit assumption of this
analysis is that the wh-feature of wh-words is suppressed.[7]

The analysis is attractive but there are problems with it. One problem concerns
the claim that wh-questions containing NPIs have a rhetorical reading. As noted by
Borkin (1971) and Lawler (1971) wh-questions starting with *why* and *how* do not
trigger rhetorical interpretations.

Lawler calls these wh-words "factives" because they presuppose that the
action itself has taken place. While in English they allow NPIs without rhetorical
reading, in Hungarian they do not even allow VALA-NPIs:

(25) (a) Miért   hívtál   meg   valakit is   a   születésnapodra?
         why   invited-2sg   VM   anybody-acc   the   birthday-2sg-on
         'Why did you intive anybody for your birthday?'
     (b) *Hogyan magyaráztál   el   valamit is   Péternek?
         how   explained-2sg VM   anything-acc   Peter-dat
         'How did you explain anything to Peter?'

If Op is freely generated in Spec of CP in all wh-question as Progovac argues, the
ungrammaticality of (25) is not expected.

Interestingly enough, the problem is reminiscent of that raised by the operator
exactly: the wh-words *why* and *how* also commit us to the presupposition of a
proposition and this is incompatible with the appearance of NPIs. These cases sug-
gest that the descriptive generalization mentioned by Progovac herself, namely that
polarity clauses share the property of not having their truth value fixed positively,
is at least as important as the question of downward and upward entailment. This
suggestion becomes even stronger once we remember our examples (4) and (5),
which we characterized in the same way.

---

[7] In fact, the argumentation is even more complicated due to the fact that Progovac claims that
wh-questions are UE. Consequently, she has to find a way to circumvent the UE-filter. I think this is
incorrect. Depending on the exact definition of what it means for a wh-question to be UE or DE, we
can get two different results: either wh-questions are DE or they are neither DE nor UE. The defini-
tion used by Progovac is based on Karttunen (1977): "for a question A to entail a question B every
true answer to A has to entail a true answer to B". This criterion gives the result that wh-questions are
neither UE nor DE, contrary to what Progovac claims. Since this has no implications for my argument
against Op in general, I will not discuss it in more detail.

Turning to embedded questions, we face the following problem: if grammatical instances of NPIs signal the presence of Op and rhetorical force is due to merging the wh-word with Op creating a NPI in Comp, then how is it possible that verbs subcategorizing for +wh Comp can still license VALA-NPIs:

(26)   Azt   kérdeztem, hogy mikor hívtál meg   valakit is   vacsorára.
        that asked-1sg   that   when   invited-2sg anybody-acc dinner
        'I asked you when you invited anybody for dinner'

Once again a negative answer is implied in the question, and the presupposition is that you never invited anybody. But at the same time the matrix verb can also satisfy its need for a +wh Comp. To account for this fact we would have to stipulate a rather strange ordering of satisfying different filters: first Comp carries the +wh feature and thus satisfies the subcategorizing requirements of the matrix verb, but then a merger takes place and gets rid of such a feature. I do not think this is a very plausible solution.

3.2. Double negation within one clause and SE-NPI-licensing

SE-NPIs also pose some problems for the claim that they must be bound in their governing category by negation. In Hungarian two instances of the negative particle *nem* are grammatical within a clause as long as they satisfy certain conditions.[8] Consider the following configurations in the focus field in Hungarian, which are directly relevant to our concerns (cf. Brody 1990 for a detailed discussion):

(27) (a) PÉTER nem   szeret   senkit.
         Peter   not   likes   nobody-acc
         'It is Peter who does not like anybody'
     (b) Senkit   nem/sem   PÉTER   szeret.
         nobody-acc   not/also not Peter   likes
             same as (27a)
     (c) *Nem PÉTER szeret senkit.
     (d) Nem PÉTER nem szeret senkit.

---

[8] The most crucial condition is adjacency of the focused element and the verb. A negative particle can in fact intervene (moving up there by head-adjunction to the verb) but not a SE-NPI. The ungrammaticality of (i) is accounted for in Brody (1990) by assuming that V (or Neg^0+V) must be adjacent to an element in spec of FP in order to transmit the feautre +focus (for more details see Brody 1990).

(i)   *PÉTER senkit   nem szeret.
        Peter   nobody-acc   not likes
        'It is Peter who does not like anybody'

The first point to make is about the nature of the negative particle occupying a position above the focused element. To exclude (27c) we could suppose that *nem* in front of the focused element is an instance of constituent negation, and therefore cannot license SE-NPIs. But such a proposal would run into difficulties when accounting for (27b). *Senkit* is licensed in this case, consequently the negative particle cannot be constituent negation.[9] Therefore, in (27b) *nem* must occupy an A' position (possibly a NegP) that is above FP.

The second point is that negation can also appear in a functional projection that is below the focused element. This is evident from examples (27a) and (27d). We conclude that in Hungarian a clause can have two NegP projections, one above FP and one below it.

Having said that, it is difficult to see how the contrast between (27a) and (27b) could be accounted for in Progovac's analysis. The first potential antecedent is negation below FP that can move from Neg^0 to adjoin to F^0 in case there is a focused element (like in (27a)). Consequently, the governing category does not extend above FP since FP already contains the NPI itself, the first potential antecedent (negation) and an accessible subject. From this it also follows that negation above FP is not the first potential antecedent. The claim that SE-NPIs are anaphors can account for the grammaticality of (27a), but the contrast between (27a) and (27b) remains unexplained. In (27b) *senkit* is not locally bound. It has moved out of its governing category into spec of NegP (above FP) and got licensed there. The claim that SE-NPIs are anaphoric and must be locally bound would predict (27b) ungrammatical.

The configurations for (27a) and (27b) are given in (28a) and (28b), respectively:[10]

(28) (a) $[_{FP}$ PÉTER $[_{F^0}$ nem$_i$ szeret $[_{NegP}$ senkit $[_{Neg^0}$ t $_i$]]]]

(b) $[_{NegP}$ Senkit$_j$ $[_{Neg^0}$ sem $[_{FP}$ PÉTER $[_{F^0}$ szeret$_i$ [.. t$_i$ t$_j$]]]]]

3.3. Licensing of SE-NPIs by matrix negation

The claim that SE-NPIs behave like A' anaphors and that, unlike English *any*-NPIs, cannot raise at LF means that licensing in any other polarity contexts is impossible for them. Certain verbs, however, seem to allow overt raising of SE-NPIs and

---

[9] The possible alternation between *nem* and *sem* is due to the fact that the SE-NPI has been moved to a position adjacent to the negative particle. *Sem* is a contracted form of *is+nem*, meaning 'also not' i.e. 'neither'.

[10] Note that the satisfaction of the Neg-Criterion is possible in (28a) since the negative element *senkit* in Spec of NegP is in a spec-head configuration with the trace of the negative head *nem* which has cliticized onto the finite verb. Similar constructions are argued for in Haegeman Zanuttini (1991) for West Flemish and in Puskás (1992) for Hungarian.

thus license them in the matrix sentence. These verbs have been traditionally referred to as bridge verbs. This set includes volitional verbs like *akar* 'want', and 'neg-raising verbs' like *hisz* 'think', *gondol* 'believe', etc.

(29) Senkit$_i$ sem akarok,   hogy meghív   t$_i$.
      nobody-acc not want-1sg that   invite-2sg-subj
      'I don't want you to invite anybody'

(30) Senkit$_i$ sem kértem,   hogy meghívj   t$_i$.
      nobody-acc not asked-1sg that   invite-2sg-imp
      'I didn't ask you to invite anybody'

(31) Senkit$_i$ sem hiszem,   hogy meghívtál   t$_i$.
      Nobody-acc not believe-1sg that   invite-2sg-ind
      'I don't believe that you have invited anybody'

Progovac deals with a similar phenomenon in case of NI-NPIs in Serbian/Croatian and NPIs in embedded subjunctive clauses in Italian and French. Her claim is that in subjunctives, independent tense is absent and this makes it possible for Infl to delete at LF. Comp can also delete if it carries no unrecoverable material. She argues that domain extension takes place only with volitional verbs, exactly because their Comp does not carry any independent truth value index. In this way, absence of independent Tense and Truth specifications render the functional projections Infl and Comp unnecessary for interpretation, giving rise to what has been called "clause union effect". Since NPIs are A'-anaphors whose potential antecedents are in embedded Infl, in Comp and in matrix Infl, both Comp and Infl deletion are required for licensing by matrix negation in case of those NPIs that cannot raise at LF like *nikoga, nessuno*:

(32) (a) Non pretendo [$_{CP}$ che tu   arresti   nessuno]
           neg require-1sg that you arrest-subj noone-acc
           'I don't require that you arrest anybody'
      (b) Ne zelim [$_{CP}$ da   mrzim nikoga]
           not wish-1sg that hate   anyone-acc
           'I don't wish to hate anyone'

Although this analysis can successfully account for the grammaticality of sentences (32a) and (32b), it faces both conceptual and empirical problems. Conceptually, it is not clear how a head that hosts agreement features can get deleted. Deletion

of Comp in case of subjunctives is also problematic since absence of deictic tense is only true in a subset of subjunctive clauses and this subset does not coincide with those that allow NPI-licensing across a clause boundary. In the following Greek examples the subjunctive subordinate clauses have independent tense:[11]

(33) (a) Prepi   na   efije   xthes.
         must-3sg subj left-3sg yesterday
         'He must have left yesterday'
     (b) Tha ithela   na to ixes   skefti   prin   su   to zitiso   ego.
         fut. wanted-1sg. subj it had-2sg thought before you it ask-1sg I
         'I would like you to have thought of it before I asked you'

The Hungarian data in (29)–(31) poses empirical difficulties for any analysis invoking deletion of Comp and Infl. As can be seen from (31), subjunctive is not the only mood that makes licensing of SE-NPI across clause boundary possible. Furthermore, when SE-NPIs are licensed long distance, overt raising is always enforced into a position adjacent to the negative particle *nem*:[12]

(34)   *Nem hiszem, hogy meghívtál   senkit.
         not   think-1sg that invited-2sg nobody-acc
         'I do not want you to invite anybody'

This is in sharp contrast with constructions that involve infinitival complements, where overt raising is optional and clause union has really taken place:

(35) (a) Nem szeretnék   senkit   megbántani.
         not   like-1sg-cond nobody-acc hurt-inf
         'I would not like to hurt anybody'
     (b) Senkit   sem   szeretnék   megbántani.
         nobody-acc not   like-1sg-cond hurt-inf
         same as (35a)

[11] These data are from Anastasia Giannakidou (personal communication).

[12] As Progovac pointed out (personal communication) NI-NPIs can (but do not have to) move up to matrix Infl in case of restructuring verbs. But overt raising is restricted to this class of verbs and no 'mixed type' sentences are allowed. Furthermore, in Serbian/Croatian negation in matrix Infl can license NI-NPIs in the embedded clause only if the subjects of the main and the embedded clause are identical. This is not the case in Hungarian.

Progovac's raising parameters involve only LF raising. The NPIs that she examines do not have to move overtly when they are licensed by matrix negation: they can either move at LF to extend their governing category or, in case of volitional verbs (restructuring verbs in her terms), the domain is extended through Infl/Comp deletion.

In fact, Hungarian is not the only language where negative elements that are in most of the cases only licensed by clausemate negation also get licensed across a tensed clause boundary. As Haegeman and Zanuttini (1991) point out about West Flemish (WF), "there are locality constraints on the relation between *en* and the negative constituent... let us say at first approximation that the negative clitic *en* is licensed by a clausemate negative constituent with sentential scope." Here are their relevant examples:

(36) (a) *...da Valere en-wist   da   zen voader geen geld   oat.
         that   Valere not knew that his father no   money had
         'that Valere did not know that his father had no money'

     (b) *...da Valere an niemand zei   da   Marie ziek en-was.
         that   Valere to nobody   said that Marie ill   not was
         'that Valere said to nobody that Marie was not ill'

In (36a), the intervening clause boundary blocks the relationship between *en* and the NPI. In (36b), the NPI in the matrix clause cannot relate to *en* in the embedded clause.

In case of neg-raising verbs, however, the negative element can move up from the lower clause to the matrix clause overtly and thus license *ne*:

(37)   Niets   en-pienzenk da   ze   wilt doen.
         nothing not think-1sg that she wants do-inf
         'I do not think there is anything she wants to do'

Examples (37) and (36a) are reminiscent of our previous Hungarian examples (31) and (34). Considering further WF data, Haegeman and Zanuttini (1991) conclude that a binding analysis is unsatisfactory. Their main argument for such a conclusion is that while a binding relation can be established between a subject NP and an element (a) within a coordinate structure, (b) in a VP-internal position, (c) in extraposition, it can be shown that VP-internal negative elements, negatives coordinated with non-negative constituents, and extraposed negative constituents cannot license *ne*. Thus the possibility of establishing a local binding relationship cannot be the relevant question.

As Giannakidou and Quer (1995a; 1995b) show, Greek and Catalan NPIs also support the claim that subjunctive modality is not the decisive factor for long distance licensing of NPIs. They also have examples for long distance licensing with indicative mood in the embedded clauses. Their claim is that transparency of the complement domain for NPI-licensing is determined by the selecting predicate and by tense dependencies.

Moreover, empirical problems arise with Progovac's typology of NPI-licensing by matrix negation when we consider the morphological make-up of Greek NPIs. Giannakidou and Quer (1995a; 1995b) show that although Greek NPIs are not morphologically negative, they are subject to diverse locality requirements. In particular, in Greek the same lexical item can be licensed by different means depending on whether it is emphatic or not: emphatic NPIs depend on the co-occurrence of sentential negation while non-emphatic NPIs do not. Such distributional differences are not predicted by Progovac's Binding Analysis.

Going back to the Hungarian data, independent evidence against the domain-extension analysis is provided by 'mixed-type' sentences where both VALA-NPIs and SE-NPIs are licensed by matrix negation:

(38) Senkinek   sem   akarom,   hogy   ezt   valaha is elmondd.
     nobody-dat not   want-1sg that   this-acc ever   tell-2sg-subj
     'I don't want you to ever tell this to anybody'

Any analysis invoking some kind of deletion or transparency of Comp and Infl would fail to account for these 'mixed type' sentences since VALA-NPIs do not allow clausemate negation. Consequently, the presence of *valaha is* indicates that clause union could not have taken place.

3.4. Factivity and licensing of VALA-NPIs

Let us return now to the question of LF raising of certain NPIs and the argument in favour of a raising analysis, namely, that NPIs seem to exhibit island effects. Our examples (20)–(22) at first sight are clear cases of CNPC. A closer look at some of Ladusaw's examples and at some Hungarian sentences containing complex NPs, however, will require reconsideration of the data:

(39) Are you appalled by the idea that anyone would lift a finger to visit Cleveland?

(40) Nem látom   annak   a   lehetőségét, hogy valamit is   megértsek.
     not see-1sg of-that the possibility that anything-acc understand-1sg-subj
     'I don't see the possibility that I can understand anything'

These examples strongly suggest that it is not the CNPC that prevents licensing of NPIs in (20)–(22). In examples like (39)–(40) complex NPs do not block licensing of *any*-NPIs and VALA-NPIs by matrix negation, contrary to what we would expect if CNPC were at work. Furthermore, in the case of some verbs, licensing of VALA-NPIs by matrix negation is impossible, though no complex NP intervenes.[13]

(41) *Nem tudtam,   hogy valamit is   elrontott.
        not   knew-1sg that   anything-acc destroyed-3sg
        'I did not know that he had destroyed anything'

(42) *Nem mondta meg, hogy valakit is   meghívott.
        no   said-3sg VM that   anybody-acc invited-3sg
        'He did not say that he had invited anybody'

A close examination of these data suggests that factivity plays an important role in licensing NPIs in negative polarity contexts.

To understand the phenomenon under discussion I would like to invoke Vendler's (1979) semantic categorization of verbs and also some analysis of factive islands in the generative framework.

Vendler's classification of propositional verbs (verbs that can take *that*-clause complements) contains three large sets: performative verbs (e.g. *mention, state, inform, admit, say, assert*), mental act verbs (e.g. *realize, assume, remember*) and mental state verbs (e.g. *know, think, believe*). While this is a useful classification for several purposes, it turns out that the whole domain of propositional verbs can be cross-classified into fully factive (like *know*), half-factive (like *tell*) and nonfactive (like *believe*) verbs. Vendler mentions three criteria which can be applied to decide which verb belongs to which class.

1. *the wh-criterion*: nonfactives reject wh-nominal complements
2. *the fact-criterion*: nonfactives reject the noun *fact*, and its kinship: *cause, result, outcome*, and *truth*.
3. *the adverb-criterion*: fully factives cannot cooccur with the set of adverbs consisting of *falsely, wrongly, incorrectly*, or simply with the denial of their *that*-clause complements.

---

[13] The element *meg-* in Hungarian is a perfective prefix which makes the verb *mond* 'say' behave like a factive verb. The gloss VM stands for verbal modifier.

Thus the picture is the following: fully factives pass criteria 1 and 2 but fail 3, non-factives fail 1 and 2 but pass 3, half-factives pass all three.

Let us see how these criteria can be applied:

Factives (e.g. *mention, know, find out*)

(43) (a) He mentioned/knew/found out where he lived.

    (b) He mentioned/knew/found out the fact that his uncle died.

    (c) *He falsely mentioned/knew/found out that his uncle died.

Half-factives (e.g. *tell, inform, report*)

(44) (a) He told me/informed me/reported who arrived late to the meeting.

    (b) He told me/informed me about/reported the fact that Jane moved out.

    (c) He falsely told me/informed me/reported that Jack stayed at home.

Nonfactives (e.g. *claim, assert, think, believe, assume*)

(45) (a) *He claimed/thought/believed where he went.

    (b) *He claimed/thought/assumed the fact that Mary failed her exam.

    (c) He wrongly/incorrectly thought/assumed that I slept home yesterday.

In Hungarian we can carry out the same tests, but some caution is necessary. The verb *mond* (meaning roughly *tell* or *say*) can be both factive and nonfactive, depending on other elements within the sentence. Namely, the emphatic pronoun *azt* indicates the nonfactive reading, while the perfective prefix *meg* enforces the factive reading:

(46) (a) Azt   mondta, hogy haza ment,   de hazudott.
       that   said-3sg that home went-3sg but lied-3sg
       'He told me that he had gone home but he lied'

    (b) *Megmondta, hogy haza ment,   de hazudott. (factive reading)
       said-3sg   that home went-3sg but lied-3sg
       'He told me that he had gone home but he lied'

Our claim is that licensing of VALA-NPIs by matrix negation is closely related to the question of factivity. If the matrix verb belongs to the class of nonfactives or half-factives, no problems arise. Negated factives, by contrast, do not license VALA-NPIs.

Analyses of factive islands have been numerous in the literature and the most influential of these (Kiparsky–Kiparsky 1971; Adams 1985; Zubizaretta 1982;

Rizzi 1990; Varlokosta 1994) work with the hypothesis that the CP complement of factive verbs is nominal and constitutes an island to extraction.

Instead of giving a nominal feature to factive complements we propose an analysis based on the semantic characteristics of factive verbs.[14] Factive verbs subcategorize for CP complements which are presupposed events. Such complements are always true, independently of the context. It is exactly this semantic difference between the truth value of sentential complements of factive and nonfactive verbs that is attested by Vendler's criteria. Although these observations are not accounted for in Progovac (1988; 1992a; 1992b), she remarks that none of the polarity clauses she examined has its truth value fixed. Her suggestion is that Op can have a +/− value, and that it is the minus value that is responsible for NPI-licensing and the plus value is given when the proposition is presupposed. Although her solution gives the right results, the +/− value of Op is rather stipulative.

Two alternative solutions are presented in Roussou (1992) and Giannakidou–Quer (1995a; 1995b). Roussou (1992) examines the (im)possibility of extraction from factive complements in Modern Greek, where they are introduced by a particular complementizer *pu* as opposed to the nonfactives which are introduced by *oti*. She argues that an empty operator in [spec, CP] together the [·definite] feature of the complementizer *pu* can successfully account for the extraction facts: both argument and adjunct extractions are blocked in factive complements.

So instead of saying that Progovac's polarity operator in Comp can have both a negative and a positive value, we could argue that the empty operator sitting in [spec, CP] of factive complements blocks NPI-licensing by matrix negation.

Giannakidou and Quer (1995a; 1995b) likewise remark that factive complements are opaque for long distance licensing of NPIs because they are presuppositional. They account for this fact by assuming that factive complements undergo QR at LF and end up in an IP-adjoined position. As a result of this LF movement, the c-command relation between matrix negation and the NPIs within the complement clause is destroyed and no licensing can take place.

No matter which of the two possible analyses we adopt, the negative operator will not play a role. This shows that the island effects in (20)–(22) can be successfully accounted for without any recourse to LF-movement of NPIs.

3.5. Conclusion from the previous facts

From the above mentioned arguments we can conclude that although characterizing SE-NPIs as anaphors that have to be bound in their governing category can cap-

---

[14] For some theoretical and empirical arguments against treating CP complements of factives as nominal see Rooryck (1991) and also Roussou (1992).

ture one of their most obvious distributional properties, it cannot account for several properties that should not be left out of consideration. Therefore, I propose that licensing of SE-NPIs and VALA-NPIs should be analysed as two different procedures in the spirit of Ladusaw (1992; 1994) and Giannakidou–Quer (1995a; 1995b). The inherent negativity of the former requires that they establish a spec-head relation with a head containing the feature +neg.

In particular, SE-NPIs are better analysed with the help of the Neg-Criterion, which must be satisfied at S-structure in Hungarian: SE-NPIs must move overtly to the specifier position of a functional projection whose head hosts the feature ·neg.[15] In Ladusaw's (1992; 1994) term this means that SE-NPIs can be licensed only strongly. Since this always implies movement, the local nature of SE-NPI licensing follows. Exceptions are those cases where the embedded clause does not constitute an independent tense domain and therefore movement of the negative element is possible, as we discussed in section **3.3.**

In the case of VALA-NPIs the situation is different. As we have seen, the claim that VALA-NPIs are anaphoric pronominals that can raise at LF faces empirical problems. These problems are further aggravated by facts from other languages. Raising does not seem to be supported by the data. Rather, VALA-NPIs behave like other indefinite expressions closed by the existential closure. They have no inherent negative feature (i.e. there is no morphological negation within VALA-NPIs), and therefore theoretically nothing forces them to move into the specifier of NegP. On minimalist assumptions, the closest c-commanding nonveridical operator unselectively binds them and licenses them. This means that VALA-NPIs never undergo LF raising out of the embedded CP.

4. Summary

Several empirical and theoretical facts have been put forward to support the claim that analysing NPIs not as quantifiers but as expressions associated with free variables which can be bound by a nonveridical operator is a desirable move towards understanding their behaviour. First of all, it enables us to get the correct interpretation when VALA-NPIs are in the embedded clause and are licensed by matrix negation. Secondly, this approach gives motivation for the adjacency requirement shown by SE-NPIs and the negative particle. Thirdly, it allows for a unified analysis of negative concord and licensing of NPIs in general: both SE-NPIs and VALA-NPIs are indefinites without any quantificational force of their own and two differ-

---

[15] Puskás (1992) also argues for satisfaction of the Neg-Criterion at S-structure in Hungarian.

ent licensing mechanisms give rise to different interpretations. In the case of
Hungarian, the motivation for the strong construal is the satisfaction of the Neg-
Criterion. Since VALA-NPIs are not subject to this criterion, they are licensed in
situ. This is a welcome result, since it becomes possible to capture the distribution-
al differences between SE-NPIs and VALA-NPIs on principled grounds, based on
morphological distinctions. While the assumption that VALA-NPIs, as opposed to
SE-NPIs, can raise at LF in order to extend their governing category is completely
ad hoc and is against the generally accepted view that Hungarian scopal relations
are reflected in the S-structure order, our proposal relies on an inherent feature of
these NPIs and is compatible with a minimalist claim that an element moves if and
only if morphological properties of that element are not otherwise satisfied.

References

Adams, M. 1985. Government of empty subjects in factive clausal complements. In: Linguistic
        Inquiry 16: 305–13.
Aoun, J. 1985. A grammar of anaphora. The MIT Press, Cambridge MA.
Aoun, J. 1986. Generalized binding. Foris, Dordrecht, Holland.
Baker, C.L. 1970. Double negatives. In: Linguistic Inquiry 1: 169–86.
Barwise, J.–Cooper, R. 1981. Generalized quantifiers and natural language. In: Linguistics and
        Philosophy 4: 159–219.
Belletti, A. 1990. Generalised verb movement. Rosenberg & Sellier, Torino.
Borkin, A. 1971. Polarity items in questions. In: CLS 7: 53–62.
Brody, M. 1990. Remarks on the order of elements in the Hungarian Focus Field. In: Kenesei, I. (ed.):
        Approaches to Hungarian, 3: 95–122. JATE, Szeged.
Brody, M. 1995. Focus and Checking Theory. In: Kenesei, I. (ed.): Approaches to Hungarian, 5:
        29–44. JATE, Szeged.
Carlson, G.N. 1980. Polarity *any* is existential. In: Linguistic Inquiry 11: 799–804.
Chomsky, N. 1981. Lectures on government and binding. Foris, Dordrecht, Holland.
Chomsky, N. 1986. Barriers. The MIT Press, Cambridge MA.
Chomsky, N. 1992. A Minimalist Program for Linguistic Theory. In: MIT Occasional Papers in
        Linguistics 1, Cambridge MA.
Giannakidou, A. 1994. The semantic licensing of NPIs and the Modern Greek subjunctive. In:
        Language and Cognition 4, University of Groningen.
Giannakidou, A.–Quer, J. 1995a. Two mechanisms for the licensing of negative indefinites. In:
        Gabriele, L. et al. (eds): Papers from the 6th Annual Meeting of the Formal Linguistics Society
        of Midamerica, IULC Publications, Bloomington IN.
Giannakidou, A.–Quer, J. 1995b. Long distance licensing of negative indefinites. Ms. University of
        Groningen and University of Utrecht.
Haegeman, L. 1992. Sentential negation in Italian and the Negative Criterion. In: Geneva Generative
        Papers 0, Université de Genève.

Haegeman, L Zanuttini, R. 1991. Negative heads and the neg criterion. In: The Linguistic Review 8: 233 51.

Haspelmath, M. 1993. A typological study of indefinite pronouns. Ph.D. dissertation, Freie Universität Berlin.

Higginbotham, J. 1985. On semantics. In: Linguistic Inquiry 16: 547–93.

Horn, L.R. 1978. Remarks on neg-raising. In: Cole, P. (ed.): Syntax and Semantics, 9: 129–220. Academic Press, San Diego

Horn, L.R. 1989. A natural history of negation. The University of Chicago Press, Chicago IL.

Hunyadi, L. 1981. Remarks on the syntax and semantics of topic and focus in Hungarian. In: Acta Linguistica Hungarica 31: 107 36.

Karttunen, L. 1977. Syntax and semantics of questions. In: Linguistics and Philosophy 1: 3 44.

Kenesei, I. 1992. On Hungarian complementizers. In: Kenesei, I. Pléh, Cs. (eds): Approaches to Hungarian, 4: 37 50. JATE, Szeged.

Kenesei, I. to appear. Finno-Ugric. In: Vincent, N. (ed.): Complementation in the languages of Europe. Mouton de Gruyter, New York.

Kiparsky, P. Kiparsky, C. 1971. Fact. Reprinted in Steinberg D.D. Jakobovits, L.A. (eds): Semantics, Cambridge University Press, Cambridge.

Ladusaw, W.A. 1979. Logical Form and conditions on grammaticality. In: Linguistics and Philosophy 6: 373 92.

Ladusaw, W.A. 1980. Polarity sensitivity as inherent scope relations. Garland, New York and London.

Ladusaw, W.A. 1992. Expressing negation. In: SALT II Proceedings. The Ohio State University, Ohio.

Ladusaw, W.A. 1994. Thetic & categorial, stage & individual, weak & strong. In: SALT IV Proceedings. University of Rochester, NY.

Laka, I. 1990. Negation in syntax: on the nature of functional categories and projections. Ph.D. dissertation, MIT, Cambridge MA.

Lawler, J.M. 1971. Any questions? In: CLS 7: 163 73.

Linebarger, M.C. 1981. The grammar of negative polarity. The Indiana University Linguistics Club, Bloomington IN.

Piñón, C.J. 1992. Heads in the Focus Field. In: Kenesei, I. – Pléh, Cs. (eds): Approaches to Hungarian, 4: 99 122. JATE, Szeged.

Piñón, C.J. 1993. SigmaP and Hungarian. In: Mead, J. (ed.): The Proceedings of WCCFL 11: 388 404. CSLI, Stanford.

Progovac, L. 1988. A binding approach to polarity sensitivity. Ph.D. dissertation. University of Southern California, Los Angeles CA.

Progovac, L. 1991. Polarity in Serbo-Croatian: anaphoric NPIs and pronominal PPIs. In: Linguistic Inquiry 22: 567 72.

Progovac, L. 1992a. Negative polarity: a semantico-syntactic account. In: Lingua 86: 271 99.

Progovac, L. 1992b. Non negative polarity licensing must involve Comp. In: Linguistic Inquiry 23: 341 7.

Puskás, G. 1992. The Negative Criterion in Hungarian. Ms. Université de Genève.

Rizzi, L. 1990. Relativized minimality. The MIT Press, Cambridge MA.

Rooryck, J. 1991. Negative and factive islands revisited. Ms. Indiana University.

Ross, J.R. 1967. Constraints on variables in syntax. Ph.D. dissertation, MIT, Cambridge MA.

Roussou, A. 1992. Factive complements and wh-movement in Modern Greek. In: UCL Working Papers in Linguistics Vol. 4.

Szabolcsi, A. 1994. Szintaxis, szemantika, nyelvi érintkezés [Syntax, semantics and language contact]
    Talk given in Budapest at the Telegdi conference, 1994.
Varlokosta, S. 1994. Issues on Modern Greek sentential complementation. Ph.D. dissertation,
    University of Pennsylvania.
Vendler, Z. 1979. Telling the facts. In: Kiefer F. · Searle, J. (eds): Speech act theory and pragmatics.
    Reidel, Dordrecht.
Zanuttini, R. 1991. Syntactic properties of sentential negation: a comparative study of Romance lan-
    guages. Ph.D. dissertation, University of Pennsylvania.
Zubizaretta, M. 1982. Theoretical implications of subject extraction in Portuguese. In: The Linguistic
    Review 2: 79 96.

Address of the author: Ildikó Tóth
                              Grammaticamodellen
                              University of Tilburg (KUB)
                              Postbus 90153
                              5000 LE Tilburg, The Netherlands
                              e-mail: i.toth@kub.nl

INFORMATION FOR CONTRIBUTORS

Please provide typewritten double-spaced manuscripts with wide margins (50 characters per line, 25 lines per page). Type on one side of the paper only and number all the pages consecutively. The following information should be provided on the first page: the title of the article (or, in case of reviews, the title of the book under discussion, the name of the publisher, the place and date of publication as well as the number of pages), the author's full name and address and an abbreviated title of the paper (not exceeding 45 characters including spaces). Only original papers will be published and a copy of the Publishing Agreement will be sent to the authors of papers accepted for publication. Manuscripts will be processed only after receiving the signed copy of the agreement. Authors are requested to send two hard copies of their manuscript + a floppy disk under DOS.

The manuscript should be accompanied by an abstract of about 100 words. It should be typed on a separate sheet of paper. Tables, diagrams and illustrations (with the author's name and an Arabic number) should be presented on separate sheets. Captions should be typed on a separate sheet and placed at the end of the manuscript. Footnotes should be typed double-spaced on a separate sheet at the end of the article. All language examples (and only these) are to be italicized (single underlining). Citations in the text should be enclosed in double quotation marks (" " in case of a paper written in English, „ " in German and « » in French).

Reference to a publication should be made by the name of the author, the year of publication and, when necessary, by page numbers in the following ways:

. . . as described by Schmidt (1967) . . .
. . . as referred to by Hoover (1967, 56 78; 1976, 43).
. . . mentioned by several authors (Elgar 1978, 77; Williams 1981, 154 6) . . .

An alphabetically arranged list of references should be presented at the end of the article as follows:

Bárczi, G. 1958a. Magyar hangtörténet [The history of Hungarian sounds]. Akadémiai Kiadó, Budapest.

Bárczi, G. 1958b. A szótövek [Word stems]. Akadémiai Kiadó, Budapest.

Lakoff, G. - Peters, P.S. 1969. Phrasal conjunction and symmetric predicates. In: Reibel, D. - Schane, S. (eds): Modern studies in English, 113 41. Prentice-Hall, Englewood Cliffs.

Leben, W.R. 1980. A metrical analysis of length. In: Linguistic Inquiry 11: 497 509.

Ross, J.R. 1967. Constraints on variables in syntax. Ph.D. dissertation. MIT, Cambridge MA.

For marking subsections decimal notation should be applied. Do not use more than four digits if possible.

Examples within the text should be marked in italics. Meanings are to be rendered between inverted commas (' '). If glosses are given morpheme by morpheme, the initial letter of the gloss should be placed exactly below that of the example. Grammatical morphemes can be abbreviated in small case letters connected to the stem or the other morphemes by a hyphen. No period should be applied in such abbreviation. For example:

(1) (a)   A   sólymaid   elszálltak
          the falcon-gen-pl-2sg away-flew-3pl
          'Your falcons have flown away.'

Examples can be referred to in the text as (1a), (1a d), etc.

One proof will be sent to the author. Please read it carefully and return it by air mail to the editor within one week.

ACTA LINGUISTICA HUNGARICA

VOLUME 46, NUMBERS 1–2, 1999

CONTENTS

23.

Acta Linguistica Hungarica

AN INTERNATIONAL JOURNAL OF LINGUISTICS

Akadémiai Kiadó
Budapest

Kluwer Academic Publishers
Dordrecht/Boston/London

ACTA LINGUISTICA HUNGARICA

AN INTERNATIONAL JOURNAL OF LINGUISTICS

ACTA LINGUISTICA HUNGARICA

AN INTERNATIONAL JOURNAL OF LINGUISTICS

GUEST EDITORS' NOTE

For most workers of the scientific order, it is especially enjoyable and rewarding to study topics that unite scientific interest with social, political and/or practical utility. Our conviction is that the study of the Romani language and culture represents such a domain of research for the international scientific community. This challenging and fascinating field, however, has remained largely unexplored so far.

The present issue of ALH, dedicated to the Romani language and culture, would like to add further material to the growing body of academic research on Romani. Topics of the papers in the volume embrace several domains including linguistics proper, as well as language-related interdisciplinary fields: descriptive linguistics, historical and contact linguistics, and issues of sociolinguistics and linguistic anthropology.

As to the first domain mentioned above: the descriptive study of Romani may be challenging for linguists in more than one respect. The challenge may partly emerge from the fact that, in spite of international efforts, Romani so far exists mainly as a non-standard spoken language, in a range of regional dialects. As a consequence, linguists studying Romani (who themselves are mostly nonnative speakers of one or another Romani dialect) can rely neither on a homogenized, elaborated and standardized (written) code, nor on their own linguistic intuition of the native speaker. However, as so often in history, virtue may arise out of necessity: inevitably, the analytical approach to Romani has to be more sensitive to data of spoken language, more aware of the historical factors and the social and communicative constraints shaping language, and more conscious of the heterogeneity of linguistic systems than current "mainstream" linguistics usually is.

At the same time, the varieties of Romani have certain typological and grammatical features that are only scarcely represented in the languages most studied in contemporary Western and American linguistics. Thus, aspects of Romani may be of great interest to theoretically oriented linguists as well.

The paper by Yaron Matras in this volume, **Subject clitics in Sinti** represents such a complex approach to the description of a typologically peculiar grammatical feature of one variety of Romani (the Sinti dialect, mainly spoken in German-speaking regions). In analyzing the problem of subject clitics, the paper discusses formal, pragmatic and historical arguments as well, thus arriving at a new interpretation of

this grammatical phenomenon. The author also gives an outlook on the possible phases of the formation of the systems of subject clitics in Sinti.

The following three papers in the volume demonstrate possibilities the study of the Romani may offer for contact linguistics—both in synchronic and diachronic respects. As one of the authors of the present volume remarked: "... within Europe, there is no other language that underwent such different influences as Romani", particular dialects having been influenced by particular European languages (see Boretzky, this volume, 169). Norbert Boretzky's paper **Grammatical interference in Romani: Loan formations for foreign categories** gives a detailed survey of the process and the ways different contact languages shaped selected grammatical aspects of different varieties of Romani (with special emphasis on the formation of the verbal system). Birgit Igla's paper, **Disturbances and innovations in the case system in Bulgarian Romani dialects** is in several ways complementary to the former one, as it discusses the factors contributing to the formation of the case system common to Romani dialects in Europe, and its peculiarities in the Romani dialects as spoken in contemporary Bulgaria. Endre Tálos' paper **Etymologica Zingarica**, while tracing the origin and history of selected lexical items, may shed light, from a different perspective, on the formation of varieties of Romani through the history of language contacts: the paper is a collection of etymologies of a number of lexical entries that, in the available classical and recent etymological dictionaries of Romani, have been registered as having unknown or uncertain origin. Hopefully, the stock of etymologies presented in this paper will represent a contribution to future etymological dictionaries.

Shifting perspective in our approach to language, the last three papers of the volume are dedicated to the study of Romani in its social contexts. Ethnographic studies of language and language use emphasize that the world's speech communities may greatly differ in a range of basic features related to the social use of language. Such differential features of a community's "speech economy" may include, for example, the value and importance attached to different forms of speaking, or even to speaking in general, the extension and elaborateness of the stylistic repertoire available for speakers, the overall quantity of speech, among others (see e.g. Hymes 1972). Anthropological descriptions of Romani (see e.g. Stewart 1997) suggest that, at least for Vlach Gypsy communities in Europe, language represents the primary value of culture, and the individual's ability to speak in culture-specific ways (*romanes* 'in the Gypsy way') is considered the basic mark of identity.

Katalin Kovalcsik's paper (**Aspects of language ideology in a Transylvanian Vlach Gypsy community**) analyses the "language ideology" revealed in a traditional Romani song sung by a speaker of a Vlach Gypsy community in Transylvania. In this largely improvised song the singer develops his arguments proving the unique

purity and authenticity of his own dialect, contrasted with other varieties of Romani and other languages spoken around him.

Some of the earlier anthropological analyses focusing on culturally specific features of speaking in traditional communities also suggest that for the Roma, language has a high practical value as well (e.g. Kaprow 1982; Stewart 1997). According to these studies, for example, it is mainly through the use of different verbal strategies that the Roma try to achieve control in intra-group, as well as in interethnic, communication. Perhaps the most important one of these basic interactional verbal strategies is **teasing**, a way of speaking that is extremely widespread in traditional Gypsy communities, and is used in a variety of situations and genres. The paper by Zita Réger (**Teasing in the linguistic socialization of Gypsy children in Hungary**) focuses on the use of the verbal strategy of teasing in adult–child communication, and analyzes some of its quantitative and qualitative aspects in the linguistic socialization of babies and young children growing up in a traditional Romani-speaking community in Hungary.

Beyond the fields mentioned above, linguists' contribution is much needed in a few related, more practically oriented domains such as the survey of the linguistic status quo of present-day Romani-speaking communities, language maintenance, and language planning, language rights and related educational problems. The last paper of the volume (Victor A. Friedman's **The Romani language in the Republic of Macedonia: Status, usage, and sociolinguistic perspectives**) analyses the linguistic situation of Romani in the multilingual context of Macedonia around these topics. It also exemplifies some of the successes and pitfalls of recent efforts for Romani language planning in general. The issue of ALH dedicated to research on Romani therefore ends with a paper that opens perspectives on the burning question of 'What can be done' in the interest of Romani and its speakers. Our hope is that aspects of the model provided here will be used by linguists and policy makers and perhaps by the members of the respective communities as well, in their efforts to develop the Romani language and literacy in regional as well as international respects.

References

Hymes, D. 1972. Foundations in sociolinguistics. University of Pennsylvania Press, Philadelphia.
Kaprow, M. 1982. Resisting respectability: Gypsies in Saragosa. Urban Anthropology, 3-4.
Stewart, M. 1997. Time of the Gypsies. Westview Press, Boulder CO.

Zita Réger
Institute for Linguistics
Hungarian Academy of Sciences
Budapest

Katalin Kovalcsik
Institute for Musicology
Hungarian Academy of Sciences
Budapest

SUBJECT CLITICS IN SINTI

YARON MATRAS

Abstract

Holzinger (1993) has recently defined subject clitics in Sinti as markers of high reference continuity. Structural aspects of subject clitic distribution in Sinti lead me to a new interpretation of the functions of this referential device, especially as regards its role in the typology of the dialect. I argue that clitics are employed primarily in constructions in which verb–subject order is obligatory. There are two main patterns for such constructions in the language. The first, in *te*-constructions, is inherited. The second, verb–subject inversion, is largely an outcome of syntactic convergence with German. The specialization of clitics for certain constructions is taken as an indication of their beginning retreat in the dialect.

1. Adjectival subject agreement in Romani

Verbs in Romani are generally inflected for person and number, while adjectival agreement is marked for gender and number. There are however two types of adjectival subject agreement with finite verbs in the language:

(a) The finite use of participles. In some Balkan dialects, this is the only way of forming the simple past tense of some intransitive verbs, especially of verbs indicating motion or change of state, as well as passives and inchoatives (*gelo/geli* 'he/she went', *arakhadžilo/arakhadžili* 'he/she was found'). Active past participles are always restricted to the third person. In Lovari and other Vlach dialects they exist alongside inflected simple past tense forms of the same verbs (*gelas* 'he/she went', *arakhadžilas* 'he/she was found'). Here, the active past participle has acquired an evidential meaning. Its distribution in discourse is determined by the pragmatics of the interaction. It is mainly used to stress non-confirmative aspects of the proposition such as surprise, disbelief, unexpectedness, or irony (see Matras 1995a).

(b) Subject clitics. Like the active past participles, they appear in Romani only in the third person. Sampson (1926, 161) regards the nominative enclitic pronominal forms *-lo* (m), *-li* (f), *-le* (pl) as derived from the Old Indic pronominal stem *ta-* and

so as part of the same historical paradigm as the oblique forms of the third person pronouns *les, la, le(n)*. This implies that they are older than, and in most environments have been replaced by the current nominative pronouns, which show dialect variants in *ov/oj/on, jov/joj/jon, vov/voj/von*. Such an analysis is plausible, as it is the nominative form of the third person pronoun which is universally most often subjected to structural renewal. Boretzky (1994, 63) however suggests that we might be dealing with a later development of subject clitics, one which copies the oblique forms.

Subject clitics appear with lexical verbs in varieties of Sinti–Manuš and in the Central dialects of Hungary, while in the Vlach and Balkan dialects they are restricted to existential constructions, where they either supplement or substitute for the existential verb *si/hi* 'is', *naj* 'is not' (cf. Boretzky 1995, 32–3): *vo sî lo phuro* 'he is old', *kaj lo?* 'where (is) he?', *eta lo* 'there he (is)' (see also Boretzky 1994, 62–4). A similar distribution appears in the English and Welsh dialects (Smart–Crofton 1875, Sampson 1926).

Both types of adjectival agreement patterns with the subjects of finite verbs, the active participle and the subject clitic, appear to have been more productive in earlier stages of the language. In the related Central languages of subcontinental Indo-Aryan, such as Hindi, the active participle still forms the simple past tense of all verbs, transitive and intransitive; agreement in transitive constructions is generally with the direct object, a feature of ergativity in those languages. In Romani, which is not ergative, participle agreement is only possible with the subject, and is therefore restricted to verbs of the unaccusative type, where the current state of the subject allows inference about the underlying process or action. But the emergence of a new inflected past tense paradigm with personal affixes leads to a competition of forms even here, and the 'finite' participle generally becomes dispensable. It disappears entirely in the western, northern and central Romani dialects (Sinti, Baltic, North-Russian, Slovak). In the Vlach dialects its specialization for evidentiality or non-confirmation copies a distinction which is made in the past tense systems of some of the congruent languages of the Balkans (cf. Friedman 1986; see discussion in Matras 1995a). This process may be regarded as a case of contact-induced grammatical recycling, or exaptation, as Lass (1990) refers to the opportunistic exploitation of a grammatical item the use of which is only indirectly connected to its original meaning.

Much like the active participle, subject clitics with adjectival agreement are retreating in the language as a whole. Thus, while Boretzky (1994, 62–4) cites several examples for the use of clitics in existential constructions in the Kelderaš dialect, my own observations on a Kelderaš/Lovari contact variety (cf. Matras 1994) show only one single case of clitic usage, *kaj-lo/la* 'where (is) he/she', where

the clitic itself initiates a predication, and the verbal copula is missing. Such specialization for deictic locatives is a general tendency in certain types of non-verbal predications, as is the appearance of pronominal copulas in equatial predications (see Hengeveld 1992, 208–12). One might draw a connection between the restriction of such pronominal clitic predications to the deictic center and the tendency to restrict active participles to deixis-related functions, as seen in the evidential usage of such participles with unaccusative verbs. In both cases, the adjectival ending identifies the subject on the basis of a highly accessible domain of reference.

This paper deals with the distribution of subject clitics in the Sinti dialect spoken in Germany, where they are much more frequent and productive than in the Vlach dialects. I examine data from two corpuses: The transcribed oral narratives presented by Holzinger (1993, 318–26) in an appendix to his grammar of the Sinti dialect (henceforth 'Holzinger corpus'), and a recent translation of the Mark Gospel into Sinti, published in Florshain, Germany, in 1994 (henceforth 'Mark corpus'). In Sinti, as well as in the closely related variety of Manuš described by Valet (1991), subject clitics are not confined to deictic or situational predications, but there is agreement in the literature that they express more accessible subject referents. Thus Valet (1991, 121) refers to clitics as the "normal forms" of the pronoun, while the full pronouns *jop, joj, jon* are described as "emphatic forms". Similarly, Holzinger (1993, 290–308) defines subject clitics as highly continuous referential markers.

While both the distribution of the forms in the data, and the general pattern of subject clitic usage and its retreat in the language as a whole (in comparison with other dialects) support this view on a connection between subject clitics and subject accessibility, formal–structural aspects of subject clitic distribution in Sinti, if reconsidered, may lead to a new interpretation of the functions of this referential device, especially as regards its role in the typology of the dialect. I argue below that subject clitics in Sinti are employed primarily in constructions in which verb–subject order is obligatory. There are two main patterns for such constructions in the language. The first is inherited, and appears in other dialects of Romani as well. It includes embedded constructions initialized by the non-factual subordinating conjunction *te*, which, as a marker of modality, is always immediately followed by the verb: modal complements, manipulative complements, and purpose clauses. The second VS pattern is a result of language contact with German, and involves verb–subject inversion in constructions in which the first position in the sentence is occupied by an ad-verbal element other than the subject (a direct or indirect object, an adverb or an adverbial clause), or in connective constructions where the finite verb appears in initial position (for the latter in spoken German cf. Rehbein 1992, 544–9). Although VS constructions are frequent in other dialects of Romani as well

(cf. Matras 1995b), their formalization in some environments in Sinti is clearly a result of ongoing convergence with the word order patterns of German.

In the case of *te*-constructions, Sinti is unique among the dialects of Romani in largely avoiding the use of free pronouns. At the same time subject clitics in non-factual *te*-subordinations allow for the retention of a pattern of linear ordering very different from German, and so in a way they make it easier to 'resist' syntactic convergence. In the case of verb–subject inversions, the effect of clitic retention with respect to Sinti–German convergence phenomena appears to be just the opposite: Clitics are exploited following rules that are compatible with the syntax of verb–subject inversion in German. Thus, while the mere occurrence of clitics cannot in itself be directly associated with syntactic borrowing, clitics nevertheless assume a crucial role in regulating the special dichotomy of syntactic autonomy versus adaptation in a language contact situation.

The frequency of clitic occurrences in VS constructions of these types suggests that formal aspects of subject clitic distribution are at least as relevant, and perhaps even more so, than the discourse-functional or pragmatic features associated with their appearance (cf. Holzinger 1993). Nevertheless, high topic continuity being a feature of many VS constructions due to their connected character, the correlation observed between the appearance of subject clitics and high referential continuity or subject accessibility is not particularly surprising.

2. Clitics and participant tracking in Sinti

In the more conservative varieties of the Sinti–Manuš group of dialects, clitics show considerable variation as regards their status on the hierarchy of reference devices, as well as referential distance. In the dialect described by Valet (1991, 130), all forms are based on the *l*-stem and they all follow the verb:

(1)   Har vejan   le paš leste,   o   biboldo dikas   lo ku kova, dejas   lo lende
       how came-3pl cl near him-loc the Jew   looked-3sg cl at that   gave-3sg cl them-loc

       krat i pisla love.
       just a little money

       'Just as they came to him, the Jew looked at it, he gave them just a little money'

(2)   O   puredar čavo čivas   les   an peskri posita un bistras   les   lo dren.
       the older   boy put-3sg it-acc in his   pocket and forgot-3sg it-acc cl inside
       'The older boy put it in his pocket and forgot it inside'

Clitics are shown in (1)–(2) to be a productive device for participant tracking in various syntactic environments: adverbial clauses, following topicalized subjects, in constructions involving inverted subject pronouns, and for reference to continuous subjects in paratactic chains. A special feature of clitics in the short text presented by Valet is their employment for reference to topicalized subjects, as in the second clause in (1). Such usage does not appear in either the Holzinger or the Mark corpuses. In Vlach and Balkan Romani, it is common however with object pronouns, which are etymologically related to the subject clitics. Consider the following example from a Cerhari-Kelderaš (Vlach) dialect from Transylvania (Kovalcsik–Tálos 1991, 114):

(3)   The, le dò raklèn   xutjildàha-le   the phanglàh-le   ke le   grastèhki pouri.
        and the two girls-acc caught-3sg-them and tied-3sg-them to the   horse-gen tail
        'He took the two girls and tied them to the horse's tail'

Subject clitics cannot occur in such positions in Vlach, and they are confined either to the deictic center or to equatial predications. The fact that subject clitics are not used to support topicalization in the Sinti corpuses considered here either, although they do appear there in other constructions, might be taken as a first indication of their beginning retreat in Sinti. The constraints on their appearance as 'floating' clitics are partly represented in the morphophonological reduction of the forms. Holzinger (1993, 292–98) distinguishes between 'enclitic pronouns' (*lo/li/le*), which usually follow an object or an adverb, and 'verb suffixes', also referred to as 'subject suffixes' (*-o/-i/-e*), which appear without an object and are usually directly attached to the inflected verb:

(4)   Vajaso   pal   mende.
        came-3sg-cl   after us-loc

        Rodehs   men   lo   (Text 1: 32–3)
        searched-3sg us-acc cl

        'He came after us.
        He was looking for us'

Holzinger defines these structures as two functionally distinct referential devices. Enclitic pronouns are placed higher than 'suffixes' on the scale of topic continuity and thematic coherence. Holzinger's continuity scale is based on an evaluation of referential distance measured by the number of sentences between the present and the preceding reference to the same topic, as well as of the frequency of occurrence

(a) with subject switches, (b) in paragraph-initial position, and (c) in foregrounded positions. But the scores obtained for the two forms of clitics differ only slightly.

The argument in favour of a discourse-functional distinction between the two forms of clitics is further weakened by their structural distribution. According to Holzinger, "enclitic pronouns only rarely appear in texts" (296). The positioning of the *l*-form after an object or adverb obviously prevents it from appearing in the first or foregrounded position in the sentence. Thus it is less likely to indicate a shift in topic or thematic discontinuity than the 'subject suffix'. Furthermore, in *te*-embeddings, where the verb immediately follows the conjunction, subjects following the verb will generally be expressed by the 'suffix'. In complement clauses involving manipulation, the 'suffix' and not the 'enclitic pronoun' or *l*-form will therefore be used to indicate a subject switch. As a result, the 'suffix' ranks lower on the topic continuity scale than the *l*-form. An inherent link between tighter continuity and the *l*-forms would, however, violate an iconicity principle, which seems to apply elsewhere along the continuity hierarchy of referential devices in Sinti, with parallels in other languages as well, and according to which structurally more complex forms are generally employed when more effort is needed to track down the referent. Although not an imperative, this tendency is often regarded as a universal of language, representing universals of cognition and communication (cf. Givón 1995, 50–66).

The structural features show a tendency towards complementary distribution of the two forms of clitics which justifies their treatment as two realizations of the same reference device. Clitics that are attached to the inflected form of the verb tend to assume the form of adjectival suffixes, while those that remain distant from the verb retain the full, consonantal stem in *l*-. I therefore refer to these two forms of the subject clitic as 'short' and 'long' forms respectively. This view is partly supported by the data from the Mark corpus, where likewise the vowel suffixes always follow an inflected verb, while *l*-forms display a tendency (over 50%) to follow conjunctions, reflexive and object pronouns, and the existential verb.

The two forms of clitics assume adjoined positions on Holzinger's hierarchy of continuity. They are immediately preceded by the top position on the hierarchy (most continuous), which is occupied by the verb with no overt marking of the subject/topic other than the person/number inflection (pro-drop), and they are followed by the free personal pronoun. Further lower positions on the hierarchy show demonstratives, followed by noun phrases with different degrees of definiteness (cf. Holzinger 1993, 308). Clitics and pronouns are thus competing devices for overt anaphoric reference. But what is the nature of the opposition between them?

Free pronouns according to Holzinger (298) are much less frequent in Sinti than they are in German. This can be explained by the availability of other referential devices, including pro-drop. Although pronouns follow short-form clitics on

Holzinger's continuity hierarchy, there are no significant measurable differences between them as far as referential distance is concerned. Rather, their positioning on the scale is determined by the higher frequency of free pronouns in constructions involving a subject switch (different-subject constructions). Here Holzinger notes a correlation with word order patterns. The preverbal position is reserved for the thematic accentuation of participants and so it is used to direct the hearer's attention to a topic. Free subject pronouns are said to appear preverbally in most cases (299–300), and in fact a careful look at the Holzinger corpus revealed not one single postverbal occurrence of a free subject pronoun in the third person. On the other hand, enclitic pronouns (long-form clitics) cannot appear in preverbal position according to Holzinger (296), and short-form clitics, being suffixes to the inflected verb, naturally only occur in postverbal position.

This raises the question whether subject clitics and free subject pronouns might indeed be functionally equivalent, but appear in complementary distribution, free pronouns occurring in preverbal position, clitics in postverbal position. This is tentatively supported by the little data provided in Valet's (1991) description of Manuš where likewise full pronouns always occur in preverbal, clitics in postverbal position. In the Mark corpus, however, this is not entirely so, and we have occasional appearances of free subject pronouns in postverbal position. But there are only 56 cases in the entire corpus, compared with numerous (for a manual evaluation, indeed countless) instances of free pronouns occurring in preverbal position. It will be shown below that there are no obvious pragmatic–textual constraints on their occurrence. Long-form clitics, on the other hand, may appear in the Mark corpus in preverbal position. It is important to note, however, that the language of the Mark corpus tends to replicate German word order patterns much more consistently than that of the Holzinger corpus. This includes the placement of the finite verb in final position in subordinated clauses. Statistically, this means that anaphoric devices are more likely to appear in preverbal position here than in the Holzinger corpus, and indeed all instances of long clitics in preverbal position in the Mark corpus (altogether only 28% of the total occurrences of long clitics) involve clitics attached to the conjunction in a subordinate clause.

To summarize: The Holzinger corpus shows complementary distribution of short and long clitics, but also of clitics and free pronouns. The Mark corpus supports this general tendency, but it allows for competition of short clitics, long clitics, and free ponouns in postverbal position, as well as for a competition of long clitics and free pronouns in preverbal position in subordinate clauses. In the following sections I examine these distribution patterns more closely, paying special attention to the use of anaphoric reference devices in several types of complex constructions involving clause and referent integration.

3. Distribution in the Holzinger corpus

The two texts presented by Holzinger (1993, 318–26) constitute a limited corpus, but they nevertheless display some basic distributional tendencies which apply to the more extensive Mark corpus as well. The salient feature in the Holzinger corpus is the lack of any free pronouns in postverbal position, and so the lack of competition between anaphoric devices here. The distribution of clitics in the Holzinger corpus is thus rather straightforward.

Short clitics rarely appear in adverbial subordinations. Their occurrence here, however, renders a distinct word order type with the subject clitic assuming the final position in the clause:

(5)   ... deš   krone, pantš krone, jenachdem har   but   lehso.   (Text 1: 9)
         ten   crown five   crown   depending   how   much took-3sg-cl
         '... ten crowns, five crowns, depending on how much he was getting'

Conditionals, though formally *te*-constructions, often serve in the Holzinger corpus as temporal adverbial clauses, due to convergence with German *wenn*-clauses:

(6)   Und ko   te   tikehso,   dann ...   (Text 1: 22)
         and that if   saw-3sg-cl then
         'And when he saw it, then ...'

These subordinations are complemented by the class of *te*-clauses. *Te* in Romani dialects introduces non-factual complements, such as those of modality and manipulation, as well as purpose clauses. It generally corresponds to the respective non-factual complementizers of the other Balkan languages, and its distribution in Sinti is one of the most obvious traces of the syntactic Balkanization which early Romani had undergone before the divergence of its various dialects. Although Sinti has partly generalized the 3sg form of the verb as an infinitive-like structure, while Vlach and Balkan Romani generally lack an infinitive, even this 'new' infinitive (cf. Boretzky 1996) in Sinti is introduced by *te*. The common modality feature of all constructions with *te*, including conditionals, supports an integrated view of such constructions as *te*-enhancements to main or core clauses (cf. Matras 1994, 224–36).

While the modality feature is lost in cases like (6) as a result of convergence with German, the structural condition on *te*-enhancements prevails, and *te* always immediately precedes the verb. In modality clauses, subject clitics are a convenient device for anaphoric reference which suits the particular structural features of Romani modality. Modality is marked by the choice of the conjunction itself, and

so no insertion is permitted between the conjunction and the verb in the embedded clause. Short clitics, rather than (emphatic) free pronouns, assume the function of back-reference in same-subject constructions, as well as in cases of object raising:

(7) Dann sikevaues les, har te ulevelo   (Text 1: 58)
     then showed-1sg him how comp drive-3sg-cl
     'Then I showed him how to drive'

Alongside adverbial subordinations and complement clauses, we find the majority of occurrences of short clitics in subject inversion constructions. Sinti follows the rule for German word order, confining the finite verb in simple, thematic declarative sentences to the second position. The subject, represented in our cases by a subject clitic, follows the verb if the first sentence position is occupied by an adverbial clause, an adverb, or an object:

(8) Har dajam les i grai, hiso demfig, dšineh   (Text 1: 31)
     how gave-1pl him a horse was-cl broken-winded know-2sg
     'When we gave him a horse, it was broken-winded, you know'

(9) noch nicht mol ko tserdehso   (Text 1: 15)
     not even that pulled-3sg-cl
     'it didn't even pull that one'

In addition to this formally triggered inversion, Sinti also exhibits an inversion based on clause connectedness, or connective inversion. While absent in written German, connective inversion is a common feature of standard spoken German (cf. Rehbein 1992, 544–9), as well as regional (northern and central) varieties of German, Yiddish (cf. Reershemius 1997, 157–88), and Jewish dialects in southern Germany (cf. Matras 1991, 278). It signals thematic supplementation, often resulting in a consequential interpretation, and is often employed as a connecting strategy in serial chaining in narratives (Rehbein 1992). The Sinti constructions in (10)–(11), as well as in (4) above, thus conform to spoken German word order patterns:

(10) Dšajas paš miro kamlo dadeste, rakedas mit leha.
      went-3sg by my late father-loc spoke-3sg with him-soc

      Phenaso: 'Hoi pheneh tu?'   (Text 1: 44–5)
      said-3sg-cl what say-2sg you

      'He came to my late father, he spoke to him.
      He said: 'What do you say?''

(11)   Miro kamlo dad   ap koi rig,   me ap kai rig.
        my   late   father on that side   I on   this side

        Pandelo   kote fest, me kate fest.   (Text 2: 21–2)
        tie-3sg-cl there tight, I   here tight

        'My late father on that side, me on this side.
        He ties it there, I (tie it) here'

Now in Vlach and Balkan Romani, connective inversion of a similar type may occur too (cf. Matras 1995b). But while it is possible that Sinti did not actually acquire the construction due to contact with German, its presence is nevertheless likely to have been reinforced by the ongoing convergence of word order patterns with those of (spoken) German. The employment of the clitic here, as in the formally triggered inversion in (8)–(9), may thus be regarded as an opportunistic exploitation of a reference device which typically follows the verb, and therefore is suitable for replicating the verb–subject sequence now adopted.

The distribution of long clitics in the Holzinger corpus is essentially similar, except that they usually follow an object or a reflexive pronoun (see Table 1). Interestingly, two cases of clitic doubling occur, showing traces of the floating character of the historical subject clitic in its long form:

(12)   Phenaso   tšimone   lo, haieveh   (Text 2: 40)
        said-3sg-cl something cl   understand-2sg
        'He said something, you understand?'

(13)   Dann mangehso   miro kamlo dadester   lo, te   mukelo   man
        then asked-3sg-cl my   late   father-abl   cl   comp let-3sg-cl me-acc

        doch dui voxe   paš leste   (Text 1: 66)
        part two weeks by   him-loc

        'Then he asked my late father to leave me with him for two weeks'

Table 1 summarizes the distribution of short and long clitics in various syntactic constructions in the Holzinger corpus.

Free pronouns were found in none of the syntactic constructions considered here, a result which is expected given their confinement to preverbal positions. It can be seen that the great majority of clitics appear in inversions, which copy, or at least are congruent with, both the formal and communicative rules (i.e. interclausal connectedness) for verb–subject inversion in (spoken) German. Ranking second in

Table 1

Holzinger corpus: Distribution of clitics in syntactic constructions

	Short clitics	Long clitics
Total	21 (100%)	9 (100%)
Adverbial subord.	2 (10%)	
Follows obj. pron.		1 (11%)
te-compl./modal compl.	4 (19%)	
te-adverbial clauses	1 (5%)	
Attaches to verb		1 (11%)
verb–subject inversion	14 (66%)	
Follows refl. pron.		2 (22%)
Follows obj. pron.		3 (33%)
'Doubling'		2 (22%)

the frequency of distribution across syntactic constructions we find clitics in complement clauses, joined to some extent by adverbial clauses, partly due to the extended meaning of *te* copying German *wenn*. Unlike German, the word order here is verb–subject.

This brings us to a first interpretation of the typological role which subject clitics assume in the dialect, and which, I argue, must be seen in connection with language contact and the ongoing processes of syntactic convergence with German. Subject clitics help reduce structural friction between inherited and borrowed syntactic patterns. In complements, they are employed as a reduced type of anaphoric reference, which downplays the conflict with German word order patterns. In inversions, they allow to compromise zero anaphora while still keeping 'real' pronouns out of the game. The use of clitics thus constitutes a compromise in both types of syntactic constructions, one that exploits their in-between status as a semi-bound and yet overt anaphoric reference device. In the next section, this tentative interpretation is examined on the basis of data from the more extensive Mark corpus.

4. Distribution in the Mark corpus

The hypothesis suggested in the previous sections implies that the structural distri-
bution of clitics, rather than depend on the inherent properties of clitics as markers
of greater or lesser referent continuity, arises from the tendency of certain types of
syntactic constructions to rely on clitics for anaphoric reference. This in turn is a
strategy for regulating a cohabitation of convergent and non-convergent structures
in a language contact situation. Let us therefore begin this section, in which the bulk
of the corpus is evaluated, by considering each of the relevant syntactic construc-
tions and the anaphoric devices they show.

4.1 Adverbial clauses and embeddings

In adverbial subordinations, the Mark corpus shows almost exclusively free pro-
nouns in a word order type that is compatible with that of German, that is, SV, with
the finite verb assuming the final position in the clause:

(14)   Har job noch jake rekeras,   wajan   i paar dran ko kher von kolester
        how he still so   spoke-3sg came-3pl a few from this house of   this-abl
        'While he was saying this, some people came out of this person's house'   (Mk 5: 35)

Postverbal clitics in adverbial clauses are strictly a marginal phenomenon. Free pro-
nouns are also the preferred strategy in factual embeddings and relative clauses, as
illustrated in the third clause in (15). Here too, word order is compatible with German:

(15)   Und mangan   lester,   te   krelo   kowa, hoi job immer ap koi feira
        and asked-3pl him-abl comp do-3sg-cl this   what he always at this ceremony

        krela
        do-3sg-fut

        'And they asked him to do what he would always do at this ceremony'   (Mk 15: 8)

Clitics also appear in embeddings of this type, however. Short clitics, as in
(16)–(17), attach to the verb, rendering a VS arrangement, while long clitics in most
cases either follow the conjunction, as in (18), or follow a pronoun, thus showing a
tendency towards SV, though in some cases they too attach to the verb:

(16)   Har wel   job koi zu, kai chalo   mit kol zöllnaria   und sindaria?
        how come-3sg he this to   that eat-3sg-cl with these tax-collectors and outcasts
        'Why does he eat with these tax-collectors and outcasts?'   (Mk 2: 16)

(17)   Koia   krajas,   hoi   naschte   krajasi.   (Mk 14: 8)
        this   did-3sg   what   could   did-3sg-cl
        'She did what she could do'

(18)   Und har   job   an   ko   kher   dren   dschajas,   schunas   i   dschuwel   von   lester,
        and how   he   in   this house inside went-3sg   heard-3sg the woman   of   him-abl

        kai   lo   koi   hi.   (Mk 7: 25)
        that cl   there is

        'And as he went into that house, the woman heard about him, that he was there'

In evaluating the position of VS order and short clitic appearance, as in (16)–(17), one must take into account the replication of modality-type, i.e. of *te*-type structures in infinitival constructions: (16) is clearly calqued on German *Wie kommt er dazu, ... zu essen?*, lit. 'How does he come to it, to eat?'. The statistical representation of VS and short clitics in embeddings is further obscured somewhat by instances such as (17), where again the clitic actually appears in the modal complement.

The picture so far is therefore as follows: There is an almost complete replication of German word order patterns in adverbial subordinations, which carries with it a shift to an almost exclusive use of free pronouns as anaphoric devices. As a general tendency, this drift is found in embeddings as well. Short clitics and VS order are retained in some environments which come close to modal embeddings. Long clitics retain a typical position following conjunctions and pronouns, where they nevertheless yield to the drift in word order patterns, and so display SV.

4.2 Modality and *te*-constructions

The infiltration of modality, it could be seen, motivates resistance to the drift to SV and the subsequent takeover of all anaphoric functions by free pronouns. And indeed in genuine modality constructions, i.e. those introduced by *te*, clitics constitute by far the majority of anaphoric occurrences:

(19)   Aber   job   dschajas   krik   und   fangas an,   te   rakerelo   but   von   kowa
        but   he   went-3sg   away   and   began-3sg comp speak-3sg-cl much   of   that
        'But he went away and began to tell alot about that'   (Mk 1: 45)

(20)   rodan   tschimone khate   o   Jeuseste,   te   marene   les.
        searched-3pl   something against the Jesus-loc   comp kill-3pl-cl   him
        'They looked for something (to use) against Jesus, in order to kill him'   (Mk 14: 55)

(21)   Tschib tire wasta ap late,   te   weli   sasto und te   dschiweli
        put   your hands on her-loc comp come-3sg-cl well and comp live-3sg-cl
        'Put your hands on her, so that she may recover and live'   (Mk 5: 23)

Note that the question of subject identity (equi) in the two parts of the construction does not affect the choice of anaphoric device in the subordinated clause, and that same and different subject constructions behave alike. Nor is the choice affected by semantic integration (single versus non-single event), which might have been expected to motivate a difference between modal complements and purpose clauses.

Similarly, the choice of long clitics in *te*-constructions does not appear to be motivated by pragmatic considerations of referential continuity either. Following object and reflexive pronouns, as in (22), the conditioning is clearly structural, while a case of a long clitic attaching to the verb is demonstrated in (23):

(22)   Und job phenas   kol menschenge, te   beschen pen le ap i   phub.
        and he   said-3sg these people-dat comp sit-3pl refl cl on the ground
        'And he told those people to sit down on the ground'   (Mk 8: 6)

(23)   Und jon bisteran, te   lenle   maro   peha.
        and they forgot-3pl comp take-3pl-cl bread refl-soc
        'And they forgot to take along bread'   (Mk 8: 14)

The choice of a free pronoun in *te*-constructions, though marginal in the corpus, is accompanied, interestingly, by an adaption of the pronoun to VS order:

(24)   Job rodas   desch ta   duien wi kai dajaso   ko   lab apostle,
        he   searched-3sg ten   and two-acc out rel gave-3sg-cl this name apostle

        te   wen   jon pasch leste   und te   bitscherelo len wi,
        comp come-3pl they by   him-loc and comp send-3sg-cl them out

        te   phenene   o dewleskro lab   durder.   (Mk 3: 14)
        comp say-3pl-cl the Lord-gen   word farther

        'He picked out twelve people, whom he called apostles, to accompany him, and to send out
        to spread the holy word'

This is also found in *te*-clauses which, calqued on German *wenn*, express adverbial subordination. Most of those show short clitics, but also allow free pronouns in VS order:

(25) Und te wajaso maredo, dann stel job pal trin diwesa pale pre
and if became-3sg-cl killed then stand-3sg he after three days again up
'And when he is killed, he will rise to life again three days later' (Mk 9: 31)

(26) Te schunen jon o lab, lene les sik mit freuda pre.
if hear-3pl they the word take-3pl-cl it fast with happiness up
'As soon as they hear the message, they receive it gladly' (Mk 4:16)

We see that modality in Sinti, as in Romani in general, is linked to the choice of *te* as a complementizer, and relies on the proximity of the conjunction to the verb, triggering VS order. The retention of this feature results in word order patterns which resist convergence with German, despite the general trend in this direction that can be observed in the corpus. The fact that *te*-constructions rely on clitics, rather than on free pronouns, for anaphoric reference to subjects is connected to their being the natural choice of anaphor, on structural grounds, in postverbal position. The occasional occurrences of free pronouns in VS order in modality clauses shows just how stable VS is in these constructions, unlike adverbial subordinations, where the use of free pronouns is linked, as in German, to SV. Clitics, however, remain a convenient choice as they help avoid a structurally more complex VS construction with free pronouns, and so reduce the friction between inherited and convergent structures.

4.3 Verb–subject inversion

Finally, the Mark corpus shows similar rules as the Holzinger corpus for verb–subject inversion. Linear or formal inversion is triggered, as in German, by a third entity occupying the first sentence position. Short clitics are the preferred anaphoric device here:

(27) Und har job jake rakeras, phenaso ap lende: (Mk 4: 2)
and how he thus spoke-3sg said-3sg-cl on them-loc
'And as he thus spoke, he said to them:'

(28) Und an i rati dschajaso wi nach Betanien (Mk 11: 11)
and in the night went-3sg-cl out to Bethany
'And in the night he went out to Bethany'

(29) Und kol dui matsche dajaso nina i zele menschende. (Mk 6: 41)
and these two fish gave-3sg-cl also the all people-loc
'And he also distributed those two fish among all the people'

(30)   Dann dschajaso   an kol   gaba   (Mk 6: 6)
        then went-3sg-cl in these villages
        'Then he went to those villages'

Frequently, inversion is triggered by the deictic *koi* copying German *da*, as in (31),
where the corresponding German structure is *da sagte er*:

(31)   Und job putschas   lester: 'Har   khareh   tu?'
        and he   asked-3sg   him-abl   how   call-2sg you

        Koi phenaso: (Mk 5: 9)
        dct said-3sg-cl

        'And he asked him: 'What is your name?'
        He said:'

Connective inversion of the type encountered in the Holzinger corpus appears as
well, though less frequently, possibly since we are dealing with a translation of a
written text, and not with a narrative which is continuously being restructured and
rearranged by the speaker:

(32)   Und jon   luran   koi pre, ob   job les nina ap o   heiligo
        and they waited-3pl dct up   whether he   him also on the holy

        diwes sasto   krela.
        day   healthy make-3sg-fut
        ( )
        Phenaso   ap   lende: (Mk 3: 2–4)
        said-3sg-cl on   them-loc

        'And they waited (to see) whether he would also heal him on the Sabbath.
        ( )
        (And so) he said to them:'

Alongside clitics, we find free pronouns in inversions as well. Again there are no
obvious pragmatic conditions triggering the use of pronouns. Thus, in (33), the
inverted subject pronoun refers to a continuous subject, but a disrupted topic. In
(34) the pronoun represents a continuous topic, but a subject switch. Finally in (35)
it is a continuous subject-topic:

(33) Job krajas lauter mischto. Kolen, kai schunen naschti gar, krel
he did-3sg all   good   those-acc rel hear-3pl can not do-3sg

job, te   schunene   (Mk 7: 37)
he   comp hear-3pl-cl

'He did everything well. He causes those who can't hear to hear.'

(34) Kol verspotteren les und tschungeren les   an und peitscheren les wi
those mock-3pl   him and spit-3pl   him on and whip-3pl   him out

und maren les. Und pal trin diwesa stel   job pale pre. (Mk 10: 34)
and kill-3pl him and after three days stand-3sg he again up

'They will mock him, spit at him, whip him, and kill him. And three days later he will rise
back to life.'

(35) Und an koi momenta, har job dran o pani dschajas, dikas   job, kai o
and in that moment, how he from the water went-3sg saw-3sg he that the

bolepen krajas   pes pre.   (Mk 1: 10)
heaven   did-3sg refl up

'And just as he came out of the water, he saw heaven opening'

Thus, in the Mark corpus, we find free pronouns gaining ground in postverbal position as well..

4.4 A quantitative evaluation

It is now time to turn to a quantitative evaluation of the extent to which complex syntactic constructions rely on specific anaphoric devices in the corpus, and of the distribution of subject clitics. The first is illustrated by Table 2.

Adverbial subordinations overwhelmingly draw on free pronouns with SV order. A similar tendency can be observed in factual embeddings (complements and relative clauses), though here short clitics with VS order also occur (often triggered by modality-like constructions, as pointed out above), while long clitics tend to attach to the conjunction, showing in such cases SV order. On the other hand, both *te*-constructions and inversions rely heavily on short clitics, which is also the general tendency in conditional clauses introduced by *te*. In short, there is a clear preference for clitics in syntactic constructions that demand VS order.

Conversely, the employment of free pronouns in complex syntactic constructions depends on the degree to which the subject is permitted to assume a prever-

YARON MATRAS

Table 2

Mark Corpus: Use of anaphoric devices in syntactic constructions

	Total	Short clitics	Long clitics	Free pronouns
Adverbial subord.	48 (100%)	2 (4%)		46 SV* (96%)
Compl./relat. clauses	82 (100%)	16 (19.5%)		46 SV (56%)
Follows conj.			12 (15%)	
Attaches to verb			6 (7%)	
(in 1 case to *hi*)				[1 (1.2%)]
Follows refl. pron.			1 (1.2%)	
Follows obj. pron.			1 (1.2%)	
te-compl/modal compl	87 (100%)	75 (86%)		6 VS (7%)
Follows refl. pron.			3 (3.5%)	
Attaches to verb			3 (3.5%)	
Conditional clauses	18 (100%)	11 (61%)	1 (5%)	6 VS (34%)
VS inversion	142 (100%)	88 (62%)		44 (31%)
Attaches to verb			9 (6.3)	
Follows refl. pron.			1 (0.7%)	

* Free pronouns show SV order in adverbial clauses, factual complements and relative clauses, but VS in *te*-constructions (modal, purpose, and conditional clauses) as well as, by definition, VS inversions

bal position. The more rigid the rule for postverbal placement of the subject, the stronger the tendency to use clitics. In the case of inversions, however, the rule on postverbal positioning of the subject is borrowed from German. It appears to have been initially realized entirely by clitics, but is gradually admitting free pronouns, thus conforming even more closely to the German model. Yet one must keep in mind that the occurrence of full pronouns in inversion constructions is marginal considering their general distribution in the language, and that the 44 cases of inverted free pronouns are an almost meaningless fraction of the total occurrences of free pronouns in the corpus.

The results could be considered to harmonize with Holzinger's (1993) continuity hierarchy if adverbial clauses were assumed or found to show emphasized subjects or subject switches, whereas purpose clauses and modal clauses were to be

Table 3

Mark Corpus: Distribution of clitics

	Short forms	Long forms
Total	193 (100%)	39 (100%)
Adverbial subord.	2 (1%)	0
Compl./relat.clauses	16 (8%)	
Follows conj.		12 (31%)
Attaches to verb		6 (15%)
Follows refl. pron.		1 (2.5%)
Follows obj. pron.		1 (2.5%)
te-compl./modal complements	75 (39%)	
Follows refl. pron.		3 (8%)
Attaches to verb		3 (8%)
Conditionals	11 (6%)	1 (2.5%)
VS inversion	88 (45.5%)	
Attaches to verb		9 (23%)
Follows refl. pron.		1 (2.5%)
Other	1 (0.5%)	1 (2.5%)
Follows *hi*		1 (2.5%)

considered as having high topic and subject continuity, and inversions to constitute a mixed type. However, above we have already seen that the use of pronouns and clitics in the Mark corpus does not necessarily follow the continuity hierarchy postulated by Holzinger.

Let us now approach the same data from a different perspective, and examine the distribution of subject clitics in various syntactic environments (Table 3).

The crucial observation is that the distribution of clitics is connected to specific structural devices. Short clitics mainly appear in inversions, and in *te*-enhancements, which, if one adds conditional clauses to modal complements and purpose clauses, amount to as high a score as attained for inversions. In other words, some 90% of short clitics occur in constructions in which VS order is not an option, but

is obligatory. Of those, only the case for connective inversion could actually be challenged on the grounds that they are subject to speaker's choice, rather than syntactic constraints or conditions. But the choice pertains to the type of serialization or chaining device itself; once alternative devices such as adverbs or conjunctions are rejected, treating the preceding clause as a thematic point of departure will automatically trigger VS order in the following clause. Such constructions are quite straightforward and easy to recognize in the corpus. They always appear in clause-initial position and show a tight thematic link with the preceding clause. But they are also much less frequent than cases of formalized (linear) inversion, and so even if interpreted differently than along the lines followed here, they are unlikely to influence the general picture obtained.

Long clitics are somewhat more evenly distributed, but their most salient feature in opposition to the short clitics is their possible placement following conjunctions and pronouns, which altogether accounts for more than 50% of their occurrences. Their striking presence, compared with the short forms, in embedded complements and relative clauses is due to this feature. On the whole, the confused picture that emerges for long form clitics may be taken to reflect their general retreat in the language. With merely 39 occurrences in the Mark corpus, long clitics are a marginal phenomenon in the dialect, a result supported by Holzinger's observations. Short clitics, on the other hand, may be said to be able to survive through specialization for certain syntactic constructions.

5. Conclusion

While clitics are generally retreating in Romani, short form clitics are able to survive in Sinti through specialization for certain syntactic constructions, namely those in which postverbal placement of the subject anaphor is obligatory. There are essentially two such devices: The first is inherited from Common Romani and includes *te*-enhancements in their various functions, such as purpose clauses, modal complements, and conditionals. The other, verb–subject inversion, especially in its formalized version, is borrowed from German. The role that clitics assume is still connected to their older function as anaphoric reference devices, though it is argued here that their distribution is now governed predominantly by syntactic rules, rather than by their hierarchical status on a scale of referential continuity. It is therefore difficult to regard their change in function as a genuine case of exaptation in the sense proposed by Lass (1990), and documented for the active participles with adjectival agreement in other dialects of Romani (Matras 1995a). One might

instead choose to view clitics and free pronouns as an inherited opposition gradually assuming a new complementary distribution.

Thus, the following scenario can be postulated for the role of subject clitics and their development in Sinti: The language inherits two sets of overt anaphoric reference devices: emphatic free pronouns that appear mostly in preverbal position, and continuous (long) clitics that appear in postverbal position. These devices are rearranged as the language undergoes convergence of word order and other syntactic patterns with those of the contact language, German. Clitics are gradually restricted to syntactic constructions in which VS order is obligatory. One of those is inherited, as Sinti preserves the rule on VS order in *te*-constructions. Here, it was argued, clitics help reduce the friction which arises through the retention of a non-convergent structure. The other, verb–subject inversion, is a result of Sinti-German convergence of word order patterns. Clitics now appear here, while zero anaphora is compromised. Free pronouns are reserved for other constructions. They gain ground in environments which now, due to convergence with German, favor SV order: adverbial clauses, (factual) embeddings, and relative clauses. In addition, of course, they assume the function of pronominal or overt anaphoric reference in thematic sentences. As a result of this rearrangement in the distribution of subject clitics and free pronouns, and especially as a result of the formalization of their distribution, at least in some environments, clitics on the whole become less continuous, and free pronouns become less emphatic. Finally, at a stage the beginning of which may be observed especially in the Mark corpus, we are confronted with the gradual appearance of free pronouns in postverbal position as well. This may be motivated by the German model for inversions, which, German having no subject clitics, makes use of free pronouns; but it can also be observed in some *te*-clauses which have lost or partially lost their modality feature. At any rate, the infiltration of postverbal positions by free pronouns reinforces the beginning retreat of subject clitics, a development already evident through the reduction of long forms to short forms in positions immediately following the verb, and the fact that the original long forms are rather scarce.

References

Boretzky, N. 1994. Romani. Grammatik des Kelderaš-Dialekts mit Texten und Glossar. Harrassowitz, Wiesbaden.

Boretzky, N. 1995. Die Entwicklung der Kopula im Romani. In: Grazer Linguistische Studien 43: 1–50.

Boretzky, N. 1996. The "new" infinitive in Romani. In: Journal of the Gypsy Lore Society, Series 5, 6: 1–51.

Friedman, V. 1986. Evidentiality in the Balkans: Bulgarian, Macedonian, and Albanian. In: Chafe,
    W. – Nichols, J. (eds): Evidentiality. The linguistic coding of epistemology, 168–87. Ablex,
    Norwood.

Givón, T. 1995. Functionalism and grammar. John Benjamins, Amsterdam.

Hengeveld, K. 1992. Non-verbal predication. Theory, typology, diachrony. Mouton de Gruyter,
    Berlin.

Holzinger, D. 1993. Das Rómanes. Grammatik und Diskursanalyse der Sprache der Sinte. Institut für
    Sprachwissenschaft, Innsbruck.

Kovalcsik, K. – Tálos, E. 1991. The little maple tree: A Transylvanian Gypsy folk tale with songs. In:
    Journal of the Gypsy Lore Society, Series 5, 1: 103–25.

Lass, R. 1990. How to do things with junk: exaptation in language evolution. In: Journal of Linguistics
    26: 79–102.

Matras, Y. 1991. Zur Rekonstruktion des jüdischdeutschen Wortschatzes ehemaliger 'Judendörfer' in
    Südwestdeutschland. In: Zeitschrift für Dialektologie und Linguistik 58: 267–93.

Matras, Y. 1994. Untersuchungen zu Grammatik und Diskurs des Romanes. Dialekt der
    Kelderaša/Lovara. Harrassowitz, Wiesbaden.

Matras, Y. 1995a. Verb evidentials and their discourse function in Vlach Romani narratives. In:
    Matras, Y. (ed.): Romani in contact. The history, structure and sociology of a language, 95–123.
    John Benjamins, Amsterdam.

Matras, Y. 1995b. Connective (VS) word order in Romani. In: Matras, Y. – Sasse, H.J. (eds) Verb–
    subject order and theticity in European languages, 189–203. (Sprachtypologie und Univer-
    salienforschung 48). Akademie, Berlin.

Reershemius, G. 1997. Biographisches Erzählen auf Jiddisch. Grammatische und diskursanalytische
    Untersuchungen. Niemeyer, Tübingen.

Rehbein, J. 1992. Zur Wortstellung im komplexen deutschen Satz. In: Hoffman, L. (ed.): Deutsche
    Syntax, 523–74. De Gruyter, Berlin.

Sampson, J. 1926. The dialect of the Gypsies of Wales. Clarendon, Oxford.

Smart, B. C. – Crofton, H.T. 1875. The dialect of the English Gypsies. Asher, London.

Valet, J. 1991. Grammatical sketch of Manouche. In: Bakker, P. – Cortiade, M. (eds): In the margin of
    Romani. Gypsy languages in contact, 106–31. Institute for General Linguistics, Amsterdam.

Address of the author: Yaron Matras
                            Department of Linguistics
                            University of Manchester
                            Oxford Road
                            Manchester M13 9PL
                            England
                            e-mail: y.matras@man.ac.uk

GRAMMATICAL INTERFERENCE IN ROMANI: LOAN FORMATIONS FOR FOREIGN CATEGORIES

NORBERT BORETZKY

Abstract

Romani has been influenced by a number of European contact languages, as a rule particular dialects by particular European languages. The extent of borrowing is greatest in the lexicon, but the contact languages have left their traces in the grammar as well. An important number of grammatical categories appears to have been created or at least shaped by this influence, not only via direct borrowing of markers ot other structures, but also by calquing foreign expressions, either by literal translation if this was feasible, or by rendering them according to the original sense. In this paper, some of these processes are discussed, among them the creation of expressions for future, permissive, modal constructions, superlative, infinitive, and perfect.

1. General remarks

During its European history Romani has experienced a lot of foreign influences, which as a rule did not alter the overall character of the language (cf. Boretzky 1996a), but which in individual dialects amounted to a high degree of extrasystemically conditioned change. For contact linguistics, the study of Romani interferences is especially interesting because particular dialects got into contact with various languages of Europe, thus opening up the possibility of comparative studies. We can state that, within Europe, there is no other language that underwent such different influences as Romani. Influence affected all levels of the language, most vigorously the lexicon (cf. Boretzky 1992; Vekerdi 1980), less importantly the sound types and the phonological system of the dialects (cf. Boretzky–Igla 1993), and to a varying degree the morphological system. Foreign morphemes of word formation can be found in nearly all dialects, whereas inflectional morphemes in the strict sense of the word are seldom borrowed (cf. Boretzky–Igla 1991). A little more frequent is borrowing of independent markers (functional words). Dialects spoken north of the Balkanic region appear to have been heavily influenced by the word order rules of the contact languages (Slavic languages, German, Hungarian; cf. Boretzky 1996d). In this paper, indirect influences upon morphology, i.e. the creation of new grammatical categories under the influence of contact languages, but

formed by indigenous means, are examined. Most of these processes can be char-
acterized as grammatical calquing. We are going to discuss here some of the more
important and more widespread phenomena that have not been dealt with to date in
special publications.[1] For reasons of space, we have restricted ourselves to collect-
ing the data and providing evidence for the interference argument as opposed to the
internal development argument, but we did not attempt to show how the new struc-
tures interacted with the existing ones.

2. New categories in Romani

For reasons of space we are concentrating on categories that can be assumed to
have not existed or have not been fully shaped in pre-European times of Romani,
leaving aside categories that must have existed but have been reshaped under the
influence of European languages. Some items will remain controversial, mainly
because we have no Romani texts written prior to or immediately after the Romas'
immigration to Europe, and because language material appropriate for linguistic
studies is only available from the 19th century on. What we can do, then, is to draw
conclusions from the state of the various dialects about the state of the language
immediately prior to the immigration, not about the whole period between the emi-
gration from India and the arrival at Anatolia.

2.1. Future formation

2.1.1. Future formed by *kam-* 'love' and *mang-* 'ask, demand'

We cannot claim with certainty that, before contact with Greek, there was no
future category in Romani, since the expanded form of the present ending in *-a*
(*ker-av-a̱*) is used in some varieties of Vlach as well as in Central dialects for
expressing future events. In other Central dialects and in Sinti, however, it is not
strictly separated from the true present, and in South Balkanic dialects, as e.g. in
Arli, Erli, Bugurdži, Drindari etc., it is reserved to present tense. Therefore, it is not
very clear what the original function of this form was like. In my opinion, the *a*-
form was an unspecified present that could be used as a future too, whereas the
short form (*ker-av*) originally functioned as a subjunctive but later on became con-
fused with the old present form in practically all dialects. In dialects receiving a

---

[1] From the very beginning of Romani studies scholars have drawn attention to the extensive
amount of loan elements in this language. Details have been mentioned in many books and articles,
but no special study has been dedicated to this topic. For linguistic orientation about the phenomena
of interference in general cf. Boretzky–Igla (1994b)

new future morphology the long form specialized for the present, but in Central Europe it continued to serve for both present and future; cf. Table 1.

Table 1
Types of future formation

	present	subjunctive	future
Type I	*ker-el-a/ker-el*	*te ker-el*	*ker-el-a*
Type II	*ker-el*	*te ker-el*	*ker-el-a* (and innovations)
Type III	*ker-el-a (ker-el)*	*te ker-el*	(innovations)

Central dialects preserved type I, type II is represented by some Vlach dialects (varieties of Kalderaš and Lovari), and type III with an obligatory new future is found among the South Balkanic dialects.

In this reconstruction, the *a*-form is conceived of as an old (perhaps emphatic) present, which in default of a specific future form served as a future as well (cf. the unmarked present-future in German).

Under the influence of the Balkanic languages, and Greek in the first line, a periphrastic future was formed with the aid of the verb 'love, want', as has been known for a long time.[2] Today this verb appears in reduced form in the Balkanic languages, but in former times it was recognizable as such and could be translated into Romani; cf. the following constructions in Greek: θέλω νά > θέ να > θά (the intermediate stage is partially preserved in Greek dialects); Alb.: *do të* (*do* is not distinguished from the full verb 'love, want'!); Bulg.: *šte* + present, in dialects also *šte da* + present < *xášte da*; in dialects inflected *šta/štes/šte* etc. functions as an independent verb 'want, will' even today; Rum.: *voi/vei/va* + infinitive, but in the colloquial language *o să* + subjunctive is used more frequently.

Thus it was not difficult for the Romani speakers to copy this model in the form of *kamav te kerav, kames te keres, kamel te kerel,* etc. Later this construction was reduced to *kama-, kam-* and *ka-,* respectively, and combined with the mere subjunctive, i.e. *kama-kerel/kam-kerel/ka-kerel,* under foreign influence rather than in an internal process of grammaticalization. The most frequent form is *ka-,* but the other two also survived. Besides this grammaticalized form the old unreduced construction *kamel te kerel* has been preserved with its literal sense, which means that there occurred a functional split.

---

[2] Pott, who had little knowledge of the Balkanic dialects, mentions *kam-ela* only as a modal verb (1844; I, 360ff), but already Paspati (1870, 101f) and Miklosich (1881, 101) derived *kam/ka* from *kam-el* relating this process to Greek θέλει να > θά.

One might argue against this view on the basis that a verb 'want, will' has been used in many languages for forming a future and that it is not cogent to assume Balkanic influence, but the details speak in favour of the Balkanic option, and moreover, this future form is not found to the North of the Vlach dialects, where there were no languages that could have served as a model. Also, it is rather unlikely that the Northern dialects should have known this future formation, but lost it later under the influence of the new contact languages without leaving traces. Apparently, the Balkanic Romani dialects did not develop this future at the first contact with Greek, but only after a long-lasting coexistence with the Balkanic languages.

Besides this construction there is another, especially in Drindari of Bulgaria, making use of the verb *mang-* 'ask, demand, want', which is reduced to *ma- (mə-)* (see Gilliat-Smith 1913–14, 277). From a semantic point of view, this formation is as plausible as that with *kam-*. There is, however, another possibility for deriving *ma-*, namely from *kama-* (cf. above) by reducing the first syllable. (By the way, the origin of the second *a* is not clear, and *kama-kerel* e.g. in Southern Gurbet has the function of a conditional rather than of future tense, but it cannot be doubted that it is derived from *kam-el*.) A decision in favour of one of these suggestions will only be possible if forms less reduced than *ma-*, something like *mang-kerel*, are found. In view of the fact that for Bulgaria there are no texts prior to the last century, the question might never be settled.[3]

2.1.2. Future formed by an equivalent of 'have'

Another future formation seems to be copied from Bulgarian–Macedonian (and perhaps Albanian) models. As is known, there is another possibility in Bulgarian for expressing the negated future, not by 'will', but by the verb 'have':

(1)   njama da ida / nema da idam 'I won't go'

i.e. with the aid of an auxiliary already reduced to a non-inflected particle *njama/ nema*.

In Macedonian dialects even positive constructions containing *ima da* are in use for the future. To be sure, there is no verb for 'have' in Romani, *si / naj* 'is / isn't' + accusative being used instead; cf. *si / naj man / tut / les*.[4] Apparently, *naj* for

---

[3] Kenrick (1967, 77) has *mâ* for the positive, and *nâma ma* or *nâma tâ* for the negative future. At least the last form looks much like a contamination of Romani *na ma* (NEG-FUT) and Bulgarian *njama da* (dial. *nema da*)! See also the following section.

[4] Modern Indic languages do possess a verb for 'have', but in some of them a construction comparable to the Romani one is in use; e.g. in Punjabi (local) 'be' + genitive: *munde da bəut vəḍḍa sir si* 'the boy had a very big head'. This may be an old construction in the languages of the Indian subcontinent, and therefore an inherited category in Romani.

'haven't' was utilized in order to copy the Bulgarian and Macedonian construc-
tions, thus resulting in *naj te džav / te džas / te džal* etc.

Sometimes even a positive future is formed, i.e. the construction *si te džav* may
be used with both necessive and future meaning ('I have to go' and 'I will go'). The
positive meaning may result from a process of generalization, but it is equally pos-
sible that it goes back to Slavic dialectal usage. Some examples:

(2) ma dara, nane te čhingarol tut o phral tuke
     'don't be afraid, your brother will not scold you' (Prilep)[5]

(3) tut me nisar ninaj te mukav
     'I won't abandon you under any circumstances'
     (Vlach, Bulgaria; Romane svjati gilja 1933, 43)

In Gegue (Northern) Albanian the future is generally formed with the aid of 'have'
+ infinitive; e.g. *kam me shkue* 'I'll go' as well as *s' kam me shkue* 'I won't go'.
This had repercussions in various dialects spoken in Kosova:

(4) dik so hi ma te kerav tumara dizake
     'look, what I will do to your town!' (Prizren)

(5) kan ka dikhel mo dad o šandani, ov si te mangel les tuke
     'as soon as my father will catch sight of the candle-stick, he will demand it from you' (Bugurdži)

2.1.3. Future formed by 'go'

In Welsh Romani a sort of future is formed by *džava te* etc. 'I am going to...'. It is
self-evident that here the English construction has been the model for Romani,
since there is nothing comparable found in any other dialect: wherever 'go' is con-
structed together with the subjunctive, it preserves the original meaning. Examples
from Wales (cf. Sampson 1926, 191):

(6) ake me džava te xå 'now I am going to get my dinner'

(7) brišindo džala te del 'it is going to rain'

---

[5] Where no sources are given, the examples are taken from Boretzky 1993a (Bugurdži), Boretzky
1994a (Kalderaš) and from unpublished recordings of N. Boretzky and B. Igla (for Džambazi, Gurbet,
Prilep dialect, Arli).

In the second case 'go' has lost the connotation of moving from one place to another (cf. Hopper–Traugott 1993, 1ff and 61f), and it is clear that by this step the construction has become a futurum proximum. This construction of English could not be imitated in all details by the Romani speakers, simply because an equivalent for *-ing* in Romani, *-indos/-indoj*, has fallen out of use in Welsh Romani. It has to be emphasized here that such incongruencies between languages in contact are nearly never a hindrance for interference. Some linguists have claimed that in order to prove interference (as opposed to internal development) the rules of the model language must have been copied in all details, but this view is not tenable in view of our experiences with language interference (for the arguments cf. Boretzky 1986; 1993c).

2.1.4. Future formed by 'take'

In dialects spoken in Russia and the Ukraine a future construction containing *l-el* 'take' is found, the auxiliary being inflected for person; cf. *lav te kerav / les te keres / lel te kerel* etc. (cf. Wentzel 1983, 72, 77).

Similar constructions can be found in other dialects too, e.g. in South Balkanic ones, although here paratactic constructions of the type *lav thaj kerav* 'I take and make' seem to prevail. However, the meaning is different from that found in Russia: 'begin to do' or 'take to doing' and similar. It cannot be excluded that in the course of natural, internal evolution an erstwhile ingressive has changed into a future, but in our case this is less likely since there is a model for the future function of the construction discussed here. Besides the normal East Slavic future formed by *budu* + infinitive or by the simple present form of the perfective verb there is an Ukrainian formation making use of a verb 'take'; cf. *čytaty-mu / čytaty-meš / čytaty-me* etc.[6]

(8)   me na lava tusa [te] paruvav 'I'll not change with you' (Dobrovol'skij 1908, 24).

Apparently, the marker is derived from an old inf. *jati*, pres. *imu, imeš, imet'* etc. 'take', which survived in a prefixed form only: Ukr. *uzjaty, prijnjaty* and Russ. *vz-jat'*, *pri-n-jat'* and others. This Ukrainian construction may have influenced Romani, whereby the semantic content, but not the morphosyntactic rules of the model had been adopted. We have to take into account that future expressions formed with the aid of 'take' are rare in Europe. Sure, the preconditions for such a future form to come into being were favourable since there were the inherited ingressive constructions mentioned above, but we have to ask ourselves why it is only in Ukrainian and Russian environment that this future form arose. We may add here

---

[6] This connection has already been detected by Miklosich (1881, 101).

that even in Old Russian from the 14th–15th century on future-like constructions emerged that made use of *jati*, pres. *imu* etc. (not to be confused with *imam* etc. 'have'), and further with *počnu/učnu* 'begin' which besides their future function all preserved the connotation of beginning (see Kiparsky 1967, 234). Thus it is not unlikely that Gypsies in South Russia became acquainted with such constructions.

2.1.5. Future formed by 'become'

In Russia, there is a different construction competing with 'take', formed with the aid of 'come = become' *av-ela*. According to Eloeva–Rusakov (1990, 15) it is found only rarely in North-Russian Romani, but there is much evidence for it in the texts of Dobrovol'skij (1908). The two authors derive *avela te kerel* 'he will do' correctly from Russ. *budet delat'*:

(9)   išče fedyr tut avela te žaline 'she will love you even more'

In the same paragraph there is a future formed by *lela* 'take' in a sentence with nearly identical wording:

(10) ta jou tut išče fedyr lela te žaline 'and she will love you even more' (p. 11)

Besides those two possibilities the present ending in -*a* can be used for future, which means that no standard form has been developed.

2.1.6. Non-marking of future

In Sinti (Finck 1903, 10; Holzinger 1993, 98) future is not formally distinguished from present; the long forms ending in -*ava*, -*eha*, -*ela* etc. are in use for the present and the future as well, whereas the short forms -*av*, -*es*, -*el* etc. have predominantly present value. But even to them future meaning may be attached. This merger of functions within one form is in congruence with the usage of the present form in German, where in colloquial style the marked future *ich werde gehen* occurs rather seldom, and this makes us believe that the merger was effected by German influence or, at least, that a possible non-distinction of the two tenses was reinforced by German. Furthermore, this argument is supported by what will be called here cumulative evidence. As we were able to demonstrate, there is a number of different future formations all displaying parallels to their respective contact languages.

Only those dialects that use the long form in -*a* exclusively as future or in a modal sense with future connotation, developed forms clearly independent of contact languages. These are varieties of Kalderaš and Lovari, and perhaps some of the Central dialects. Lípa (1965, 37f) claims that pres. *phen-av* and fut. *phen-ava* are

strictly distinguished in Slovakian Romani (or in some varieties only?). According to Barannikov (1934, 95), this holds for the dialects spoken in the Ukraine and South Russia, but in Dobrovol'skij (1908) the long form is used indiscriminately for both tenses. It is perhaps for that reason that future variants formed by the auxiliaries *lela* and *avela* were gaining ground.

The dialect of Finland is similar to the Central dialects (cf. Thesleff 1901, 120), and in the dialect of Wales both forms are used for the present, but for the future the long form is preferred (Sampson 1926, 189). Here the question arises if this dialect has been for some time under the influence of German. Apparently, the English future formed by *will* (*shall*) has not been copied.

2.2. Permissive constructions

In the Balkanic languages the verb 'let' has been grammaticalized for expressing adhortative, permissive, optative and even concessive meanings, the imperative form of 'let' serving as a basis for grammaticalization. As is usual in such processes this form was partially phonetically reduced; cf. Greek ἄφησε > ἄσε ἄς, Slavic *nechaj/nech* + -*ka* (particle) > *neka/nek* (*da*); Alb. *le të* and Rum. *lasă* (the last two without reduction).

Romani has utilized the imperative form of *mek-el/mukh-el*, *mek/muk* in the same way as the Balkanic models did but, as far as I can see, not with the same multitude of functions. The Greek model ἄς was transparent enough to be translated into Romani, since it does not differ much from the imperative aor. ἄσε even in its recent form. In addition, this particle is used as a marker of both adhortative and permissive, mostly in the third person, less often in the first person, but apparently not in the second person, where simple *νά* 'that' is sufficient. The concessive marker is κι ἄς 'even if'. This particle can also be combined with the imperfect, a construction which, to my knowledge, is not possible in Romani. It is conspicuous that the finite verb is immediately linked with ἄς, i.e. an earlier *νά* 'that' has been abandoned, if the verb immediately follows ἄς. Examples from Greek (Mackridge 1987, 298):

(11) ἄς πᾶμε κι ἐμεῖς 'let us go too!' (adhortative)

(12) ἄς μήν ἔρθουν, ἀφοῦ δέ θέλουν
     'may they not come, if they don't want to!' (permissive)

(13) ἄς τόν νά πάει 'let him go!'

(14) ἔλα στό πάρτυ μας, κι ἄς φέρεις καί τόν ἄνδρα σου
     'join our party, even if you bring with you your husband!' (concessive)

It is possible that in Albanian this element has even more functions, but since this language was a contact language for only a few Kosova dialects of Romani, we will not go into details. For the time being nothing can be said about possible further influences of Albanian upon the local dialects.

In Bulgarian, the third person forms are used with adhortative, permissive and optative senses, and with the 1sg a demand addressed to the ego is expressed, whereas concessive use is not possible. As in the two other languages, the particle *neka* can be combined with past tenses (cf. Gramatika III, 1994, 70f).

Rum. *lasă că* is restricted to the concessive function: 'even if; it's not only that ... but'. Rumanian differs from the other languages in that factive *că*, not the non-factive *să* is used.

For Romani I was not able to find other than the third person forms combined with *mek*. This restriction may be caused by the fact that the simple subjunctive has a variety of functions, including that of the adhortative. However, in other dialects constructions of *mek* with 1sg can be found (see below). Examples from Kalderaš (Boretzky 1994a):

(15) muk te żan ande kutari kutari plajin
       'let them go to that and that mountain' (adhortative)

(16) muk te żantar, phenkə, von i kagja mule
       'may they go, he said, they are dead (i.e. they will die) anyway' (permissive)

(17) mek te del o del kagja i majangle
       'may God arrange it that way in future too' (optative)

Whenever *mek/muk* appears in combination with other persons, like in (18), it must be understood as a request addressed to another person rather than to the ego or to the speaker's own group; for the latter, the mere subjunctive would suffice; cf. (19):

(18) mek te żas khərc 'let/permit us (to) go home' (1pl.)

(19) te żas khərc 'let's go home'

Uhlik (1974, 76) gives some examples for the 1sg that are difficult to translate, because the antecedent clauses are not real conditionals:

(20) kana dži kathe peradilem, muk mardivav jodži dži agoreste
       'if I have been turned down to this point, let me perish completely'

(21) te xasardem mo ćhelipe, mek xasardem les, naj mae khonik bango
      'if I have spoilt my playing, let it be spoilt, nobody else is responsible for it!'

A better example can be quoted from Slovakian Romani (Lípa 1963, 131):

(22) mi koral'uvav, te na čačipen, te pheras kerav!
      'let me become blind, if it is not true, if I am joking!'

In Kalderaš, there seems to exist a kind of concessive construction in addition to the permissive and adhortative ones, but since the sentences are elliptic, with the main clause lacking, we cannot be sure if this is a standard function of *mek*:

(23) ma mek tu kɔ popravisajli, kɔ naj više kodeja so sas, ...
      approximately: 'but even if she mended her ways, if she isn't any longer what she was, ... (to be added: 'even that wouldn't help')

(24) ma mek tu, phrala, phenkɔ, i te sutjan (sc. munřɔ gaźasa), ...
      'but even if you slept (sc. with my wife), brother, - he said - (to be complemented by: 'even that wouldn't matter')

In view of the fact that this dialect is heavily influenced by Rumanian, it is very likely that the model for the constructions quoted here has been Rum. *lasă că*, the more so as *mek ke* (with factive 'that') instead of the more widespread *mek te* has been used.

    Mek has been reduced to *mi* in Slovakian Romani and in Southern Poland—a normal process in progressive grammaticalization (cf. Lípa 1963, 131):

(25) u amaro čavo paš amende mi bešel
      'and our son may stay with us' (Kopernicki 1930, 87)

It is an open question if negated **mek te na avel* is possible, as for instance in Greek (cf. ἄς μήν ἀπελπιζούμαστε 'let us not despair!') and in other contact languages.

2.3. Modals and modal constructions

Modal auxiliaries and particles have been borrowed from the contact languages into a variety of Romani dialects. Besides the inherited *kam-el te* 'love, wish, want' and *mang-el te* 'want, ask' for volition, *šaj* 'can'/*našti (te)* 'cannot' for possibility and *džan-el te* 'know how to' for ability, *si te* 'must' for necessitiy and some other periphrastic expressions like *kampel* 'have to', a lot of direct loans are in use:

možinel and *dali* 'can', *trom-al* (in all dialects) and *darfte* / *tref te* 'dare to, to be allowed to', *birin-el*, *sabadno*, *hodno te* 'to be able to', *trubul* / *trebola* / *treba te* 'it is necessary', *mora*, *musaj/moste* and *prepi* 'must', *valjazla* 'it is appropriate'.

The multitude of direct loans gives reason to suspect that even among the expressions formed by Romani elements some may be of foreign origin, calqued from Greek and other languages, e.g. the constructions *si te*, *džan-el te*, *kampel*, as well as a peculiar dative construction met with in various Balkanic dialects:

(26) so hala pes tuke?
     'what would you like to eat?' (Bugurdži, Boretzky 1993a).

Doubtlessly, this is a calque of Serbian *ne jede mi se* 'I don't want to eat' and the like.

Since the means for expressing modalities have been described in detail in Boretzky (1994b; 1996b), nothing more will be said here on this subject.

Conditional sentences are constructed in Romani dialects according to a variety of models, some of them giving the impression to be calqued, too. This is most probable for one construction found in Greek and in Macedonian Romani being made up of the future particle plus the imperfect tense, e.g. *ka keravas* 'I would do/would have done', Greek θά έκανα (for the details cf. Boretzky 1993b).

2.4. The formation of the prepositional system

The Romani noun inflection is characterized by a system of eight cases that are nominally identical with those of Old Indic, but do not continue the old case forms, most of the case morphemes going back to former postpositions. In addition, European Romani developed a rich system of prepositions from former locative adverbs and to a lesser degree from locative nouns. Old postpositions are not extant in any dialect, but in Finnish Romani there are tendencies to convert prepositions into postpositions, apparently under the influence of the contact language.

As far as we can see, the prepositional system was created by the concurrence of some factors, but it is completely clear that it was not achieved by a shift of old postpositions into pre-nominal position. First it has to be emphasized that even in Modern Indic languages at least some prepositions can be found. There is an ambivalent case in Hindi and Punjabi, where *bina* functions either as a post- or a preposition, cf. for Hindi *N-ke binā* or *binā N-ke* 'without N'.

Also, Dardic languages of Northwest India, which are said to have played a role in the formation of Romani, have a few prepositions that can even be agglutinated to the noun (cf. Edel'man 1983, 306). Therefore, it cannot be excluded that a few prepositions of Romani directly originate from old prepositions. This may be the case with *bi* 'without' and *vaš* 'because of, for', which are unusual in governing other cases

than nominative/locative: *bi* is constructed with the genitive, *vaš* with the dative. (Perhaps the old adposition *andar* 'from, out of' < OI *antarāt* 'id.' was in use as a preposition, too.) It is possible that for some time *bi* could occur in both positions, i.e. *N-GEN bi* and *bi N-GEN*. Under the influence of Persian, which uses *bī* (older *bē*) as a preposition (and as a prefix) it might have been consolidated in prenominal position, but this assumption would not explain why *bi* governs the genitive.

The prepositional system of Persian (see Jensen 1931, 179ff) shows some similarities with that of Romani, but we cannot claim with certainty that it exerted an influence on Romani. We will give here some details together with their Romani parallels:

(a) There are a few old prepositions combined directly with the noun, as *az* 'from', *ba* 'in, at, on', *dar* 'in', *bī* (*bē*) 'without', *tā* 'to, till', which are roughly paralleled by the Romani local cases and related categories as well as by the prepositions *ke* (*kia*), *tar* (*kotar*) and *te*, the latter being etymologically related to the case morphemes *-ke* (dative), *-tar* (ablative) and *-te* (locative). Since except for *-rā* there are no postpositions in Persian, there is a slight possibility that this or closely related languages may have triggered the development towards a prepositional system in Romani.

The rest of the Persian prepositions are construed together with the so-called Izafet particle.

(b) Some of these elements stem from locative adverbs or nouns, as *pēš* 'front, foreside', *zēr* 'lower part', *pahlū* 'side', *tū* 'inner part, middle', *bālā* 'above, up'. When Romani came into contact with Persian or related languages, it had elements very similar in character to the Persian ones: the adverbs *upre/upral* 'above', *tele/telal* 'below', *angle* 'in front', **pale/palal* 'behind', and the nouns *paš* 'side' and *maškar* 'middle'. We can imagine, therefore, that the Romani speakers identified their elements with the Persian ones and began to use them not only as adverbs but as prepositions. We have to take into account that the eight cases of Romani did not suffice to express all the meanings conveyed by the adverbs/prepositions listed above.

(c) Some Persian elements are derived from other nouns, especially from terms of body parts: *sar* 'head' > 'on, above', *rōy* 'face' > 'on', *pušt* 'back' > 'behind'; without Izafet construction *bar* 'top of the head, height, bosom' > 'on, over'. For them, no corresponding elements can be found in Romani (perhaps with the exception of *ma-muj* 'opposite' < *muj* 'face, mouth').

As we have seen, there are some parallels with Persian, but the similarities between Romani and Greek are by far more conspicuous. It has to be pointed out that Greek underwent a far-reaching change. Whereas the Old Greek system displays typically Indo-European structures with one preposition governing more than

one case and having a couple of quite different meanings, Modern Greek differs radically from the old language: it preserved only a few simple prepositions, and these correspond to the cases or to some simple prepositions of Romani. All other meanings are expressed by secondary prepositions composed of locative adverbs (of the type 'near, below, above') plus one of the basic elements σέ, roughly 'in', and ἀπό, roughly 'from'. In the course of change σέ became a neutralized, semantically vague element capable of expressing a variety of local and purely grammatical relations (e.g. as a dative marker).

Now, much speaks in favour of the assumption that the secondary prepositions of Greek served as a model for Romani. Romani had locative adverbs corresponding to those of Greek, and since in Greek these adverbs continued to be used as adverbs, it was possible to identify with them the Romani adverbs and create a set of new prepositions. This was only possible with Greek as a model; the Slavic languages preserved the simple prepositions of the Indo-European type, which could not be translated or indirectly imitated. In what follows we will give a detailed comparison of Greek and Romani.

(a) simple prepositions of Greek and their equivalents in Romani (Table 2):

Table 2
Primary prepositions

Greek	Romani	Greek	Romani
γιά 'for'	= dative	σέ 'in; dative'	= locative/dative
μέ 'with'	= instrumental	ὡς 'to, until'	= *dži*
ἀπό 'from'	= ablative or *katar* 'from' or *andar* 'from, out of'	χωρίς 'without'	= *bi-* (with genitive)

In terms of function, these are the abstract grammatical cases as well as the basic local cases. The genitive of Greek (at least in the sg.) cannot be replaced by prepositions, and that is probably the reason why it was preserved in Romani too. As for the dative, it has been abandoned very early in Greek, but it is in full vigour in Romani as a grammatical case. There are nearly no tendencies to substitute prepositions for it.

(b) prepositions derived from adverbs in both languages (Table 3):

Table 3
Secondary prepositions

Greek	Romani
μέσα 'in it' > μέσα σέ 'in'	= andre > ande/an/dre
ἀνάμεσα 'in between' > ἀνάμεσα σέ 'between'	= maškar 'middle' > maškar
κοντά 'near' > κοντά σέ 'near' and πλάγι 'side, beside it' > πλάγι σέ 'at, near'	= paš 'side'/paše 'beside it' > paš(a) 'at'
πάνω 'above' > πάνω σέ ἀπό 'on, above'	= opre/opral > pe/ap/upr
κάτω 'below, down' > κάτω ἀπό 'under'	= tele/telal > tal(al)
μπροστά 'in front' > μπροστά σέ 'in front of'	= angle/anglal > angla
πίσω 'behind' > πίσω ἀπό 'behind'	= palal > pal(a)
γύρο 'around' > γύρο ἀπό 'around'	= trujal > trujal
πέρα 'over there' > πέρα ἀπό 'beyond'	= perdal (adv.?) > perdal

As can be seen from this comparison, the special local cases of both languages agree in detail. In Greek, the simple preposition ἀπό plays an important role as a second constituent, and it may be therefore that in Romani the new prepositions have been derived from the ablative adverb in -al rather than from the locative adverb in -e. It is very unlikely that there exists a Modern Indic language which would have a comparable set of adpositions, and that makes Greek influence highly probable. When these Romani prepositions came into being under Greek influence, no restructuring from a postpositional to a prepositional system was necessary, simply because the elements under discussion were innovations. As for the prepositions tar, ke and te, corresponding etymologically with the ablative, the dative and the locative respectively, there remains a diachronic problem. We do not know whether they once were postpositions or adpositions that could be used both before and after the noun, but even if they became established as postpositions, they might have changed their position under the new circumstances. That this is possible can be demonstrated by a syntactic change of precisely this kind in Finnish Romani (see Valtonen–Gilliat-Smith 1967).

The temporal prepositions μετά and πρίν ἀπό are rendered in Romani by the originally local pala (palo) and anglal. Overall, then, there is an equivalent in Greek for each Romani secondary preposition, the most striking difference being that in Greek the adverb must be combined with a simple preposition (σέ or ἀπό, which in many instances are no longer in semantic contrast). At a first glance, it may seem that Romani has no equivalent for this σέ, but we have to take into account that the secondary prepositions of Romani govern the locative (today with pronouns and sporadically also with nouns, but earlier perhaps throughout), which

is a good equivalent for the semantically bleached σέ. Thus, with all the differences on the expression side there is a considerable functional similarity.

In both languages the distribution of simple and complex expressions is in agreement with the assumptions of markedness (or naturalness) theory: the less marked categories have simpler or shorter expressions, and vice versa. Therefore, universal factors may have played a role in forming this distribution (cf. Mayerthaler 1981).

It is possible that some Romani dialects present a transitional stage closer to the Greek original. In the texts published in Paspati (1870) the simple new prepositions seem to be lacking, complex constructions being used in their stead. The first elements preserved the full, unreduced adverbial form, e.g. *andre* instead of *ande / and / an* or *dre*, which shows the ancientness of the construction; the second element is *ke* or *te*, i.e. old prepositions (etymologically identical with dative and locative respectively); cf. *anglal t-i rakli* 'in front of the girl'; *andre t-o saraj* 'into the palace'; *mamuj t-o dakar* 'vis-à-vis the king' (598); *katar k-o maškareder* 'to the second (brother)' (600); *palal t-o raklo o mulo* 'behind the dead boy' (602); *opre k-o česmes* 'upon the fountain', *katar t-i flori* 'out of gold coins' (606); *telal t-o šeran* 'under the pillow' (614); with old *dži: dži t-o koča* 'up to the knees' (610).

In the Paspatian dialect, this seems to be the only possible construction; I was not able to find simple *andre / telal* etc. + noun. To be sure, the distinction made in Greek between σέ and άπό cannot be rendered by Romani *ke* and *te*, but apart from this both languages do behave very similarly. The dialect of the Izmir Sepetčides (Heinschink–Zambakli-Heinschink 1994, 2), similar to Paspati's variety in many other details too (prep. *te!*), shows the same constructions: *talal k-i prošik* 'under the lawn', *andre k-o džuva* 'among lice'; *andre k-o mela* 'amidst the filth'; *katar k-o them* 'out of the country', *ko maškar k-o them* 'to the center of the country'. [According to Windfuhr (1970, 277), in Persian Zargari the local adverbs seem to be used as postpositions (*anri* and *opre* with locative, *avri* with ablative, and *andama* 'together' < άντάμα with genitive). Cf. also postpositional *andre* in *duj masekende andre* in Paspati (1870, 626). Is this due to Turkish influence?]

The fact that the new prepositions stem from original nouns or adverbs combined with older prepositions, was already noted by Miklosich (1881, 67) and by Sampson (1926, 221), but those authors did not relate the rise of the new prepositions with the structures of the Greek system.

There is another preposition in Romani that probably owes its existence to foreign influence: *kaj, kate*. Originally *kaj* was an interrogative, 'where?', and *kate* an adverb and in some dialects an interrogative, 'here' and 'where?', but now both acquired additional funcions: *kaj* is in general use as 'at, with', and *kate* occurs with the same meaning in the Arli dialect. This double function is paralleled by Bulgarian (and Macedonian); cf. Bulg. *kăde (kaj)*, Maced. *kade (kaj)*, being both

interrogatives and prepositions. The short forms are identical with Romani *kaj*, and Slavic *kade* is similar to Romani *kate*, and it is probably this similarity that made interference easier. Moreover, *kaj* is also similar to the old preposition *ke* (*kia*).

In practically all dialects prepositions have been directly borrowed from the respective contact languages. Since we are interested here in processes of calquing we will not elaborate on this topic, but one special case should be mentioned. Varieties of Romungro have a preposition *miste* 'for, because of', which goes back to Slavic *město* or rather to the PP *na městě* 'instead of' (the *i* in *miste* pointing to an Ikavic dialect of Serbo-Croatian as the immediate source). Since in Slavic this secondary preposition never occurs with the meaning 'for, to the benefit of', we have to look for another explanation for this meaning. Perhaps, this can be found in the double meaning of the old Romani preposition *vaš* occurring in Central dialects with both meanings: the second meaning of *vaš* was transferred to the loan *miste* because the two elements were associated by one common meaning—an analogical extension. For *vaš* 'instead' cf. Hübschmannová *et al.* (1991):

(27) gejl'as odoj vaš mange (= Czech šel tam naměsto mne)
      'he went there in my place'

An example for *miste* 'for' (Knobloch 1953, 30):

(28) akor ništa ovla te počinel miste o siklibe
      'then nothing will have to be paid for the instruction'

In some dialects the PP *pe than* / *po than* / *ko than* + genitive is used for 'instead (of)'. This is in all likelihood a translation of Slavic *na městě* or *na-mesto* or simply *mesto*, as in Serbo-Croatian. Also, Rumanian *în locul* + genitive may have served as a model. However, this calque appears not to be a really popular element, since instead of it loan-words are preferred. In Hübschmannová *et al.* (1991) it is lacking, and in the dialects examined by the author it is rare, giving the impression of a conscious *ad hoc* translation.

A similar process has taken place in the word formation of Sinti, where the particles (the movable "prefixes") of German verbs have been translated into this dialect (cf. Igla 1992).

2.5. New complementizers

There are only a few old elements in Romani (*thaj*, *vaj*, *vi*, *te*), most complementizers being borrowed from the contact languages. Among the indirect loans some are translated, some others consist of a translated and a directly borrowed constituent.

2.5.1. Relative clause markers

As is frequently found in the languages of the world, the interrogatives function as complementizers too; cf. *kon* 'who?' > rel. 'who; which', *so* 'what?' > '(that) what', *kaj* 'where?' > 'where; which; that', *kana* 'when?' > 'when', *keti* and similar forms 'how much/many?' > 'how much', *savo* 'which?' > 'which'.

Since transitions like these are nearly universal, they need not have come about in Romani as late as after the contact with the Balkanic languages, but may have existed potentially, as *ad hoc* possibilities, for a long time. The situation is different, however, with *kaj* in its function as a general relative marker 'which'. Here, we must not disregard the functional identity of *kaj* with Greek πώς, where both elements appear with both meanings. Therefore, it is most probable that *kaj* acquired this function under the influence of Greek, which is preserved in nearly all dialects:

(29) ehin jek kupcos, kaj les ehin dešuduj sklepi
     'there was a merchant who had twelve shops' (Kopernicki 1930, 87)

The oblique case is marked indirectly by the inflected personal pronoun *les*.

Of the Central dialects Slovakian Romungro has *so* 'what' with broad relative function:

(30) o raj so o love našadja, dinja ari te hangozinen ...
     'the gentleman who lost the money, made public ...'

(31) o love so zakernas pr'o bijav
     'the money (which) they earned at the wedding' (Lípa 1963, 135)

Normally, in the Slovakian translations the relative pronoun *ktorý* occurs for *so*, but in popular speech *čo* 'what' (and in Czech *co*) can also be found, and this seems to be the immediate model for the Romani relative marker; cf. from the Czech literature:

(32) lampa ještě hořela, co nad klekadlem visela
     'the lamp that hang over the desk was still burning'

and in Slovakian:

(33) tento Bartoš, čo šantí na dvoře, je najmladší
     'this Bartoš who is playing in the yard, is the youngest'

(34) niet takej pesničky, čo by konca nemala
     'there is no such song that wouldn't have an end'

2.5.2. Temporal clause markers

The use of *kana* 'when?' for temporal 'when' may be an indigenous development, since this transition has countless parallels in other languages. However, expanded *kana-tu* (as well as *kon-tu* 'who', *so-tu* 'what' etc.) found in Kalajdži and Drindari can have only been formed by copying the Bulgarian model, especially *koga-to* 'when', derived from *koga* 'when?':

(35) ti kana-tu nakhinas-li manuša opra phrucjatar ...
      'and when(ever) the people passed the bridge, ...' (Kostov 1973, 107)

and further *kon-tu* following Bulgarian *koj-to* in

(36) kon-tu naj-erkin aila, lá mɔ-mukin andu temej
      'who (sc. of the wives) comes first, they will let her down into the foundations
      (sc. of the bridge)' (Drindari; see Gilliat-Smith 1962, 127)

It is another question whether *kana-tu* has been formed on the basis of a pre-existing relative *kana* 'when' or whether it turned up as a relative immediately in this shape.

Also, *so-m* 'when, as soon as' of Erli is an imitation of Bulgarian *što-m*; cf.

(37) som aljan khere ka čhinav tut 'as soon as you come home, I'll kill you'

Many dialects of the former Yugoslavia as well as Central dialects use *sar* 'how?' for temporal 'when, after, as soon as' and for conditional 'if'. Instances of temporal *sar*:

(38) katar kada vakci sar čorde mange o pasoši
      'from the time (when) they stole my passport' (Bugurdži, Boretzky 1993a, 100)

(39) rakh že adatre mirija romnja, syr lela te bijanel
      (= Russian smotri že zdes' za ženoj mojej, kak ona rožat' stanet)
      'look after my woman, as soon as she begins to bear' (Dobrovol'skij 1908, 14)

(40) sar les mudargjas, gelja kia leste
      'after he killed him, he went to him' (Kopernicki 1930, 4)

(41) ta sar amen dine oda grastano mas, ta gejljom andro pani, kaj les te thovav avri
      'and when they gave us that horse meat, I went to the water in order to wash it'
      (Romano Džaniben 1994, 1; 27)

For this use of 'how' we find close parallels in Greek and in the Slavic contact languages:

(42) σάν βραδυάση θά φύγω 'when/as soon as it darkens, I'll leave'

(43) τόν συνάντησα (κατά) πώς ἔβγαινε ἀπό τό σπίτι
     'I met him when he left the house'

(44) Serbo-Croatian
     kako me ugledao, on mi priđe
     'when he saw me, he came nearer'

(45) Polish
     jak żyję, nie szłyszałem
     'as long as I live I didn't hear (sc. something like this)', and

(46) jak pódziesz, to wstąp do mnie
     'when you arrive, drop in (to me)'

(47) Czech
     jak to pověděla, duši vypustila
     'after having said this she breathed her last'

It is not quite clear if even *so* 'what?' used temporally is due to interference:

(48) ratjaha, so gelja tele pal o vodro, cidinja len po pre
     (= Slovakian ráno, len čo zliezol z postele, obul si ich)
     'in the morning after getting out of bed, he put on (sc. his shoes)'
     (Romano Hangoro 1993, 55)

2.5.3. Conditional clause markers
In these constructions *sar* is used too:

(49) sar amenge denašeja, amen tu tosarla livinaha
     'if you run away, we will shoot you tomorrow' (realis; Kopernicki 1930, 3)

(50) sar me tutar denašavas, ta man mindre duj phralora livinenas
     'if I had run away from you, my two brothers would have shot me' (ibid. 4)

Parallels from the contact languages:

(51) Greek

σάν θέλεις όλα γίνονται
'if you want, everything will come true'

(52) Czech

jak odejdeš, budeš potrestán
'if you leave, you will be punished'

(53) Polish

jak będę mógł, dam ci pieniądze
'if/as soon as I can, I will give you the money'

Note that in Slovakian *ak* 'if' has been derived from *ako* 'how?'.

2.5.4. Object clause markers

Even the common Romani *kaj* 'that' introducing factive object clauses has its parallel in Greek ποῦ 'where?; that' (cf. above ποῦ and Romani *kaj* for 'where?' and relative 'which'), although in Greek ποῦ is less frequent than πώς and ὅτι. Since this use of *kaj* is so well known we can dispense here with examples.

Less frequent is the use of *so* 'what?' for 'that'; since it occurs in dialects spoken in Russian and Slovakian and Czech linguistic environment, there is little doubt that it has been equated by the Romani speakers with Russian *čto*, Czech *co* and Slovak *čo* occurring with the same combination of functions; cf.

(54) dykhela, <u>so</u> leste štar gara

(= Russian i smotrit, <u>čto</u> u stola četyre nogi) (Dobrovol'skij 1908, 5)
'he sees that there are four feet at it/at the table'

(55) a me dumali, so rašaj 'and I thought that (it was) the preast' (ibid. 7)

From Slovakian Romani:

(56) so tut pal leste, so jou odova kerlas

'what does it matter to you that he has done it?'

which corresponds functionally to Czech

(57) je tomu 14 dní, co odjel
    'it's 14 days (ago) that he departed'

2.5.5. Causal clause markers
Romani *sar/har* can even be used for introducing causal constructions:

(58) he har sas o dilino solaxardo, xudin'as e god'averes he čid'as les avri a xevaha
    'and since the stupid one got angry, he grabbed the clever one and threw him out of the window'
    (Moravia; Mann 1947, 28)

There appears to be an overlapping in *har* between temporal and causal senses, which is also true for English *since*. We have to do here with nearly universal transitions between the functions mentioned, and this weakens the interference argument. Nevertheless, 'how' can be used in West Slavic languages with this sense; cf. from Czech:

(59) ještě se mi kolena třesou, jak jsem se lekla
    'my knees are still shaking because/as I was scared'

The agreement between *soske* 'why?', dative of *so*, > 'because' and equivalent Greek γιατί, Bulg. *zašto* (Maced. *zošto*) and to a lesser degree with Alb. *(se-) pse* < *për se*, all meaning literally 'for what?', is greater and more convincing, since two elements are involved.

2.5.6. Final clause markers
In Kalderaš, *sar te* 'that, in order to' (literally 'how that') is widely used with final function in the two Yugoslavian variants described by the author (Boretzky 1994a). It is structurally similar to Serbo-Croatian *kako bi*, but much closer to Rumanian *ca să* which, apparently, has served as a model for this Vlach dialect:

(60) i vo cərdel les sar te šinel lesko šoro
    'and he dragged him (away) in order to cut off his head'

(61) mekəl palal o praxo sar te kərəl jek trago
    'and he strews the ashes behind himself in order to leave a trace'

We find complex conjunctions in other dialects as well. Since one of the constituents is a direct loan, the origin of the constructions need not be discussed; cf. *hotj / hoj te* in Hungary. Hungarian *hogy* has the primary meaning of 'how' and

'that', i.e. *hotj te* is equivalent to *sar te* of the Kalderaš dialect both structurally and semantically; an example from the Vend dialect:

(62) soske na mukes hodj panj te pijav
      'why don't you allow (me) to drink water?' (Vekerdi 1984, 76)

There is another equivalent to *sar te* and *hotj te* in Macedonian dialects: *za te*, the model being Macedonian *za da*:

(63) ratilo; avile jekhe vešeste on dujdžene phrajlja za te soven
      'night has fallen; the two brothers came into a forest in order to sleep there' (Prilep)

In Russia, another complex marker, *so-b* 'that, in order to', occurs. There can be no doubt that it is formed after Russian *čto-b(y)*; cf.

(64) u dylyno priphendja, soby čaven na len
      'the fool has said that they should not take along their children'

(65) te prizlyžal sop te mar (!) o rom e romnja
      'that he should make the man beat his woman' (Dobrovol'skij 1908, 8, 10)

As far as I can see, there is no direct equivalent for the *kaj te* 'that, in order to' of the Central dialects to be found in the respective contact languages, since Czech *a-by* and Polish *że-by* display a different structure.

2.6. The definite article

We do not know to what degree an article has been in use in Romani prior to the contact with Greek. In Nuri, there are the elements *uhu, ihi, ehe*, which served as demonstratives and perhaps as articles too, but they are not very frequent (cf. Macalister 1914, 8). Sure, these elements must be affiliated in some way with the article of European Romani, *o i e*, but from this it does not follow that they had article functions in early Romani. Therefore, it cannot be excluded that a definite article as an obligatory category developed only under the influence of Greek, the more so that at least masc.sg. *o* and fem.sg. *i* were identical in both languages. On the other hand, we have to take into account that the Greek article shows no similarity with αὐτός and ἐκείνος, the demonstratives of Modern Greek, and this is to say that Greek did not provide the Roma with a model for forming the article on the basis of the demonstrative 'that' (as is normally the case in languages having developed an article).

2.7. Superlative formation

Romani has an inherited comparative ending in *-eder*, which is in full vigor in Central dialects and in dialects spoken north to it, whereas it is in retreat in the Balkanic dialects and in Italian Romani where loan elements are applied instead. This distribution of conservative and innovatory behaviour appears to be conditioned by the respective contact languages: West Slavic languages, German and Hungarian make use of inflectional means (suffixes) that cannot be easily borrowed or copied in an indirect way into other languages. On the other hand, the comparative particles of Rumanian, *mai*, of Italian, *più*, of Greek, πιó, of Bulgarian, *po* for the comparative and *naj* for the superlative, present no obstacles to borrowing, which enabled (but did not make inevitable) the replacement of old *-eder* in the Southern dialects. An indirect calquing of the periphrastic comparative was not possible because there is no equivalent for 'more' in the Balkanic dialects, the formal comparative *but-er* being restricted to the Central dialects. To my knowledge, *buter* has not been utilized for marking the comparative in any dialect, but it can be added to an old comparative as reinforcement. The distribution of old and borrowed elements gives reason to assume that the foreign particles were not borrowed because the older means of expression were no longer available or because the foreign elements had properties more advantageous than the old ones, but simply because they were at hand and could be adopted without difficulties. This view is supported by the fact that in northern dialects *-eder* is even applied to borrowed adjectives; cf. in Dobrovol'skij (1908) *daludyr* 'farther', *prytkedyr* 'quicker', *gromkedyr* 'louder', and in the Slovakian dialect *goreder* 'worse' (Lípa 1963, 80).

Whereas the abandonment of the old comparative morphology has no reasonable motive, interference in the realm of the superlative has to be judged differently. As in Romance languages, it had no form of its own, and the shaping of a special expression made it possible to distinguish between comparative and superlative unequivocally even without further context. Therefore, it should not be by chance that perhaps all Romani dialects that had access to a foreign superlative marker, did borrow it. In West Slavic languages, Latvian and Hungarian, the superlative is derived from the comparative by a preponed particle, and this has been copied in the respective Romani dialects; cf. *naj-terneder* after Slovakian *naj-mladši*, *vis-baredir* after Latvian *vis-lielākais*, and *leg-terneder* after Hungarian *leg-ifjabb*; (the particles *nek-* and *jekh-* seem to be derived from *leg-* in an internal process). In Russian, it is *samyj*, roughly 'same, very', combined with the positive, and this too can be found in Romani (cf. Eloeva–Rusakov 1990, 17).

Expressions for the superlative may have been borrowed earlier than those for the comparative. This can be inferred from the appearance of reinforced constructions in the Balkanic dialects: we have *naj-bareder* in non-Vlach dialects, which is

an unambiguous superlative. Later on in *naj-bareder* the double marking was simplified to *naj-baro*, and a new comparative *po-baro* was added. The Vlach dialects, however, did not come to systematically differentiate between comp. *bareder* and superl. *maj-bareder*, apparently because this distinction is not made in Rumanian.

So far no calquing of foreign superlatives has been mentioned, and direct translations of foreign markers were not possible, since in Romani no equivalent for 'most', 'utmost' or similar elements is available. Some formations, however, give the impression of having been created under the influence of foreign superlatives, even if no element has been translated in the normal sense of the word.

(a) Rozwadowski (1936) gives forms containing *nok-*; if this is not a variant of *nek* < Hungarian *leg-*, it might be taken from or at least secondarily influenced by German *noch* 'still'. This would be a direct loan, but what is peculiar about it would be that it does not copy a German construction, since in German *noch* can only be used as reinforcement of true comparatives (not superlatives!). Thus, *nok-feder* might continue German *noch besser* structurally, but not functionally. It must have become a superlative by an endogenous process.

(b) For Bohemia, an element *balo* has been recorded (Ješina 1886), e.g. *balo choreder*. Pott (I, 1844, 210) has *bala*, which he compares with Hindi *bhalā* 'good'. The origin of this element remains unclear.

(c) There is another opaque element recorded for Bohemia, *kon-o/kon-i*, and for Finnish Romani, *koni*; cf.

(66) koni pxureder romni 'the oldest woman'

To this *koni* another element found in Sinti seems to be related: *one* in Slovakian Sinti (Lípa 1965, 32f) and in the Lalere dialect now spoken in Western Germany; cf. *one bare-den* 'the greatest' (Holzinger 1993, 55f).

While it is true that the creation of this new category was triggered by the contact languages, it cannot be said to what degree Central European models have been copied, since the origin of the markers remains unclear. Valtonen (1972, 64) analyzes *koni* as *kon* 'who' + *hi* 'is', and if this is correct, the gender-inflected *kon-o/kon-i* must be explained as a reinterpretation of the copula constituent, whereas *one* is a reduction of the grammatical marker *koni*, a process to be observed quite universally. The geographical distribution points to a rather old element, which must be seen as an endogenous development, independent of language contact (see also Pott I, 1844, 211).

(d) In Wales *buteder* 'more' is sometimes used to form a superlative on the basis of the old comparative or the positive (Sampson 1926, 151). This procedure, too, does not render English *most* + positive in a direct way.

2.8. The 'new' infinitive

It is not clear whether Romani had an infinitive or not when it first got into contact with Greek, but since there are no traces of infinitive morphology in the conservative Balkanic dialects, we are not in a position to clarify if an inherited infinitive was abandoned under Greek influence or if this category did not exist in pre-European times. However this may be, the infinitive met with in Central and northern dialects of Romani is a late creation formed by interference with Slavic languages, Hungarian and German. We are compelled to draw this conclusion in view of the fact that in Romani an infinitive is present only in those contact regions where an infinitive does exist and, moreover, is in frequent use. Infinitive morphology is dervided from finite subjunctive constructions of the type *kam-av te l-av*, in which the subjunctive agrees in person and number with the governing verb. A non-finite form was created by generalizing one of the personal forms of the subjunctive: *kam-av te l-el* (3sg); *kam-av te l-en* (3pl or 2pl); *kam-av te l-e(s)* (2sg).

Morphologically, these forms are totally independent of the infinitive forms of the contact languages, the reason being that for the lack of comparable elements in Romani the bound infinitive morphemes of Slavic and other European languages could not be translated or imitated in a similar way. Nevertheless, distribution and particular functions in the dialects do not leave room for an explanation other than by interference. This is a good example for demonstrating that calquing need not be accompanied by copying morphological rules.

For further information see Boretzky (1996c), where historical problems, distribution, forms, functions etc. are treated in detail.[7]

2.9. Perfect and pluperfect active in South Balkanic dialects

Romani differs from the Balkan and other languages in that it does not distinguish a simple unmarked past from a marked (resultative) perfect, and this state has not been changed in most of the Balkanic dialects of Romani. There is one exception, however, in dialects spoken in Macedonia. Potentially, all dialects are capable of forming a perfect passive or a stative by combining the participle with the copula, e.g. *kerdo si* 'is made' or 'has been made'. A perfect active might have been formed in the same way if there were a verb for 'have', thus providing equivalents for e.g. English *has made* or French *a fait*, but since a word for 'have' was not available, the perfect active of Greek or Bulgarian could not be imitated by simple calquing. Despite these difficulties a perfect active (and a pluperfect active) turned up in the dialect of Prilep and perhaps in varieties of Macedonian Arli, but it is identical in form with the perfect/pluperfect passive. With intransitive verbs, this identity did

---

[7] Cf. also Pott (1844, 329f), where early sources are collected and discussded, and Soravia (1978).

not create confusion because intransitive verbs do not normally form passives, but with transitive verbs the context must be clear enough in order for the hearer to decide what is meant. It is interesting that, although the Prilep dialect does have a special verb for 'have', *ther-el*, it does not make use of it to copy Macedonian constructions as *imam zemen-o* 'I have taken' (trans.) and *imam dojdeno* 'I have come' (itr.), but it has imitated Macedonian constructions formed with the copula 'is' like *e dojden* lit. 'he is come' or *e umren/-a* lit. 'he/she is died':

(67) nasine alo o doktori 'the doctor hadn't come (yet)'

(68) ma te si nekoj mulo 'somebody has died, perhaps?'

As has been said before, the sentences are completely clear because the perfect is formed from intransitive verbs. This is not the case, however, with transitive verbs, which are also found in these constructions:

(69) sinan havdo? 'did you understand?'

(70) sigo sinum bisterdo 'I have forgotten it quickly'

(71) me sinum tumenge vakerdo 'I have said to you'

If the context does not suffice to disambiguate such expressions, (70) might also be understood as 'I have been forgotten quickly'. The speakers put up with this difficulty, apparently because in Slavic Macedonian dialects both auxiliaries are confused and, as bilinguals, they became used to this confusion; the following examples taken from Koneski (1967, 178f and 220f) have 'have' instead of 'be' even in passive constructions:

(72) vo Ser *imalo* ubien i eden srpski car
      'in Serres a Serbian Czar had been killed'

(73) vo toj grob *imalo* nekoj derviš zakopan
      'in that grave a derwish had been buried' (*imalo* is literally neutral 'it had')

In the same way, transitive verbs may be constructed together with 'be', e.g. verbs of eating as *sum jaden(a)/večeran(a)/poručan(a)* 'I have eaten/I have taken supper/I have taken breakfast', and other verbs:

(74) petlite se peani
     'the roosters have crowed'

(75) da ne ste zaboraveni kaj ste ja ostavile kleštata
     'didn't you forget that you left behind the pliers?'

In all likelihood, these were the immediate models of the Romani perfect/pluperfect active.

Thus the tense system has been expanded via interference at least in this particular dialect. The old pluperfect active of the type *ker-d-um-as* 'I had done' is falling out of use, being replaced by the new *sin-um-as kerdo*.

2.10. The Bulgarian narrative (evidential) in Romani

Bulgarian as well as Turkish have an elaborate system of narrative forms expressing events and actions that are not directly witnessed by the speaker. According to Kostov (1973, 107f) this distinction between direct and indirect statements has been transferred to some Bulgarian Romani dialects too but, similar to the infinitive category, not by simple translation. The Bulgarian morphological system is characterized by the principle that each narrative category is pushed one more step into the preterit than the corresponding direct category; cf. Table 4:

Table 4
Narrative in Bulgarian

	direct	narrative
pres.	*xodja*	*xodel săm*
ipf.	*xodex*	*xodel săm*
aor.	*xodix*	*xodil săm*
perf.	*xodil săm*	*bil săm xodil*
plqu.	*xodil bjax*	*bil săm xodil* etc.

This intricate system of dependencies based upon more than one principle could not be copied in all details, mainly because there was no equivalent for the Slavic *l*-participle. According to Kostov, the Roma extracted the *-l* from this form and added it as *-li* to the Romani imperfect, which resulted in the dialect of Sliven in the following system (Table 5):

Table 5
Narrative in Bulgarian Romani

pres.	*kerava*	*keravas-li*
ipf.	*keravas*	*keravas-li*
past	*kerim*	*kerimas-li*
plqu.	*kerimas*	*kerimas-li*

If this is correct, Bulgarian and this Romani dialect would agree in not discriminating between narrative present and imperfect and between past and pluperfect. According to Kostov the following sentences are in contrast:

(76) oda vakirjas mangi či tu phirsa
      'he told me that you (will) go' (directly witnessed)

(77)   oda vakirjas mangi či tu phirsas-li 'id.' (indirect statement)

Tales are told in Bulgarian most often as if the speaker knows about the events reported only from hearsay, and this can be copied in Romani (Kostov 1973):

(78) Romani:
      ti kana-tu nakhinlas-li manuša opra phrucjatar, šunsejlas-li racjasa: Pavljo, Pavljo!
      Bulgarian:
      ti koga-to minavali xora po mosta, čuvalo se prez noštta: Pavljo, Pavljo!
      'and when(ever) people passed the bridge, one heard at night: Paul, Paul!'

While it is undisputable that the examples quoted above render a categorial distinction, it is doubtful whether -*li* goes back to the -*l*/-*la*/-*lo*/-*le* of the Bulgarian participle or rather to the interrogative particle *li*; we have to take into account that, with the participle, a form ending in *li* does not exist in Bulgarian and Macedonian. On the other hand, the particle -*li* serves for marking conditional-temporal clauses, replacing normal *ako* or *da* (Gramatika III, 1994, 393ff):

(79) kanjat li te – ež, gonjat li te – bež!
      'if they invite you—eat; if they drive you away—run!'

(80) počneše li zimata, počvaxa se i veselbite
      'every time winter began, festivities began too'

We do not know how Bulgarian *li* could have been re-interpreted as a narrative marker, but what can be stated is that both *li* and narrative are not strictly indicative, i.e. they do not report on events that positively happened. Perhaps a contamination of the two elements has taken place.

2.11. A verb for 'have'

Nearly all Romani dialects have an expression for 'to have' which must be considered an old formation since similar constructions exist in Modern Indian languages as well. In Romani it is copula + oblique case of noun/pronoun for the possessor, e.g. *si man* lit. 'is *me*' = 'I have', and in Hindi various postpositions (*kā* 'of' for permanent possession or relationship, *ke pās* 'at, beside' for actual possession) cf. for Hindi

(81) zamīndār ke do gāmv the
   'the land owner owned two villages' (McGregor 1987, 52)

(82) us-ke pās paisā nahīm hai
   'he has no money' (McGregor 1987, 51)

Therefore, it is surprising that a new verb for 'have' came up in dialects influenced by Greek; it is *ther-el*, which originally meant 'hold, preserve' (< ai. *dharati*). In Kalderaš it means 'get, receive', and in various other dialects (Central and Northern) its passive form *therdjol* or *therdo si* was specialized for 'stand', a verb otherwise missing in Romani. The Balkan languages did not supply Romani with a direct model for shaping 'have', since in Greek or Bulgarian the semantic–etymological tie between 'hold' and 'have' is blurred and can no longer be detected. Therefore, the Roma had to find their own way, following the same, nearly universal, path that speakers of many other languages followed. This is to say that people take the easy way by direct copying, if there is such a possibility, but that they become inventive taking recourse to something like innate faculties if no other possibility is left.

What has been said here about universal paths of lexicalization might induce us to assume that no Greek or Balkanic model was needed in order to form a verb for 'have', but the model may have been the fact that, in the Balkanic languages, there existed a simple verb for this central element.

References

Barannikov, A.P. 1934. The Ukrainian and South Russian Gypsy dialects. Izdatel'stvo Akademii Nauk, Leningrad.

Boretzky, N. 1986. Regelentlehnung und Substrateinfluß in Kreolsprachen. In: Akten des 2. Essener Kolloquiums über Kreolsprachen und Sprachkontakte 1985, 9–39. Brockmeyer, Bochum.

Boretzky, N. 1988. Zum Interferenzverhalten des Romani. In: Zeitschrift für Phonetik, Sprach-wissenschaft und Kommunikationsforschung 42: 357–74.

Boretzky, N. 1992. Zum Erbwortschatz des Romani. In: Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 45: 227–51.

Boretzky, N. 1993a. Bugurdži. Deskriptiver und historischer Abriß eines Romani-Dialekts. Ost-europa-Institut an der Freien Universität Berlin. Balkanologische Veröffentlichungen, 21. Otto Harrassowitz, Wiesbaden.

Boretzky, N. 1993b. Conditional sentences in Romani. In: Sprachtypologie und Universalienforschung 46: 83–99.

Boretzky, N. 1993c. The concept of rule, rule borrowing, and substrate influence in Creole languages. In: Mufwene, S. (ed.): Africanisms in Afro-American language varieties, 74–92. University of Georgia Press, Athens and London.

Boretzky, N. 1994a. Romani. Grammatik des Kalderaš mit Texten und Glossar. Osteuropa-Institut an der Freien Universität Berlin. Balkanologische Veröffentlichungen, 24. Otto Harrassowitz, Wiesbaden.

Boretzky, N. 1994b. Zu den Ausdrücken für Notwendigkeit in den Romani-Dialekten. In: Romano Džaniben 3: 23–34.

Boretzky, N. 1996a. Sprachwandel im Romani: Systemdefinierende Eigenschaften und Interferenz. In: Dow, J.R. – Wolff, M. (eds): Languages and Lives. Essays in Honor of Werner Enninger, 137–53. Peter Lang, New York.

Boretzky, N. 1996b. Zu den Modalia in den Romani-Dialekten. In: Zeitschrift für Balkanologie 32: 1–27.

Boretzky, N. 1996c. The "new" infinitive in Romani. In: Journal of the Gypsy Lore Society, ser. 5, vol. 6, 1–51.

Boretzky, N. 1996d. Entlehnte Wortstellungssyntax im Romani. In: Boretzky, N. – Enninger, W. – Stolz, T. (eds): Areale, Kontakte, Dialekte. Sprache und ihre Dynamik in mehrsprachigen Situationen, 95–121. Beiträge zum 10. Bochum-Essener Symposion Juli 1995. Brockmeyer, Bochum.

Boretzky, N. – Igla, B. 1991. Morphologische Entlehnung in den Romani-Dialekten. Arbeitspapier Nr. 4 des Projekts "Prinzipien des Sprachwandels"; 50 p. Essen.

Boretzky, N. – Igla, B. 1993. Lautwandel und Natürlichkeit. Kontaktbedingter und endogener Wandel im Romani. Arbeitspapier Nr. 15 des Projekts "Prinzipien des Sprachwandels", Essen.

Boretzky, N. – Igla, B. 1994a. Wörterbuch Romani–Deutsch–English. Mit einer Grammatik der Dialektvarianten. Harrassowitz, Wiesbaden.

Boretzky, N. – Igla, B. 1994b. Interferenzlinguistik. Teil 1: Interferenz und Sprachwandel. In: Sprachdynamik. Auf dem Wege zu einer Typologie sprachlichen Wandels. Aus dem VW-Projekt "Prinzipien des Sprachwandels". Brockmeyer, Bochum.

Dobrovol'skij, V.N. 1908. Kiselevskie cygane. Tipografija Imperatorskoj Akademii Nauk, St. Petersburg.

Edel'man, D.I. 1983. The Dardic and Nuristani languages. Nauka Publishing House, Moscow.

Eloeva, F.A. – Rusakov, A.Ju. 1990. Problemy jazykovoj interferencii. Leningradskij Gosudarst-vennyj Universitet.

Finck, F. N. 1903. Lehrbuch des Dialekts der deutschen Zigeuner. Elwertsche Verlagsbuchhandlung, Marburg.

Gilliat-Smith, B. 1913–14. The dialect of the Drindaris. In: Journal of the Gypsy Lore Society 7: 260–98.

Gilliat-Smith, B. 1962. A new version of the song of the bridge. In: Journal of the Gypsy Lore Society 41: 124–33.

Gramatika na săvremenija bălgarski knižoven ezik 1994. t. III: Sintaksis. Bălgarskata Akademija na Naukite, Sofija.

Heinschink, M. – Zambakli-Heinschink F. 1994. Izmirští romové – Sepetčides. In: Romano Džaniben 2: 2–15.

Holzinger, D. 1993. Das Romanes. Grammatik und Diskursanalyse der Sprache der Sinte. Innsbrucker Beiträge zur Kulturwissenschaft. Sonderheft 85, Innsbruck.

Hopper, P. – Traugott, E. 1993. Grammaticalization. Cambridge University Press.

Hübschmannová, M. – Šebková, H. – Žigová, A. 1991. Romsko-český a česko-romský kapesní slovník. Státní pedagogické nakladatelství, Praha.

Igla, B. 1992. Entlehnungen und Lehnübersetzungen deutscher Präfixverben im Sinti. In: Erfurt, J. – Jeßing, B. – Perl, M. (Hrsg.): Prinzipien des Sprachwandels I. Vorbereitung. Beiträge zum Leipziger Symposion des Projekts "Prinzipien des Sprachwandels" 1991 an der Universität Leipzig, 38–56. Brockmeyer, Bochum.

Jensen, H. 1931. Neupersische Grammatik. Winter, Heidelberg.

Ješina, J. 1886. Románi čib oder die Zigeunersprache. List und Franke, Leipzig.

Kenrick, D.S. 1967. The Romani dialect of a musician from Razgrad. In: Balkansko Ezikoznanie 11/12: 71–8.

Kiparsky, V. 1967. Russische historische Grammatik, Teil II. Winter, Heidelberg.

Knobloch, J. 1953. Romani-Texte aus dem Burgenland (Burgenländische Forschungen, hrsg. vom Landesarchiv und Landesmuseum. Heft 24.) Eisenstadt.

Koneski, B. 1967. Gramatika na Makedonskiot literaturen jazik. II. Prosvetno Delo, Skopje.

Kopernicki, I. 1930. Textes tsiganes. Polska Akademia Umiejętności. (Mémoires de la Commission Orientaliste 7). Kraków.

Kostov, K. 1962. Aus der Syntax der Zigeunersprache Bulgariens. In: Balkansko Ezikoznanie 4: 131–46.

Kostov, K. 1973. Zur Bedeutung des Zigeunerischen für die Erforschung grammatischer Inter-ferenzerscheinungen. In: Balkansko Ezikoznanie 16: 99–113.

Lípa, J. 1963. Příručka Cikánštiny. Státní Pedagogické Nakladatelství, Praha.

Lípa, J. 1965. Cikánština v jazykovém prostředí slovenském a českém. Rozpravy Československé Akademie Věd 75, 11.

Macalister, S. 1914. The language of the Nawar or Zutt – the Nomad Smiths of Palestine. Journal of the Gypsy Lore Society Monographs No. 3.

Mackridge, P. 1987. The Modern Greek language. Clarendon Press, Oxford.

Mann, S.E. 1947. Two Moravian Romani folk tales. In: Journal of the Gypsy Lore Society 26: 24–36.

Mayerthaler, W. 1981. Morphologische Natürlichkeit. Wiesbaden, Athenaion.

McGregor, R.S. 1987. Outline of Hindi grammar. Clarendon Press, Oxford.

Miklosich, F. von 1881. Über die Mundarten und die Wanderungen der Zigeuner Europas 12. Syntax der Zigeunermundarten. (Denkschriften der phil.-hist. Classe der Kaiserl. Akademie der Wiss. Bd. 31.) Wien.

Paspati, G. A. 1870. Études sur les Tchingianés ou Bohémiens de l'Empire Ottoman. Imprimerie Koroméla, Constantinople.

Pott, A. F. 1844–45. Die Zigeuner in Europa und Asien. I & II. Heynemann, Halle (repr. Leipzig 1964).

Romano Džaniben 1994. Časopis romistických studií (ed. by M. Hübschmannová). 1–4. Praha.

Romano Hangoro 1993. (comp. by D. Banga). Ministerstvo Školstva a Vedy Slov. Rep.

Romane svjati gilja 1933. Romane Lila No. 2. Izdanie na Săjuza na Bălg. Evang. Baptistski Cărkvi. Sofija.

Rozwadowski, J. 1936. Wörterbuch des Dialekts von Zakopane. (Mémoires de la Commission Orientaliste 21). Kraków.

Sampson, J. 1926/1968. The dialect of the Gypsies of Wales. Clarendon Press, Oxford.

Soravia, G. 1978. L'infinito in Romanes. In: Istituto Lombardo. Academia di Scenze e Lettere. Rendiconti, Classe di Lettere e Scenze morali e storiche 112: 35–54. Milano.

Štrukelj, P. 1980. Romi na Slovenskem. Cankarjeva Založba, Ljubljana.

Thesleff, A. 1901. Wörterbuch des Dialekts der Finnländischen Zigeuner. Acta Societatis Scientiarum Fennicae 29, Helsingfors.

Uhlik, R. 1974. Kategorija imperativa u romskom jeziku. In: Godišnjak 12, 75–125. Akademija Nauka i Umjetnosti Bosne i Hercegovine. Sarajevo.

Valtonen, P. 1972. Suomen mustaliskielen etymologinen sanakirja. Suomalaisen Kirjallisuuden Seura, Helsinki.

Valtonen, P. – Gilliat-Smith, B. 1967. Finnish Gypsy texts. In: Journal of the Gypsy Lore Society 46: 82–97.

Vekerdi, J. 1980. Numerical data on loan words in Gypsy. In: Acta Linguistica Hungarica 30: 367–73.

Vekerdi, J. 1984. The Vend Gypsy dialect in Hungary. In: Acta Linguistica Hungarica 34: 65–86.

Wentzel, T. 1983. Die Zigeunersprache (Nordrussischer Dialekt). Enzyklopädie, Leipzig.

Windfuhr, G.L. 1970. European Gypsies in Iran. In: Anthropological Linguistics 12: 271–92.

Address of the author: Norbert Boretzky
                          Ruhr-Universität Bochum
                          Institut für Sprachwissenschaft
                          Postfach 10 21 48
                          D–44780 Bochum
                          Germany

DISTURBANCES AND INNOVATIONS IN THE CASE SYSTEM IN BULGARIAN ROMANI DIALECTS

BIRGIT IGLA

Abstract

Against the backround of the expressions for case in European Romani dialects in general, the article reviews recent developments in Bulgarian Romani dialects. Romani displays both inflectional forms and analytic means for the expression of case. Although there is a tendency of replacing inflectional forms by analytic ones, the inflectional forms have remained unchanged. On the other side we establish a lot of innovations within the system of analytic case expression. The same preposition replacing the "locative" in both its functions (place and direction) is used for expressing the dative, and even the genitive case. This levelling of the analytic case expressions seems to be triggered by the contact language. Finally, some irregularities in the government of prepositions are considered that provide further evidence with respect to ongoing changes in the case system or Bulgarian Romani dialects.

1. The case system of Romani

As a result of typological restructuring, the Romani dialects can express case relations by either inflectional means (case suffixes)[1] or by prepositional phrases. The inflectional pattern must be considered the older one, i.e. as indigenous or inherited from Indic. Some of the case affixes correspond to the postpositions of Modern Indic languages, but a good etymology has not been found in all cases. The analytic paraphrases are partly formed with prepositions that correspond to the case suffixes.[2] Therefore, it is quite plausible that case suffixes and prepositions have a common origin. Since it is hardly probable that affixes that have already merged with the stem could have been separated later and used as prepositions, one has to assume elements that originally had no fixed position, i.e. that could either precede the noun or follow it. In any case, the prepositions must have been created before the merger of the postposed elements with the noun.

---

[1] Some authors—e.g. Cortiade (1990)—consider the case affixes as postpositions. We will not deal here with the morphological status of the elements in question.

[2] The government behaviour of prepositions is presented in **3.1.**

The case system (both the inflectional and the prepositional one) presents some peculiarities that invite a comparison with the Balkan languages. I will discuss features common to all dialects of Romani spoken in Europe as well as some peculiarities of Romani dialects spoken nowadays in Bulgaria.

1.1. The common case expressions

Let us first consider the case expressions as present in many conservative dialects.[3] The singular paradigm of *o raklo* 'the boy' is as follows:

Table 1

	inflectional	prepositional
nom	o rakl-o	
acc	e rakl-es	
gen	e rakl-es-ko/ki	–
dat	e rakl-es-ke	ka o/k-o raklo
loc	e rakl-es-te	te o/t-o/k-o raklo
abl	e rakl-es-tar	tar (atar/katar) o raklo
instr	e rakl-e-sa(r)	? sa(r) o raklo

General remarks: the accusative relation is not expressed in any of the dialects by a preposition. It is broadly accepted that the so-called genitive is not a pure case.

---

[3] For Bulgaria we can regard the dialect of the Erlides in Sofia as an example of a well-preserved dialect. A grammatical description of this dialect is presented in Kostov (1963), a small dictionary is given by Malikov (1992). The translation of the New Testament by Metkov (1995) gives a sample of that variety, too. If not otherwise indicated, the data presented here stem from a broad field research on Romani dialects spoken in Bulgaria that has been carried out by me since 1993, and since 1995 in cooperation with students and colleagues of mine from the University of Sofia. I am especially indebted to Evelina Grigorova, Elisaveta Manolova, Ivelina Čobanova, Karamfilka Getova, Nadežda Buχova and Kaspar Krikorjan. For general linguistic support I am indebted to Vladimir Filipov.

Abbreviations used in the tables are as follows:

abl	ablative	instr	instrumental
acc, A	accusative	io	indirect object
AGR	Ancient Greek	loc, L	locative
art	article	MB	Modern Bulgarian
Bg	Bulgarian	MGR	Modern Greek
c.g.	casus generalis	nom	nominative
d	direction	OCS	Old Church Slavonic (Old Bulgarian)
dat, D	dative	p	place
gen, G	genitive	poss	possessive pronoun/possessive affix
Gr	Greek	pp	prepositional phrase
infl	inflectional	prep	prepositional

The reason is its special syntactic behaviour which is the same as for adjectives. The inflectional genitive is in general not replaced by a prepositional phrase—for such a replacement we find, however, evidence in Bulgarian varieties of Romani (cf. **2.4**). The dative can be replaced by a prepositional phrase with *ka* (*k-o/k-i*). Generally, there are restrictions for the replacement of cases by pp's following (i) the animate/ inanimate distinction (with animate nouns prepositional phrases are less favoured) and (ii) the word class (replacement with substantives rather than with pronouns). Determination seems to be of no or at most minor importance with respect to the case realisations selected. In this field, however, especially the function of the zero article demands further research. The prototypical function of the dative case is to express the indirect object which, more often than not, is animate. This seems to be the reason why the dative is less often expressed by a prepositional phrase. The preposition used for the replacement of the dative (*ka/k-*) has extended to the area of other cases—this development is to be discussed below (**2.3; 2.4**). The locative is frequently replaced by a prepositional phrase: the dialect described by Paspati (1870) has *te*, in some Bulgarian dialects we find the shortened form *t-o/t-i*.[4] Most dialects I am acquainted with, however, have *ka/k-* which by its origin is a dative preposition. The extension of the function of this marker, which goes still further in Bulgarian dialects, seems to date back to the pre-divergent phase of Romani, i.e. the time when Romani presumably was a more or less homogeneous language that had not split into dialects following their own tendencies of development. The ablative is, together with the locative case, the case that is most often replaced by a prepositional phrase. There are some indications that point to a replacement of the instrumental case by an indigenous preposition in Anglo-Romani and in Calo, but they are few and not very convincing.[5]

1.2. Case expressions realised by borrowed prepositions

Some dialects express case relations with a prepositional phrase that contains a borrowed preposition. The instrumental case may be expressed by *ku* (< Rum. *cu*) in Kalderash dialects, or by the German *mit/met* in Sinti varieties.[6] In many instances, especially in combination with pronouns, the preposition 'with' is used alongside the instrumental ending, i.e. the result is a double marking of the case relation ((1), (2b)).

---

[4] In these varieties *t-o/t-i* often alternates with *k-o/k-i*. We have met with forms in *t-* in Plovdiv, Pazardžik, occasionally in the Rhodopes and even in Sofia.

[5] Some examples of 'the commitative postposition in prepositive function' can be found in Cortiade (1990, 25).

[6] The following examples are taken (1) from Boretzky–Igla (1994), (2a) and (3) from Soravia (1981, 4), (2b) from Holzinger (1993, 86) and (2c) from Igla (1992, 42).

(1)   ku   sa   oge-sa
         with whole heart-instr
         'with whole heart'

(2) (a) e   bale-ha/   mit   u   balo
         art pig-instr/ with art pig:nom
         'with the pig'
     (b) leha/   mit   leha
         3.sg.masc:instr with 3.sg.masc:instr
         'with him'
     (c) met   o   kaš
         with art wood:nom
         'with the wood'

While in Sinti dialects the replacement occurs fairly often (though the inflectional forms also exist), in Kalderash *ku* is only rarely used. Many varieties of Sinti also replace the genitive case with the borrowed preposition *fon* (German *von*).

(3)   e   bales-kro/   fon u   balo
         art pig-gen/   of   art pig
         'the pig's (of the pig)'

2. The locative

2.1. The functions of the locative

The locative case—both the inflectional and the prepositional one—expresses both place/locality (*ubi?*) (4a) and direction/goal (*quo?*) (4b).

(4) (a) sinjom e   gaves-te/   and o   gav
         I was   art village-loc/ in   art village:nom
         'I was in the village'
     (b) džav   e   gaves-te/   and o   gav
         I go   art village-loc/ in   art village:nom
         'I am going to the village'

The merger of these functions in one expression is generally accepted as a characteristic of the Balkan languages. It is a special case of the merger between the dative and the accusative case: in Ancient Greek we have *en* + dat for place (p) and *eis* +

acc for direction (d). In Modern Greek 'place' is expressed by *eis* + acc which thus serves both for place and direction. In Old Bulgarian two prepositions, *na* and *vă*, express place when governing the locative and direction when governing the accusative case. Both prepositions are used in Modern Bulgarian, too: after the restructuring of the case system, i.e. the replacement of inflectional cases by prepositions + casus generalis (c.g.), both *na* and *v(ăv)* (+ c.g.)[7] are used for place and direction (for the distinction between *na* and *v(ăv)* see **3.3** below).

The fusional processes in the two languages, though not identical, have led to very similar results. Schematically [p for place, d for direction] they are as follows:

<div align="center">

Table 2

</div>

	AGR	OB
p:	en + D	na/vă + L
d:	eis + A	na/vă + A

	MGR	MB
p + d:	eis + A	na/v(ăv) + c.g.

2.2. Differences in the inflectional and the prepositonal expressions

In Romani the picture is more complex, because we have to consider both the inflectional and the prepositional expressions.

<div align="center">

Table 3

</div>

	ROMANI	
	infl (old and new)	*prep (old)*
p + d:	-este	te + N
		prep (new)
		ka/k- + N

Thus, in Romani the inflectional system has remained unchanged while there has been restructuring in marking the prepositional case: the preposition *ka*, originally used for paraphrasing the dative case (with its primary function of expressing the

---

[7] In Modern Bulgarian inflected forms exist only in the pronominal system—here we may disregard fossilized forms and the so-called vocative case.

indirect object) has replaced the locative preposition *te*. In order to understand this innovation we have to compare the expressions for the indirect object in Romani and its contact languages.

2.3. Syncretism of dative and locative

After the restructuring of the case systems in Greek as well as in Bulgarian the same expressions serve for 'place', 'direction' and 'indirect object', e.g. Gr. *pao/imun/ipa ston jatro*, Bg. *otivam/bjaχ/kazaχ na lekarja* 'I go to/ was at/ said to the doctor'.[8]

While Greek and Bulgarian use the same forms, Romani, at least during an earlier stage, has different prepositional phrases for the indirect object and place/direction. By identifying the expression of the indirect object with those for place and direction in Romani the prepositional system of Romani has been brought closer to one of the two contact languages, cf.:

Table 4

	ROMANI		GR	BG
	old	*new*		
p	t-i daj	k-i daj	ston jatro	na lekarja
d	t-i daj	k-i daj	ston jatro	na lekarja
io	k-i daj	k-i daj	ston jatro	na lekarja

Marking place and direction in the same way seems always to have been a constituent feature of the inflectional locative of Romani—we do not know of any dialect where this is different. The levelling of the prepositional paradigm, however, is a later development as can be seen from the functional distribution between *ka/k-* and *te/t-* which still occurs in some dialects. The preposition levelling corresponds to the Greek (merger of *eis* and *en*), as well as to the Bulgarian development (merger of locative and accusative).

2.4. Syncretism between dative/locative and genitive

We want to draw the reader's attention to some recent, not widespread phenomena in Bulgarian Romani dialects. Along with the dative and the locative case, the gen-

---

[8] A purely synchronic comparison between Greek and Romani reveals an interesting parallel between the two languages that will not be considered in detail here: both languages use inflectional and prepositional expressions for marking case relations. However, in Greek the two expressions are interchangeable only for the indirect object, cf. Rom. *phenav e dajake/ k-i daj* and Gr. *leo tis manas/ stin mana* 'I tell the mother'.

itive, too, can be expressed by a prepositional phrase with the preposition *k-*.[9] In alternation with the inflectional form as in (5a), the respective varieties use prepositional expressions as in (5b):

(5) (a) o   čhav-esko dad,   i   čhaki   daj
         art boy-gen   father, art girl-gen mother
         'the boy's father, the girl's mother'
     (b) o   dad   k-o   čhavo, i   daj   k-i   čhaj
         art father at-art boy:nom art mother at:nom girl-art
         'the father of the boy, the mother of the girl'

Thus *k-* has acquired most of the functions which in Bulgarian are realized by *na*.

Table 5

	infl	*prep*		BG
		old	*new*	
gen.	-esko	—	k-o/-i	na
dat.	-eske	k-o/-i	k-o/-i	na
loc.	-este	t-o/-i (k-o)	k-o/-i	na

By this development the degree of convergence towards the contact language has increased: the varieties of Romani that have undergone this change express the genitive, the dative and the locative (place and direction) relations by the same prepositional phrase, just as Bulgarian does. The genitive now has entered the analytical paradigm. This type has been found only in such regions where Bulgarian is the dominant contact language (and maybe the only one). In areas with dense Turkish populations, on the other hand, the Roma use a possessive construction (6) which follows the Turkish model (7). Comparing the possessive construction as used in the local Turkish varieties (7a)[10] with the Standard Turkish one (7b) it becomes obvious that the Balkan varieties of Turkish have served as a model.

---

[9] The examples for the substitution of the inflectional genitive come from the Northern parts of Bulgaria and mainly from the towns of Lukovit and Montana.

[10] The rearranging of the possessive construction appears in many (most?, all?) Turkish dialects spoken in the Balkans. It appears that the change in the Turkish word order has been triggered/evoked by the influence of the Balkan contact languages. (cf. Németh 1965). Examples (7a) and (7b) from Németh (1965, 114).

(6)   lesko   dad   e   čheskoro
         his   father   art boy:gen
         'the boy's father'
(7) (a) baba-si   qiz-in
         father-poss   girl:gen
         'the girl's father'
     (b) kızın   baba-sı
         girl:gen   father-poss
         'the girl's father'

Disturbances in the genitive expression are found in other areas as well. A dialect spoken in Prilep (Macedonia)[11] uses the preposition *katar* which is originally ablative 'from' in meaning, cf.:

(8) (a) katar o   tikno   čhavo e   sasui
         from   art young   man   art mother-in-law
         'the young man's mother-in-law'
     (b) k-o   krajo   katar o   gav
         at-the end   from   art village
         'in the outskirts of the village'

It should be noticed that the local Macedonian dialects denote the possessive relation by *ot* 'from', and not by *na* as the standard language. This can be taken as further evidence to the claim that the innovations that can be observed in Romani are to be interpreted as convergence towards the contact language.

3. Government of prepositions

3.1. General remarks

The government behaviour of prepositions in Romani depends on whether a noun or a pronoun follows the preposition: nouns are in the nominative (9a), pronouns in the locative case (9b):

(9) (a) anglal   o   kher
         in front of art house:nom
         'in front of the house'

---

[11] The examples for Prilep are taken from Boretzky (1992).

(b) angla   les-te
    in front of   3.sg.masc-loc
    'in front of him'

This invites a comparison with Bulgarian rather than with Greek: in Greek prepositions govern the accusative case for both word classes. In contemporary Bulgarian nouns are in casus generalis (which is based on the nominative case). Pronouns have kept two forms (accusative and dative) for the oblique case/ two oblique forms: after prepositions they are always in the accusative case.

3.2. Functional contents of inflectional vs. prepositional marking

One might consider the parallel existence of inflectional and prepositional means of expression as being symmetrical if there is a one-to-one relationship of prepositional and inflectional means, both on the expressional and on the functional level. If we consider Table 1 again, we see that there is no absolute correspondence, since for some cases no prepositional marker exists. After the merger of the expressions for indirect object and place/direction (on the prepositional scale) a prepositional marker for the genitive, too, has come into existence. The inflectional paradigm with its greater specialization fulfils the one form – one function condition. In the prepositional counterpart, however, with the merger of forms one preposition has acquired several, quite distinct, functions. Thus, there are arguments for interpreting the development in terms of either loss or increase in symmetry, depending on whether the expressional or the functional level is considered. To be sure, as far as the prepositional paradigm is concerned, a higher degree of equivalence with the contact language was achieved.

For Romani dialects in general, different relations between inflectional and prepositional cases concerning the functional content (specialization vs. broadness) can be observed: the examples given above (merger of *ka* and *te*; merger of the genitive with the dative/locative) suggest a lesser degree of differentiation in the prepositional marking than in the inflectional one. There are, however, arguments for the opposite interpretation. The emergence of differentiation in the prepositional system can be observed, too. A dialect spoken e.g. in Athens[12] uses the compound prepositions *katar, atar, andar* as markers of the point of departure, cf.

(10) (a) andar o kher 'out of the house'
     (b) katar i dej 'from the mother'
     (c) tar i Germania 'from Germany'

---

[12] Cf. Igla (1996).

Within the inflectional system no such differentiation is possible: in all three cases the ablative case would be used (*e kherestar, e dejatar, e Germaniatar*). In order to express the elative component in (10a), one has to add an adverbial:

(11)   e   kheres-tar andral
        art house-abl   from inside
        'out of the house'

3.3. Differentiation of prepositions in Bulgarian Romani

We observe a development towards differentiation in the use of prepositions in Bulgarian Romani dialects, too. The above mentioned Bulgarian prepositions *na* and *v(ăv)* when expressing the direction of a movement are distributed in the following way: *na* + noun designates the goal of a movement—the locality as such is of no importance—in these cases the noun is usually without a definite article. To emphasize the locality one uses, instead of *na,* the preposition *v(ăv)* and the noun with a definite article, cf.:

(12) (a) otivam na kino
         'I am going to the cinema (in order to see a film)'
     (b) otivam v kinoto
         'I am going into the cinema (i.e. the building)'

For several Romani dialects (Sofia, Plovdiv, and elsewhere) I observed a corresponding distribution between *k-* and *and,* cf.:

(13) (a) džav k-o kinos
         'I am going to the cinema (to see a film)'
     (b) džav and o kinos
         'I am going into the cinema (the building)'

Generally, we can note the fact that the convergence towards the contact language in the case system concerns mainly the prepositional expressions. The inflectional system is involved in view of the fact that its function is, to a certain degree, taken over by prepositional phrases. Since there is almost no trace of replacement of oblique case expressions with pronouns, the dichotomy of inflectional and prepositional systems keeps on existing.

Except for the loss in frequency there are other indications for the weakening of the inflectional case system. These are considered in the following section.

3.4. The preposition *bi*

bi is the only preposition in Romani that governs the genitive case. With pronouns it demands the possessive form,[13] cf.:

(14) (a) bi   e   dades-ko
         without art father-gen
         'without the father'
     (b) bi   moro
         without 1.sg:poss
         'without me'

Probably due to the destabilization of the genitive (and the case system as a whole) or maybe due to the fact that the government behaviour of *bi* is exceptional to the whole system, a number of deviations are observed.

3.4.1. 'Wrong' government behaviour

Instead of the genitive case (the possessive pronoun), we have found many instances with *bi* being used with the ablative (15a) or the instrumental case (15b), cf.:

(15) (a) bi   e   dades-tar
         without art father-abl
         'without the father' (lit. 'without from the father')
     (b) bi   tu-sa
         without 2.sg-instr
         'without you' (lit. 'without with you')

Such examples can be found in numerous dialects either alongside with the 'correct' construction or as the only forms the preposition governs in the respective variety. We have not discovered any instances of *bi* governing the accusative case.[14]

3.4.2. Borrowing the Bulgarian preposition

The Bulgarian preposition *bez* has been borrowed into different varieties of Romani. It shows no uniform government behaviour, but one can observe that the accusative is favoured. With animate nouns the form in *-es* (respectively *-(j)a* for feminines,

---

[13] Varieties of Kalderash have developed a special genitive of pronouns which is distinct from the possessive pronoun, cf. *munro dad* 'my father' and *bi mango* 'without me'. For more detail see Boretzky (1994, 52).

[14] That could be expected if the innovation was triggered off by Bulgarian influence.

and -*en* for plural) can be used (16b), but more frequently we find the unmarked form (i.e. the nominative) (16c). Inanimate nouns are always in the nominative which for them is identical with the form of the direct object (i.e. the accusative) (16d).

(16) (a) bez tu
           'without you' (tu = nom. and acc.)
      (b) bez   e   dad-es
           without art   father-acc
           'without the father'
      (c) bez   o   rom
           without art   man:nom
           'without the man'
      (d) bez   o   lil-a
           without art   paper-nom:pl
           'without the papers (documents)'

Occasionally *bez* combines with the ablative (17a) or the instrumental (17b) case:

(17) (a) bez   tu-tar
           without 2.sg-abl
           'without you'
      (b) bez   pe   rome-sa
           without poss husband-instr
           'without her husband' (lit. 'without with her husband')

One might conclude that in the first instance the former government has been desta-bilized (see **3.4.1**). Only when this had happened did the borrowing of the Bulgarian preposition—together with the government behaviour of the preposition in Bulgarian—take place. The examples given in (17) are isolated ones that cannot be explained by Bulgarian, but rather look like a substrate transfer from the phase when the preposition *bi* had been destabilized.

4. Conclusion

The restructuring of the case system common to all European dialects certainly took place at a time before the divergent development of Romani dialects set in.

Most, but not all, innovations found in Romani dialects spoken in Bulgaria today are due to the influence of the actual contact language. In the other cases one

has to check very carefully whether one is confronted with innovations within an otherwise intact dialect or with destabilizations due to ongoing language change (loss in linguistic competence).

The linguistic area of present-day Bulgaria lends itself to comparing different dialects of Romani in view of the stability of the synthetic vs. the analytical case system. Whereas the levelling of the prepositional system as well as the predominance of prepositional case expressions is found in dialects for which Bulgarian is the only or at least the most important contact language, in dialects under the influence of Turkish the inflectional system survives astonishingly well.

The analytical case expressions, or better, the dichotomy of inflectional and prepositional case expressions must be considered as one of the system-defining structural properties of Romani as a whole. Since the preservation vs. loss of the inflectional case system as a rule reflects the situation of the contact language, we may conclude that these developments follow tendencies determined by language contact rather than language-inherent properties.

References

Asenova, P. 1989. Balkansko ezikoznanie. Osnovni problemi na balkanskija săjuz [Balkan linguistics. Basic problems of the Balkan Union]. Izdatelstvo nauka i izkustvo, Sofia.

Boretzky, N. 1992. Der Dialekt von Prilep (unpublished manuscript, 34 pages).

Boretzky, N. 1994. Romani. Grammatik des Kalderaš-Dialekts mit Texten und Glossar. (= Osteuropa-Institut der Freien Universität Berlin. Balkanologische Veröffentlichungen, Band 24.) Harrassowitz, Wiesbaden.

Boretzky, N. – Igla, B. 1994. Wörterbuch Romani–Deutsch–Englisch. Mit einer Grammatik der Dialektvarianten. Harrassowitz, Wiesbaden.

Cortiade, M. 1990. Let us finish the matter for good and all with the so-called seven cases of the Romani nominal morphology. Paper presented at the 12th Annual Meeting of the JLS.

Holzinger, D. 1993. Das Romanes. Grammatik und Diskursanalyse der Sprache der Sinte. (= Meid, W. (ed.): Innsbrucker Beiträge zur Kulturwissenschaft, Sonderheft 85). Innsbruck.

Igla, B. 1992. Entlehnungen und Lehnübersetzungen deutscher Präfixverben im Sinti. In: Erfurt, J. – Jeßing, B. – Perl, M. (eds): Prinzipien des Sprachwandels. I. Vorbereitung, 38–56. Brockmeyer, Bochum.

Igla, B. 1996. Das Romani von Ajia Varvara. Deskriptive und historisch-vergleichende Darstellung eines Zigeunerdialekts. (= Osteuropa-Institut der Freien Universität Berlin. Balkanologische Veröffentlichungen, Band 29.) Harrassowitz, Wiesbaden.

Kostov, K. 1963. Grammatik der Zigeunersprache Bulgariens (Phonetik und Morphologie) [unpublished dissertation].

Malikov, J. 1992. Cigansko-Bălgarski rečnik [Gypsy–Bulgarian dictionary]. Fondacija Otvoreno obštestvo, Sofia.

Metkov, S. 1995. O Nevo Zaveti amare Devlesko o Isus Xristos. Sofia.

Mirčev, K. 1978. Istoričeska gramatika na bălgarskija ezik [Historical grammar of the Bulgarian language]. Izdatelstvo nauka i izkustvo, Sofia.

Németh, Gy. 1965. Die Türken von Vidin. Sprache, Folklore, Religion. Budapest.

Paspati, G.A. 1973. Études sur les Tchinghianés ou Bohémiens de l'Empire Ottomane. Biblio-Verlag, Osnabrück [Reimpression de l'édition de 1870].

Soravia, G. 1981. Vocabolario Sinto delle Venezie. Lacio drom, Roma.

Address of the author: Birgit Igla
                           St. Kliment Ohridski University
                           Germanistik
                           Sofia 1000
                           Bulgaria

ETYMOLOGICA ZINGARICA

ENDRE TÁLOS

Abstract

This paper begins with an introduction of the conventions I use below and my chief sources. This is followed by a sketch of the relationship of the dialects of Romani and a brief reconstruction of the phonological systems of Ancient Romani and Ḍommānī, the Śaurasēnī-like Prakrit, which became the predecessor of Romani. The larger second part contains new or amended etymologies for 87 reconstructed Ancient or Wallachian Romani words. These may be considered part of an evolving etymological dictionary of Romani.

My original intention years ago was to comment on Vekerdi's dialect dictionary, to prepare a list of corrigenda. I had not only found numerous misspelt forms but also erroneous and bad etymologies. Some of them are so absurd that I do not even bother to refute them in the etymologies that follow. To mention but a few: Vekerdi does not take the Wallachian Romani suffix *-kinja* ('-woman, female person') to be the adoption of Southern Slavonic or Bulgarian *-киня* (e.g. *ратайкиня* 'servant girl'), but designates Greek γυνή as its source — with a "?" to be sure (Ve 89) — although this has been pronounced [*jini*] for more than a thousand years. The German Romani (GR) *kuráko* 'crow' he derives from German "*Krähe* ?" instead of the obvious etymon, Greek κοράξ. Culling such examples would have been too extensive a job and, as a matter of fact, uninteresting. Instead, I have turned my attention to Romani words which were as yet of unknown origin or whose classical etymologies (in, for example, Turner) have proved unsatisfactory. During this work there emerged a collection which may command interest and form the base stock of entries in a future etymological dictionary. Its structure will, of course, be different from what is found below. According to Vekerdi "not counting the words adopted from the language of the majority people, the full vocabulary of a speaker of Romani totals about 1200. ... Slightly more than half of the stems (ca. 400–430) are Indic in origin" (1974, 14f). In another place (1981, 410), he talks about 400–500 Indic words. As always, he calculates downwards, but even if the figure given is right, it can now be amended to almost 600, since I claim to have discovered about more than fifty words that they

are also Indic in their origin. Turner's etymological dictionary suggests that many of these are lexical archaisms retained only in Romani, while other Modern Indic languages have replaced this Old Indic vocabulary. This modifies the current picture of Romani.

The headword of most entries is the reconstructed Ancient Romani form, in some cases it is a word from a present-day dialect. My four chief sources are Vekerdi 1983 (Ve), Uhlik 1983 (Uhl), Hübschmannová *et al.* 1991 (HŠŽ) and Kochergina's Sanskrit–Russian dictionary (1978), abbreviated SRSl and the supplement to its second edition, abbreviated SRSl D, on pages 896–943. It was only once that I had to consult a bigger dictionary.

In the Wallachian Romani data from Hungary I indicate word stress on polysyllabic words (with a grave accent) and length (with a macron), unlike my sources. For the sake of uniformity I have changed the consonant symbols of Uhl and HŠŽ, according to the following chart:

HŠŽ	Uhl		here	Ve if different
t'	*ć*	→	*tj*	
d'	*đ*	→	*dj*	
ň	*nj*	→	*nj*	
ľ	*lj*	→	*lj*	
dž	*dž*	→	*ǯ*	*dž*
—	*ṛ*	→	*r̄*	*rr, r*
—	*i̯*	→	*ï*	
č	*č*	→	*č*, WR *ć*	
—	*šj*	→	*ś*	*š*
—	*žj*	→	*ź*	*ž*

Uhlik's language I simply call Bosnian Romani (BR), although his dictionary cites data from several dialects, he does not indicate which. Those which are probably or certainly Wallachian Romani I mark with parenthesized (*=WR, =WR). It must be noted that the dialect forming the core of his dictionary — whose data were collected half a century ago — is mostly spoken in Vojvodina/Vajdaság today, whereas in Bosnia Albanian Romani is gaining ground and the mixture of the two dialects gives rise to new ones.

Each Hungarian Romani (HuR) and Slovenian Romani (SnR) form is cited from Vekerdi's dictionary. Vekerdi's notion of the Gurvāri (Gu) dialect is misleading: besides real Gurvāri data he labels others that may be linked to the Cerhāri, Čurāri, Kherāri, etc. tribes as Gu as well, but this merger is unjustified despite the fact that

all of these belong to the group of central dialects genetically. The latter I define as North Eastern Wallachian Romani (WRNE) dialects adding that their Wallachian Romani character is only secondary, they do not belong there genetically. The genetical relationships of European Romani dialects are shown in the tree below:

```
                                    AR
                    _____------┴------_____
               _____/     \                      E
          ____/            \             ____/        _____
         /                  \           /                         \
        W                    C         WR              non-WR dialect
   GR and all the          SkR         and             of the Balkans
   other dialects          HuR         BR
   of W Europe             SnR
                           Gu
                          WRNE
                    Ukrainian Romani
```
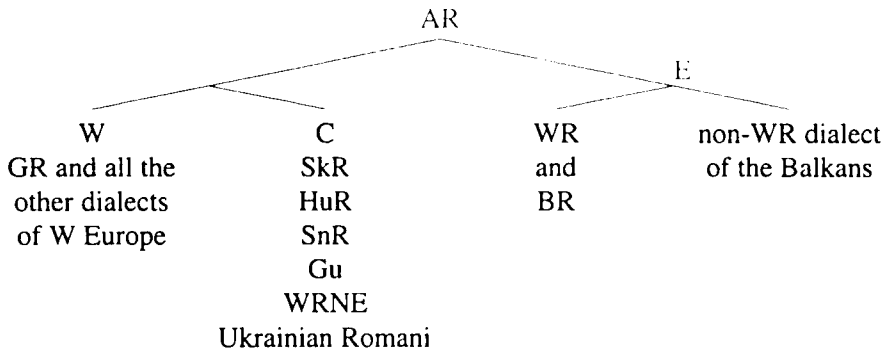
Cf. Ventcel'–Čerenkov 1976

My reconstruction of the consonant system of Ancient Romani is the following:

*ph	*th			*čh	*kh
*p	*t	*c	[ṭ]	*č	*k
*b	*d	*ʒ	*ḍ	*ǯ	*g
*f		*s	[ṣ]	*š	*x
		*z		*ž	
*m	*n		[ṇ]		
*v	*l	*r	[ṛ]	*i	*h

The *čh* column was palatal, but the *št* cluster was perhaps still realized as [ṣṭ], *nd* as [ṇḍ], and *ḍ* could have had a variant [ṛ] as well. The Śaurasēnī-like Prakrit, which became the predecessor of Romani, I have dubbed Ḍommānī. It is an important model explaining the phonological development of Romani (Tálos 1980) modelling the period from the separation of Romani until its appearance in Armenia. By Romani I only mean European Gypsy, I do not consider Bōšā (Armenian Gypsy) and Zuṭṭ (Syrian Gypsy) to be close relatives of Romani. The reconstructed consonants of Ḍommānī are given in the following charts:

Word-initial Cs

	*s	*ṣ	*ś	*x
*ph	*th	*ṭh	*ch	*kh
*p	*t	*ṭ	*c	*k
*bh	*dh	*ḍh	*jh	*gh
*b	*d	*ḍ	*j	*g
*mh	*nh	—		
*m	*n	—	*ñ	
—	—	—	—	*h
—	*l	—	—	
		*rh		
		*r		

Word-initial Cr clusters

—	—	—	—	
—	—	—	—	—
*pr	*tr	—	—	—
*bhr	—	—	—	—
—	*dr	—	—	—
—	—	—	—	
—	—	—	—	
—	—	—	—	
		—		
		—		

Intervocalic Cs

	*s	*ṣ	*ś	*x
—	—	—	—	—
—	—	—	—	—
—	*dh	*ḍh	—	—
—	*d	*ḍ	—	—
—	—	—		
*m	*n	*ṇ	—	
*vh	—	*ḷh	*yh	*h
*v	*l	*ḷ	*y	
		*rh		
		*r		

Intervocalic geminates

		*ss	*ṣṣ	*śś	*xx
*pph	*tth	*ṭṭh	*cch	*kkh
*pp	*tt	*ṭṭ	*cc	*kk
*bbh	*ddh	*ḍḍh	*jjh	*ggh
*bb	*dd	*ḍḍ	*jj	*gg
*mmh	*nnh	*ṇṇh		
*mm	*nn	*ṇṇ	*ññ	
—	—	—	—	—
—	*ll	—	—	
		—		
		—		

Other word-internal C_xC_y clusters could only be of the *NC* or *ST* type:

*mbh	*ndh	*ṇḍh	*ñjh	*ṅgh		*sth	*ṣṭh	
*mb	*nd	*ṇḍ	*ñj	*ṅg		*st	*ṣṭ	*śc

Of three member clusters there are only a few examples of the *NCr* and *STr* type. In transcribing Ḍommānī reconstructions I follow the Indologist tradition, but with Ancient Romani I switch the code and, for example, indicate (medial) palatals by a háček.

The [ə] reflex of *a* in open syllables was originally only an allophone, similarly to the short variants of *ē* and *ō* in closed syllables. Temporarily there could have existed **ai* and **au* as well as a result of contraction. Stress probably fell on the penultimate

mora, as is usual in the descendant languages, but with the exception of, for example, Marathi. The reconstruction of word-final vowel nasalization is difficult.

The five basic members of the vowel system were supplemented by an epenthetic central ə, which only occurred word-internally, never at the edge of words. Its main source is the Old Indic short *a* in open syllables (actually [ə]). After the disintegration of the unity of Ancient Romani, this vowel merged with others depending on the context, its typical reflex, however, is *e* alternating with ∅. I do not mark stress on the last syllable of the reconstructions. Enclitics following a strong morpheme boundary (+) were unstressed, examples include the secondary case markers that have developed from postpositions and the tense and mood suffixes following the person markers in verbs, e.g. *le manušèskəro* 'the man's' (genitive, possessive adjective), *le rìchèndar* 'from the bears', *la ḍomniàke* 'to the wife', *ov kə̀rla* 'he makes, he will make', *ov kə̀rlas* 'he was making (past imperfect), he would make (present conditional)', *ov kərdiàsas* 'he (had) made (plusquamperfect, past conditional)'. In vocative forms: *dàde* 'father!', *čhavèia* 'boy!'. The verb-forming suffix *+iol* (< OI *-ī bhavati*) leaves stress on the stem: *kə̀rdiol* 'rises from', *kəràdi(u)lo* 'was made', etc. Also in some adverbs: *sìgo* 'quickly, soon', *mìšto* 'well', *fèdər* 'better, rather', etc.

Let us now proceed to the entries that form the core of this paper.

1 AR *(ala)va kərəl* and *(alavo)ḍa kərəl* 'speaks, talks'

Equivalents: SnR *vakērel*, SkR *vakerel* 'mluvit, hovořit, vyprávět, povídat' HŠŽ 255, BR *vatjarav, vakerav, vrakeru* 'gòvorīm, kâžēm, kàzujēm' Uhl 89, 127, GR *rakrel* Ve 171, w i t h o u t a n e t y m o l o g y.

Wallachian Romani lacks this verb, its expected form would be *vakrel* (*vakerd*- perf.), instead a loan from Rumanian, *vorbìj, vorbisārèl*, WRNE *vorbinèl* is used, Ve 175. This famous word — cf. Hu *vakerá/ol, vaker* — is of unknown origin. The second part is suspiciously the AR *kərəl* 'makes, does', but what is the first part, and why do western dialects, for example, German Romani, have *ra-* instead of eastern *va-*? What kind of sound change is this?

At a first glance, *v-* suggests a foreign origin, but there is a more probable solution. The dialectal expressions meaning 'talks, has a conversation', with the structure *duma* f, *vorba* f, *svato* m, *lafi* m, *thavali* f 'word, speech' + *kerel,* or *del* 'gives', *thol, čhol* 'puts' — abundant examples in Uhlik — let us assume that the word under discussion has also been clipped earlier from such a structure. The most obvious possibility is that it is the composition of the plural of AR *alav* m 'word', *alava* + *kərəl*, of which the initial *ala-* has been lost. As regards the German Romani form, it obviously must be explained from the diminutive form of this word, AR *alavoḍo* m 'little word' with the plural *alavoḍa.* In this form the whole of the *alavo-* part has been lost, and the interdialectal *r-* ~ *v-* correspondence is not phonological. If the

GR *rakrel* is indeed < *∅ḍakrəl*, this is also evidence for the fact that the plural of the suffix *-oḍo* was *-oḍa even in Ancient Romani instead of the expected *-oḍe. (The stem itself is T 1373 *ālāpá-* m ← √*lap.*)

2 AR *-àle vocative plural suffix

The vocative case in Old Indic did not have a special suffix in the plural. Accordingly AR *manuš (perhaps *manuša as well) 'people' nom plur ∼ Skr *mắnuṣāḥ*, vs. *manušàle 'people!' voc plur ∼ Skr (=) *mắnuṣāḥ* again. I know of no attempt at explaining the suffix. It may be composed of a plural morpheme, to which an enclitic (therefore stressless) element *+le is added, which is connected to the definite article and is demonstrative in origin. The '+' marks an enclitic boundary assigning stress to the previous syllable. This, however, is mere conjecture.

I think the suffix may originally have been a connective-reciprocal morpheme, similarly to the Hungarian plural suffix *-ék* 'the X family, X's family/company'. (As for the term and the notion, see Hajdú 1969, 1975.) The meaning of Sanskrit *siṃhādayaḥ* is 'the lion and the other animals' (*siṃhá-* m 'lion'), and Pali has a similar construction: *rukkha-gumb-ādayō* 'trees, bushes and so on' (Mayrhofer 1951, 187). Since such an ending would regularly yield *-ādē in Ḍommānī, I consider it rather probable that R (*)*manušàle* derives from Old Indic (*)*mānuṣādayaḥ* used as a vocative.

The peculiar stress pattern of the suffix may be explained by the fact that in languages having automatic and/or simple stress rules stress placement that diverges from the main rules typically occurs in the vocative, imperative and among numerals. (Cf. e.g. WR *štàlvardeš* 'forty', but *dešuštàr* 'fourteen' ∼ Hu *tizën(`)négy* id.) In Balkanic dialects the suffix has the form *-àlen* (for example, in Bosnian Romani), the nasal of which is, of course, the infiltration of the 2nd pers plur suffix: BR *avèn, manušàlen!* 'come, people!'.

3 AR *amisiarəl 'mixes'

This verb is only known from Wallachian Romani, WR *hamìj* 'mixes, gets mixed', *hamìjpe* 'gets mixed', its long forms are *hamisārèl* 'mixes, blends, stirs', *hamisàjvel* 'gets mixed', Ve 67 without the long forms and w i t h o u t a n e t y m o l - o g y, with an erroneous initial *x-*, cf. also BR (*=WR) *hamisarav* 'mêsīm, mêšām' Uhl 168. As regards its form it is a typical loan verb, its source can nevertheless not be found in Balkanic languages. A hypothetical Gk *γαμίζω which could be its base, does not exist. The solution is that it is obviously by analogy that the word entered the class of loan verbs. The short form is secondary, the original morphemes in *hamisārèl* is *hamis-iarəl, while *hamisàjvel* has taken the place of *hamis+iol.

It is an original Indic word, but not the continuation of T 10137.1 *miśrayati*, 10137.2 *miśrāpayati* as in other Indic languages: Pa *missēti*, Hi *misnā* 'to be mixed', but the derivate of the adjective T 10135 *miśrá-* 'mixed':

OI *miśrī karôti* VIII 'mixes' SRSl 513b > AR **misiarəl*,
OI *miśrī bhávati* I 'gets mixed' SRSl 513b > AR **mìsiol*.

The only problem is the initial **ha-*. It may be a prefix, OI *āmiśra-* = *miśrá-*, in this case the *h-* is prothetic. If we take OI *saṃmiśrá-* (= *miśrá-*) as our starting point, we have to hypothesize a lenition *s-* > *h-*, but this use of *h* is unknown in Wallachian Romani. The Classical Persian equivalent of the word is *ā-mēxtan*, and not **hanmēxtan*, therefore we may not think of the influence of Persian *ham/n-* (and *hamē-*) either, the less so since AR **amal* m 'companion' (the etymon of which, ClP *hamāl* < Ir **hamahla-*, OI *samartha-* m) contains the prefix suspected, yet does not preserve the initial *h-*. The solutions seems to lie in the fact that the verb is W a l l a c h i a n Romani and was thus influenced by the Rumanian dialectal (*)*hamestecà* 'to mix' with *h-* prothesis (instead of the standard *amestecà*).

4 AR **arakhəl* tr 'finds < *comes across'

Elsewhere the equivalents of HuR *arakhel*[1], WR *(a)rakhèl* 'finds' are: SnR *lākhel* Ve 18, SkR *arakhel* '1) najít, nalézt, 2) sehnat, zjistit, 3) hlídat, střežit' HŠŽ 30b.

In its origins it may be identical to the HuR verb *arakhel*[2] 'guards, protects', whose etymology is perfect: T 10547 *rákṣati* I, *rakṣitá-* pp, T 1298 *drakṣati*, √*rakṣ* 'watches over, defends', SRSl 535a, SnR *lākhel* is the same, which is interesting because it represents the *l-* doublet of the same verb: T 10883 *lakṣati*, T 10885 *lakṣitá-* pp, SRSl 550a. Its meanings are '1) feels, senses, perceives, 2) gets to know, 3) observes, notices'. The possible connection between 'feels' and 'guards' is shown by the common origin of Hu *érëz* 'feels' and *őriz* 'guards' (see Tálos 1993, 393), and 'finds' may also be linked to them. I would still propose that the Romani verbs meaning 'finds' are the descendants of T 182, 187, 188 **aḍḍ(akk)-* 'gets stuck' > D **aḍḍakkhədi* '*gets lodged on'. Because of the regular reflex *-ḍḍ-* > AR *-r-* the merger with D **ārakkhədi* became absolute.

5 AR **as(i)an* m 'whetstone'

I myself know the word only from Uhlik: BR *(j)asan* m 'belegija, brûs, a device for sharpening scythes' (28, 40). It is certainly connected to the following: T *śānā-* f id., which is the derivate of OI *śísāti//śíśīte* III, *śyáti* IV, *śi/ātá-* pp '1) whets, 2) whets the weapon or horn' SRSl 640a, probably from an *ā-* prefixed version. But one may also think of the composition of OI *aśri-* f 'the edge of sthing' SRSl 81b and OI *áśna-* m 'stone' SRSl 82a: **aśr(y)aśna-* m > D **assannho*. All are derivates of IE **ak̂* (or **ək̂*) 'sharp, etc.'. There are similar words in Iranian languages as well,

e.g. MP /ps'n/ *pasān,* OsI *(y)ssōn,* OsD *insōn(æ)* 'whetstone' ROSl 523b. Later forms of the Persian word are: ClP /'fs'n/ *afsān,* ModP *äfsån* id. PRSl I 106b, the relation of which to the following is not clear: FK *āsyāna* DRSl 45b, KKorm *hasån* m (oblique case *hasén*) KRSl 432a, partly < Ir **upa-sāna-*. The Ancient Romani form is probably **asan* m.

6 AR **ašarəl* 'praises'

This is a well-known word, cf. Pt I 405, 431, II 233, SkR *ašarel* '1) chválit, 2) přehánět' HŠŽ 32a, BR *(a)šarav* 'hválīm' Uhl 100. Vekerdi (20) cites two comparisons of Turner's: T 14827 *āšlāghayati* with ?, *šlåghatē* I, and/or T 1641 *uccārayati.*

Neither of them can be correct; there are, however, two verbs not retained in other Modern Indic languages and not present in Turner, which can be antecedents of this word: one is OI *šålatē* I, *šalitá-* pp, √*šål* 'praises' SRSl 642b, the other is OI *šåthåyatē* X, *šathitá-* pp, √*šath* 'flatters, beguiles' SRSl 633b, to which the *ā-* prefix has to be added. Both are possible sources, *-l- > -r-* has to be hypothesized in the first, which is perfectly conceivable as shown by the existence of the **-(i)arəl* suffix, on the other hand, the regular Romani reflex of *-th-* is *-r-*.

7 AR **ašukiarəl* 'waits < *fasts'

The WR verb *ažukārèl* 'waits' has a palatalized version *ažutjārèl* as well, this version is Ve 179 w i t h o u t   a n   e t y m o l o g y, further variants: *(a)žukarèlpe* (with a complement *pe* prep) 'waits for sbody', *(a)žukaravèl* 'keeps sbody waiting', etc., in other dialects: HuR SnR SkR *užarel* '(po)čekat, očekávat' HŠŽ 274b, BR *doažutjarav, šudjarav, ušugarav* 'čekām, dòčekām' Uhl 50, 65. What is strange (and in any case secondary) in this word is the consonant *-ž-,* which normally is typical of the loan vocabulary. Earlier, I reconstructed AR **ašugiarəl* from a contamination of WR *ažukārèl* and BR *ašudjarel* and hypothesized the metathesis of voicing as an explanation.

Since we are definitely dealing with the derivates of *šuko* '1) dry, arid, parched, 2) thirsty', these variants have in fact developed by d i s s i m i l a t i o n   o f   v o i c e l e s s - n e s s. To be sure this is the *a-* prefixed verb pendent of **šukiarəl* 'makes dry', which is proved by the following Bosnian Romani data: *šutjurav* 'pòstīm' Uhl 276, *sutjaripe* m, *trušuj, trušipe* m 'křst, pôst, svétac' Uhl 144, 275, 462, where the meaning of *truš* f and its derivates is 'žêđ'. It is also instructive to consider 'glàdujēm' Uhl 83: BR *šutjarav e dand,* literally 'I am drying the teeth'. A beautiful example of the coincidence of 'hunger' and 'thirst' is Gk πεῖνα f, which originally meant 'thirst', later rather (and today only) 'hunger', besides λῑμός m/f 'hunger', δίψα f 'thirst'. An illuminating case for our etymology is the origin of AR **nerno* 'sober', which Ve 76 reports as a rare word, in the form HuR *jerno* (Gy) w i t h o u t   a n   e t y m o l o g y.

Cf. also SkR *njerno* 'střízlivý' HŠŽ 190. Izmir (Greek) Romani also has it (*nerno*), Moses Heinschink has called my attention to the word and its origin in 1987: T 7266 *niranna-* 'hungry, starving' SRSl 332, this is a derivate of *anná-* pp ← *átti* II, √*ad* 'eats'. Of the Indian relatives of **šuko* it must be noted that the derivate of Hi *sūkhā* 'dry', the verb *sūkhnā* 'become dry, etc.' has the meaning ' w a i t , be standing in the cold' as well, URSl 525b–526a; while 'keeping sbody waiting' ... *kō intizār mē sukhānā* URSl 515a, literally 'd r y sbody in waiting'.

Finally, it is worth noting that the cluster of Slavonic verbs which this Romani verb "translates", **čekati* (Vasmer IV 325, ĖSlSl Ja 4, 13) is, according to one explanation, cognate with OI *kákatē* I 'is thirsty' SRSl 144, *cakāna-* 'thirsty', on the other hand Ru жcдamь 'to w a i t' and жcadámь 'to be t h i r s t y, etc.' are related as well, cf. Vasmer II 33, 39. Of the langauges of the area Greek could have been the example of the Romani concept, cf. Gk ξεροσταλιάζω '1) I keep standing, I tarry, 2) I get tired in waiting' ÚMK 442a.

8 AR **ašvar* m 'halter'

Wallachian Romani *ašvār* m (Ve 20, without an etymology), BR *šuvar(i)* m/f, *ošvar* m 'ùlār' Uhl 404, and known in other dialects as well, cf. Pt 239–240, Mi VIII 69.

Its connection with the following is certain, but not clear: MP *(zēn)abzār* 'harness' KAP 144, ModP *äbzår* 'tool, equipment' (PRSl l 37), *(zin)äfzår* 'harness, halter' (PRSl I 106, 778). I have no explanation for the -*š*- in place of the -*z*-, the Persian -*z*- may, nevertheless, derive from an earlier -*ž/ǯ*-, which may come from -*č*-. The phonological form of the word may have been influenced by a hypothetical D **aśśəvāro* m 'rider' < T 926 *aśvavāra-* m 'horseman, groom', which is an Iranian loan there, cf. MP *aswār* id. (KAP 109), or Afgh *spôr* m (RASl 94b).

9 AR **azbavəl* 'hurts'

A Wallachian Romani word, but cf. BR (**=WR) *azbav, azbavav* 'dîrām' Uhl 63, in Lovāri the present tense always has the form *azbàl,* in the past the stem **azbav-* → *azbad-* exists as well: *mìndig azbàlasma* 'he always hurt me' imperfect, *atùnći azbadàsma* 'then he hurt me' perfect. Vekerdi's e t y m o l o g y (22): "Rum *izbi* ?" is unlikely even with the ?, as it seems to be an o r i g i n a l word.

AR **z* has no regular Old Indic source, here, however, it is before -*b*- and obviously instead of **-s-*. Turner does not know words existing only in Wallachian Romani. Another lexical archaism, not retained in other Modern Indic languages, is the OI verb *śárvati, śárbati* and *sárbhati* 'hurts, hits, kills' SRSl 638a, 712b, to which Romani has added an **ā-* prefix. Of this Early Ancient Romani had **āsabbədi* > **āsəbbādi* by metathesis and change of stem class, but for the form **azbavəl* an

OI *aśarb(h)āpayati caus starting point may also be hypothesized. OI śarbhati is related to the English verb *harm*, cf. also Av fśarəma- m 'shame', see also the entry Ru *сóром* m in Vasmer III 724.

10 AR *balval f 'wind'

A common word: WR *balval*, HuR *balval*, SnR *bāvjal*, GR *bal/rval*, each f 'wind', Ve 24, Pt 417–418, Mi VII 16.

It is undoubtedly connected to T 11491 *vātá-* m 'wind' (cf. also *vắta-* m id. SRSl 575a), it is still strange, pointing to some sort of reduplication: Ḍ *bādavāda-*. Interestingly it is exactly this word that is found reduplicated in Hazara, where it is *bōdbōt*, with the Russian gloss 'вéяние' (Efimov 1965, 29). The fact that the word has become feminine and reduplicated made it possible to keep it distinct from *bal m/pl 'hair' (T 11572) and *balo m 'swine', the latter of which is of uncertain origin.

11 AR *barvalo 'rich'

This is a well-known word, cf. Pt II 416–417, Mi VII 16–17, Ve 25, whose e t y m o l o g y : "Skr *balavant*" 'strong, mighty, predominant' SRSl 462b c a n n o t b e the antecedent of this word, since that could only lead to something like *bəl(v)o.

There must have been an AR word *bar 'richness', of which the adjective *bar-valo was derived by the common -*valo* adjectival suffix. And there is such a word! Since it does not occur in Turner, the word seems to be preserved only by Romani: OI *vắra-* m 'richness, treasure' SRSl 577a.

It may be mentioned that there are altogether four such near homophones:
(1) *bar m '*richness' — see above,
(2) *bar f 'garden' Pt II 410, Mi VII 17, Ve 24 < T 11480 *vāṭā-* f ← *vṛṇóti*
(3) *baḍ m 'stone' Pt II 409, Mi VIII 16, Ve 24 < T 11348 *vaṭṭa- < *varta- (cf. Afgh *bâṭ* m 'whetstone' RASl 724b)
(4) *-var (multiplicative suffix) < T 11547 *vắra-*[2], but it may as well be Iranian, cf. MP /w'r/ *wār.*

12 AR *beč f 'hip'

The word is reported w i t h o u t a n e t y m o l o g y by Vekerdi 26 as a Wallachian Romani word; I myself only know it from the pseudo-Wallachian North Eastern dialect of Romani (WRNE) thus: *odol bečikūnè masà* 'those thin flanks (of pork)'.

The identification of the origin of this word is connected to *maškar m 'waist', *maškar prep 'between'. The first half is T 9804 *mádhya-* n '1) middle, 2) waist', adj 'central', to which *-kar is added, which is unclear to Turner as well: "(+ ?)", thus its expected *-ǯ- became -š- because of the -k-. 'Between' in Panjabi is *vic, vicāle*

and *vickār*, the latter obviously contains the same enigmatic element as the Romani word. The *-kar* element is known in Bengali as well in the same word: /*madhyēkāra*/ *mɔddekar* 1) adj 'internal, intermediate', 2) postp 'between' BRSl 714b. This *-kar* is a tadbhava suffix usually forming an adjective from an adverb (cf. Bykova–Kolobkov 1957, 26, Ray 1966 does not mention it). Similarly in Parya: *bi/eč-a* 'in, on, within', *biškalo* 'middle (aged)', *biškar-a* 'in/to the middle' Oranskij 1977, 265b, 266a. I think that OI *madhya-* n was probably the etymon of a Ḍ adjective **majjhɔkāro* 'intermediate, middle', which by nominalization acquired the meaning 'hip'. The AR preposition **maškar*, instead of the expected **maʒe,* has developed with this detour. The Sanskrit word already had the meaning 'waist', and in the area Romani has visited during its history 'middle' and 'waist' are often connected in their roots, e.g. B /*mājha*/ mac 'middle', /*mājhā*/ *majha* 'waist', ModP *miyån* '1) middle, 2) waist' PRSl II 586a, ModGk *μέσo* n 'middle', *μέση* f '1) middle, 2) waist'. In this light, it seems probable to me that **beč* f 'hip' < (*)waist' is the etymological continuation of T 12042 *vīcya-* 'middle' via the form Ḍ **becca/i* f.

13 AR **bero* m 'boat, ship'

A well-known word (Pt II 89, Mi VII 19, Ve 26), although rare in the Central and Eastern dialects, HŠŽ 44b gives the word with *, meaning a recommended word from other dialects; BR *bero* m 'brôd, lâḍa' Uhl 39, 150. Its etymology is also known and good, but besides the reconstructed form T 9308 **bēḍa-* other variants are also to be supposed. The *-r-* of the Romani word (instead of *-*l*-) and Ne *berā* NRSl 828b (instead of **berā*) suggest that **-ḍḍ-* is to be supposed in their input.

14 AR ? **birevli* f 'bee'

Its forms are: WR *biri(v)lji, biri/ulji* f, also in Hungary there exists *birovljì* f, WRNE *birilji, birinji* f, HuR *birili, berveli* f (Ve 28), SkR *berevlji, bervelji* f 'včela', *bervel* m 'čmelák' HŠŽ 45a, BR (partly =WR) *birovl(j)i, birovn(j)i, biromni, beru(v)l(j)i, brli* f 'pčèla', *birovljoři* f 'pčèlica' Uhl 248, Pt II 287, Mi VIII 88. Turner derives it from T 11330 *varōla-*. The clusters containing *v* in the second syllable of the Romani word cannot be really obtained from an original **ō*. Based on Ḍ **bɔrōlī* > AR **bɔroli* the frequent *i* reflex of *ɔ* cannot be understood either.

Another possible source has occurred to me, which, however, raises more serious problems. I mention it, nevertheless, because of its interesting consequences. In Sanskrit — I am deliberately not using Old Indic here — there is a word explicitly meaning 'bee': *dvirepha-* m SRSl 295b. Provided that *-li* is some kind of suffix, we may assume a Ḍ **birēvhɔlī* > AR **birevli* f etymon, which yields the attested forms without difficulty. This "suffix" may be the same as that found in the word *madhūlikā-* f 'bee' SRSl 493a (*mádhu-* n 'honey'), the input to the derivation is thus

*dvirēphalikā-. The aspiration of -vh- has apparently not reached the word initial *b-, therefore the word did not become *phirevli. The change *b- > *dv- is possible in Romani, cf. AR *bərand m 'tent pole' < T 6652 *dvarānta-. The counterargument for this word being the source of the Romani word is that Skr dvirēpha- is the result of poetic word formation, it literally means 'with two r's', an allusion to the word T 3651 bhramara- m SRSl 487a also meaning 'bee', Skr rēpha- m SRSl 548a is the name of the letter r in Brahmi and its derivate scripts. It is hard to imagine that this word, meaningful only to literate people (actually only to a Brahmin) could have percolated all the way down to the Ḍōmbāḥ vegetating at the bottom of Indian society, and displaced all its synonyms. The word does not even belong to the terminology of religion, which could justify its Sanskrit origins (cf. Span Dios, Port Deus, but Jud Dioʃ without the influence of ecclesiastical Latin). If the word in question is still a loan from Sanskrit, this modifies our picture of the social position of the Ḍōmbāḥ and the ethnogenesis of the Roma!

15 AR *bivando 'raw'

The word is cited by Vekerdi (76) from Wallachian Romani in the form jivando, its etymology is "Skr vimlāna ?", that is 'withered, decayed' SRSl 523b, which does not satisfy the sound laws: it would give the form *biman.

My starting point is the privative form of the verb pádyatē, the participle of OI vipádyatē IV, vipanná- pp '*does not reach some place', which has provided a new present stem for the verb stem: Ḍ *bivannədi, the participle of which in Ancient Romani is *bivando, in which the semantic change 'unripe' → 'raw' may be supposed. The word has a wide range of variants, which can be explained by the notions of dissimilation and metathesis, but two other similar words have interfered with it: *hi/evend m 'winter' and the participle of *bianəl 'gives birth to', *biando. Cf. BR ivando, (j)ivand, v/bijand 'présan, sīrov' Uhl 289, 338.

16 AR *boləl² 'rotates tr'

We may assume two AR verbs *boləl. The first is well-known, cf. Pt II 422–423, Mi VII 23 'immerses, dips (into water), baptizes', its privative participle is biboldò 'unbaptized > Jew', hence Hu biboldó 'Jew', T 9272 *bōḍayati. There was, however, another boləl² 'turns round', whose derivates possibly include GR *bolipən m '(vault of) heaven < *curve', which in Wallachian Romani became a -del verb: WR bòldel (bòlde imperative, boldinò pp), bòldelpe 'turns him/herself', boldinès 'conversely', Ve 30 without an etymology.

Its source is the OI verb válatē I, ulyatē IV/pass, ultá- pp '1) turns, twists, curves, 2) moves' SRSl 569a. Its vocalism comes from the u of the weak forms, its initial

consonant *b-* is generalized from the forms of the full grade, similarly to the story of *bùchiol.*

17 AR *bùchiol* 'is called, has the name of'

This is a Wallachian Romani verb, *bùšol,* sometimes with the variant *bìšol,* Ve 33 w i t h o u t a n e t y m o l o g y. SkR (with the note "reg") 'jmenovat se křestním jménem' HŠŽ 56b, BR (*=WR) *bušuvav, bušivav, bučhivav, bičijau* 'zòvēm se' Uhl 459. HŠŽ 278a mention the verb *vičinel pes* 'jmenovat se (křestním jménom)', which is obviously the borrowing of SC *víkati víčēm* 'shout, call' ShRSl 52 or Blg *вùкам се* 'be called, named'. Lacková (1992, 70) has the following: *vichinlas pes Lolo* 'he was called Red' and *E čhaj pes vichinlas Sedra* 'The girl was called Bramble', with *-čh-,* which is a nice example of the contamination of the two verbs.

This is another lexical archaism, which does not occur in Turner. We again have a well-known Old Indic verb: *vácati* I, *vákti* II, *úcyatē* pass. 'speak' √*vac* SRSl 560b, SRSl D 901. The entry's antecedent is the passive form, cf. Pa *vuccati,* to which Mayrhofer (1951, 75) adds the note: "'wird genannt, geheißen' pass. zu *vac-* [. . .] mit *v-* nach *vac-.*" Based on this Ḍ *bucchǝdi* > AR *bùchiol* as if it went back to some Old Indic phrase *-ī bhavati.* Other Modern Indic languages preserve only derivates of this verbal root, e.g. T 11199 *vacaná-,* T 11200 *vácas-* n, etc.

18 AR *busno* m 'he-goat'

Vekerdi (33) reports the forms WR *busnji* and HuR *buznji* f 'she-goat', and, following the etymological tradition, derives it from (Cl)P *buz* 'goat'. The traditional etymology is made dubious by a number of problems. One is that 'he-goat' is n o t *bus/z* m, but WR *busnò* m, thus the morpheme structure of AR *busni* f is *busn-i,* and not *bus-ni:* it does not contain the *-ni* feminine suffix attaching to athematics, as the *-n-* of the word seems to belong to the stem. Another problem is that while *-sn-* > *-zn-* is a usual prenasal voicing, in case of deriving the word from Persian we would have to hypothesize devoicing, therefore the version with *-s-* appears to be more original.

We are again dealing with an original word. Turner mentions the word *vṛsan-* m 'male bull' without a number, between 12083 and 12084, which is related to the following: *vṛsa-* m '1) male, 2) bull, 3) man, husband', *vṛsana-* m '1) (sg) scrotum, 2) (du) testicles' SRSl 618a (from the latter comes Hi *basnī* f 'pouch, wallet' URSl 134b), *vṛṣni-* m 'ram' SRSl 618b, the word Ḍ *busi/ǝnau* > AR *busno* m possibly comes from the version *vṛsi/anaka-* of the latter. This word also has its Iranian equivalent, but it is not *buz,* but ModP *goš(e)n* ~ *go/äšen* 'male' PRSl II 398b < CIP *gušn,* MP /*gwšn'*/ *gušn,* Av *varšni-* 'ram' < Ir *wṛšni-* RM 33.

The stem itself is connected to 'wet(ness)', thus, through the stem √*varṣ,* with its derivatives, AR **bəraš* m 'year', and (←) **brišind* m 'rain', similarly to the connection between English *ox* and its OI cognate *ukṣan-* m 'bull' and the meaning of √*ukṣ* 'sprikle' hence 'impregnate'. Therefore, the relationship of *busno* and the other words cited may be conceived of otherwise. It may be a participle *busno* from an AR verb **busəl* (< OI *vṛṣyáti* '*moisten, tup'), with the participial suffix *-no,* marking habitual action. The most common examples are WR *xoxamnò* 'liar' (< **xoxav-no,* AR **xoxavəl* 'lies'; Hu *hóhányó* 'humbug' from its HuR or GR equivalent) and AR **bašno* m 'cock' from **bašəl* 'makes music, crows'. (*Bašno* has obviously displaced an original **kax* m id., which we may assume with great certainty on the basis of **kaxni* f 'hen'.)

19 AR **bust* f 'skewer'

In Ḍommānī the sibilant element of Middle Indic *-st-* and *-ṣṭ-* clusters is retained unlike the reflex in other Modern Indic languages, where the result is **-tth-,* **-ṭṭh-.* It is logical to expect that the Ancient Romani reflex of *-śc-* should be **-śč-*; Ancient Romani, however, does not have words with **-śč-.* This is perhaps because the regular reflex is "irregular," resulting from the simplification **-śc-* > **-st-.*

By supposing this possibility the probable etymon of 'skewer' can be spotted: T 12081 *vṛścika-* m 'scorpion, tarantula' (cf. Hi *bicchū* m 'scorpion' URSl 119b), the Old Indic word as an adjective also has the meaning 'stinging, stabbing' (SRSl 618b), thus there are no semantic difficulties with this etymology. The word is reported by Vekerdi (33) as Wallachian Romani in the form *bus* without indication of gender and w i t h o u t   a n   e t y m o l o g y. Cf. also Pt II 353, 389, BR *bus* f, *bus(t)* m 'rážanj' Uhl 320.

20 AR **čalo* 'full, satisfied'

Wallachian Romani also has it: *ćālò* ~ *ćajlò,* perhaps also with the stress **ćàjlo.* The verb itself *ćāljòl* 'eats his fill' (with the perfect stem *ćāljil-*), in fact, an expected variant also crops up: Ve 26 "*bečajvel* V (thus WR) 'gobble' ", where *be-* is the Hungarian verbal prefix. In Wallachian Romani the AR adjectival suffix **-(v)alo* became *-àjvel* (Lovári), *-àjvol* (Māśāri) by fusion with the verbal suffix **. . . -iol.* Its perfect is *-àjlas* < AR **-àlilo/i.* In this word *-alo* is part of the adjective, of an adjective which itself was once a participle.

The etymon of the word is the following: T 4535 *cakitá-* pp ← *cákati/ē* I '1) eats his/her fill, 2) gets satisfied, 3–4) etc.' SRSl 203a–b, and not Ve 37 cited from Turner, T 5019 **chādayati,* with ?, which has provided a participle (*chādítá-* pp) to another adjective, **čalo²* '*pleasing' (Ve 39).

21 AR *čerhan(i) f 'star'

Its variants: HuR *čerhan* f, SnR *čerhin* (*čerhēnja* pl) f, WR *ćerhàj* f, SkR *čerxenj* f 'hvězda' HŠŽ 68b, BR *če(h)raj(i)n*, *čerharin*, *čherajin* f 'zvézda' Uhl 460, Pt II 197, Mi VII 31, the etymology of Ve 38: "Pers *čarkh*". He surely is not thinking of ClP *čarx* 'circle, wheel' (< PIE *$k^wek^wlós$), is he?

The internal -*rh*- cluster of the word must be the result of some unusual sound change. I suppose that we are faced with a suffixed derivate of OI *ŕkṣa*- n 'star, constellation' SRSl 134b, in which multiple metathesis took place:

Ḍ I *r e cc . h ā n ī

Ḍ II *c e r . h ā n ī   instead of *čherānī

That is *-*r*- and *-*cch*- have changed places in such a way that *-*cch*- has lost its aspiration, or rather, it has stayed in "its place". The phenomenon may perhaps also be explained "theoretically" by modern syllable phonologies. Although *-*cch*- is phonemically o n e unit, it has a syllable boundary (.) within, the rhythmic template of the word has been preserved during the metathesis.

22 AR *čulo m 'drop'

Doubtlessly the original meaning and word class of SnR *čulo* '1) few, 2) cheap' Ve 44, SkR *čulo* (with the note "reg") adv 'trochu' HŠŽ 73a, was AR *čulo m '(*)drop'. This is evident on the one hand because 'trochu, málo' can also be said as *čepo* HŠŽ 68b, which is, of course, the borrowing of Hu *csëpp* ~ *csöpp*. On the other, from its derivates: SkR *čulja/ol* 'téci, k a p a t', *čuljado, čuljardo* 'tekutý', *čuljakerel* '1) neustále téci, 2) k a p a t' HŠŽ 72b–73a, and in the Czech–Romani part (354a) 'k a p a t' in the form *čuljal/vkerel* as well.

Vekerdi 44 has it w i t h o u t a n e t y m o l o g y, Turner has it under 4732, 4877 *kṣulla*-, as if its meaning, 'small, tiny' (SRSl 182b), were original, which it clearly is not. It obviously derives from the following: T 4943 *cyutá*- pp 'fallen' ← T 4948 *cyávatē* I 'drops from, trickles from, falls' SRSl 214b. Within the entry Turner cites the verb "Gy. eur (boh) *čulav*-", but as we have seen he wrongly separates the word *čulo*, which is its participle, from it.

23 A(W)R *čhel or *čhivəl f '(small)pox'

WR *šel* f '(small)pox' (often in the plural: *šeljà*, Ve 155 simply written as *šel*[2], which is, of course, a possible pronunciation) is not homophonous with the word *šel*[1] 'hundred' (< *šəl* < Ḍ *śədo* < T 12278 *śata*- m/n), with which it thus forms a minimal pair to illustrate the phonemic opposition between /ś/ and /š/ — there are no

more true minimal pairs. There exists a word *šeljà* plf 'bran' forming a minimal pair with the plural of '(small)pox' (Ve 155, T 12278 *saldkā-* f).

Other dialects also have the word: SkR (with the note "reg") *chelja* pl 'černé neštovice', with a diminutive *cheljora* pl 'plané neštovice' HŠŽ 74b, BR (=WR) *chel, šjel* f (ɔ: *šel*) 'ὸspa', *chela* 'ὸspice velike', *cheljořa* 'ὸspice mâlē', *xurde chelora* 'bòginje' Uhl 35, 234. Vekerdi follows Turner's etymology: T 12490 *śītalā-,śītalī-* f 'pox', hence Si *sīarī* f, Gu *sīlī* f id. Starting out from this word the expected Ancient Romani form would most probably be some **śi(r)li* f, therefore this etymology is not without problems.

Taking into consideration the words for pox in languages of Eastern Europe — e.g. SC *ὸspa* f < **osɐpa,* from the verb **sɐpǫll/*suti* 'sprinkle', and Hu *himlő,* again from the verb *hint//himlik* id. —I think the Romani denomination has the same notion in the background, it is the derivate of a similar verb, or at least such a verb has influenced the hypothetical **śi(r)li*. Of the numerous verbs for 'throwing, casting, hurling, scattering, sprinkling, putting' the most frequently occurring is AR **chivəl* (with **chito* pp). In its place Lovāri has two intermingling verbs: *šol* 'puts' (*š(uv)àv* 'I put' pres., *šutèm* 'I put' past perf.), and *šùdel* 'throws' (*šùdav* 'I throw', *šudèm* 'I threw', *šudinò* pp 'thrown'). Of these the first can be best linked to T 3683 *ksipáti/e* IV, *ksiptá-* pp 'throws, casts, hurls' SRS1 181a. The form that can be assumed for Ancient (Wallachian) Romani is **chivəl* f, with the plural **chiv(ə)lia*. The parallels for the simplification **chivəl* > *chel* are AR **divəs* m 'day' > WR *djes* or *djēs* (< T 6333 *dívasa-* m) and AR **devəl* m 'God' > WR *del* or *dēl* (with the oblique stem *devlès/n*) (< T 6530 *dēvata-* n 'deity').

24 AWR **-de* imperative particle

In Wallachian Romani, some verbs in the imperative mood can be present or future. Two particles, the stressless *-de* and *-ta* narrow the command to the present, adding a sense of urging to it. Their difference seems to be that the first may be added to verbs meaning departure, the latter to those meaning approach. (They cannot be added to any verb.) Vekerdi cites the example "*aven-dé* 'come along' " (47). It is not only the stress pattern that is wrong here, I feel that (*)*avèn-ta!* is better meaning 'come [here]'. The form *avèn-de!* is a 1stP plur. incl. command: 'come and let's go'. Other examples: *žà-de!* 'go [away from here], go now', *àn-ta muro šerand!* 'fast, bring my pillow', *aš-ta!* '1) wait [a little], 2) keep [quiet], 3) hush'. I do not know the origin of *-ta,* but *-de* is obviously the borrowing of Blg *ðe*. On the Bulgarian particle, see Maslov 1981, 334 (§329), on its origin BER I 328, 334.

25 AR *dilo 'mad'

A well-known word: Pt II 313, Mi VII 43–44, it has entered (partly from Rotwelsch) Hungarian as well: *dili, dilis, bedilizik,* etc. Vekerdi (49) discusses it in the entry *dilino,* to which I think is not related, see the entry *(go)di lino.* H i s e t y m o l o g y : "Skr *dina*? T 6347, 14607". If *dilino* is indeed the clipped form of an earlier *godi lino,* the origins of *dilo* must be sought elsewhere.

The most reasonable assumption is that it is the participle of the verb 'gets mad'. That is T 6512 *dŕpyati* IV 'is mad, is wildly excited', *dŗptá-* pp, √*darp.* But of the participle the expected result is Ḍ *dittau* > AR *dito,* it is very difficult to derive *dilo* from this by the regular sound laws. Obviously Romani had the corresponding verb for a long time, which I imagine to have been AR *diel,* like *piel* 'drinks'. The more regular *dil-* has taken the place of its perfect stem *dit-* by analogy, which thus became *dilo.* The hypothesized verb has been supplanted by the denominal derivation of *dilo*: *diliol* or *dilàliol* (> WR *dilàjvel*), as it has happened with *kindiol* 'soaks', *londiarəl* 'salts', and many other verbs.

Its origin may, nevertheless, be different, especially if its primary meaning was not 'mad' but 'fool', though this is not evidenced by the dialects. In the opinion of Mediaeval Europe, useless, vain, idle deeds were a type of foolishness, as shown by the meaning of WR *šūšò* 'empty, useless, vain' < AR *čučho* < Ḍ *cucchau* < T 5850 *tucchyá-, tuccha-* id. This explains the evolution of Late Latin *follis* 'fool' from *follis* '(empty) bellows' (> OF *fol* (> *fou*) > Eng *fool* 'jester, fool').

OI *dŕti-* m/f '1) skin (for water), 2) smith's bellows' SRSl 283a with the (diminutive) suffix *-aka-* would regularly yield Ḍ *didau* > AR *dilo.* ('Smith's bellows' is *pišot* m on the other hand, the origins of which is as yet unknown. What can be said about it is that it belongs to the archaic stratum and is perhaps not monomorphemic: *piš+ot* with an *-ot* suffix.)

26 AR *dil f 'pēditum audītum'

Its forms are: WR *ŕil* f, HuR SnR GR *ril* f id., SkR *rilj* f 'prd' HŠŽ 237a, BR *ŕil* f 'vĕtar iz têla' Uhl 424, Ve 142 w i t h o u t a n e t y m o l o g y . Cf. also Pt II 277, Mi VIII 61.

It is generally true that as a result of large scale phonetic simplifications and mergers the number of hom(e)ophonous word has greatly increased in Middle Indic languages, and as a consequence many words were confused with each other. We are here dealing with an expressive word, the explanation of which is an unrewarding task. I attempt the following. My hypothesis is that two words have been confused here. As for its f o r m , the word is the participle of OI *dárhati, dŕmhati* l, *dŗdhá-, dŗhitá-, dŗmhitá-* pp '1) ties (up), b e g i n s , 2) strengthens, 3) founds'. Its m e a n i n g , however, is similar to that of the OI verb *dálati* I, *dalitá-* pp ' e x p l o d e s , c r a c k s ,

b r e a k s ,   b u r s t s ' (SRSl 262a, T 6216 *dálati,* T 6508 *dṛdhá-, dṛlhá-).* The word-internal retroflexivization was extended to the word-initial consonant: Ḍ **ḍidho* > **dhiḍi/a* f > AR **ḍil* f. It is regular in all its phonological details: the voiced aspiration shifts leftwards and original -*ḍ*- always yields -*l*-.

27 AR **gad* m 'shirt'

In the Lovāri dialect of Wallachian Romani the plural, *gādà,* means 'garment(s)', which for some speakers is in opposition with a form retaining the short vowel in the plural, *gadà,* which only means 'shirts'. The e t y m o l o g y of Ve 60: "T 4125 *gātraka-*" n 'wing'. This is not a good idea even from a phonological point of view, as it might only result in something like AR **gatro* m.

It is obviously the same as Ne *gādo* 'long garment (worn by Bhutanis)' NRSl 316b. (It does not appear in Turner; neither **gadda-*[1] 'sediment, mud', nor **gadda-*[2] 'spotted, mottled' can be related to it.)

28 AWR **gərkliànos* m 'throat'

'Throat' in Wallachian Romani is *girtjàno* m.   Vekerdi (63) specifies the R u m a n i a n dialectal forms "*gîrtan, gîltean*". BER I, 288 gives the variants Rum *gîrtlan* and Banate *gîrclean,* which, especially the latter, could be the source of our word, it is more probable, however, that Blg *гръклян* (cf. also SC *gŕkljān*) id. can be specified as the i m m e d i a t e source. The sequence **-kli-* before a vowel yields long -*ttj*- in Wallachian Romani; if preceded by a consonant, it of course gives -*tj*-.

29 AR **gilabəl* 'sings'

It is a well-known word, its most original form may be **gilabəl,* the other variants are secondary. These are the palatalized one with -*lj*-, those with a change of suffix to -*av*- or -*ar*-: WRNE *gil(j)àbèl,* WR HuR *djilavèl,* SnR *gilāl, djilāl,* Gurvāri (?) *gabarel,* GR *gīvel* id. Ve 62, SkR *giljavel* '1) zpívat, 2) kokrhat, kukat, cvrlikat' HŠŽ 112a, BR *gilabav, zijabav, gilavav, djilabav* 'pèvām' Uhl 252. The WR perfect *gilabadàs* 'has sung' suggests a latent present stem variant, **gilabavel,* which is recorded by Uhlik ibid., similarly to the verb **azba(və)l.*

The word is possibly a compound, its first component is T 4167 *gītá-* n 'singing, song (acc)', and not the word **gili* f 'song' itself, which is < T 4168 *gīti-* f, *gītikā-* f; the second component is some verb, which might have been something like **(h)abb...-* in Ḍommānī.   Which is this Old Indic verb?   Because t h e e t y m o l o g y of Ve 62 T 4135 *gāpayati* i s n o t s u i t a b l e here.   I am thinking of (T ∅) OI *āhváyati/ē* I, *āhūtá-* pp '1) names, c a l l s (o u t / i n), c a l l s upon, 2) i n v i t e s, 3) causes, brings about, rouses, 4) sends for sbody, summons, wishes' SRSl 106b, 784a. Thus OI (*)*gītam āhváyati* > Ḍ **gīdo abbhēdi,* which were two

words in the time of aspiration shift, since the verb is not *khilabəl. It is also possible that the d i s t a n c e of the initial consonant is responsible for the absence of the shift, thus yielding the regular *gilabəl.

30 AR *(go)di lino 'silly, fool'

The adjective WR dilinò, GR dinelo Ve 49 is usually considered to be the derivate of dīlo 'mad': T 6347, 14607 dīná- 'sad' SRSl 271a. The word *dilo, however, is probably the participle of AR *diel 'gets mad': T 6512 dṛptá- pp ← dṛ́pyati IV SRSl 261a.

I claim that *dilino is a form which, like *vakarəl, has lost a syllable, its original form could have been *godi lino, which literally means 'mind-taken' (the participle of *lel 'take' is *lino). It is thus like Hu eszeveszett 'mind-lost, mindless → mad' or Ru сумасшедший. The etymology of GR gōdi, WR gōdji f 'mind' is known: T 4314, 14456, gōrda- 'brain'.

31 AR *gošni f 'manure, *cowdung'

Its occurrences: WR gošnji f 'dung' Ve 64 w i t h o u t   a n   e t y m o l o g y, BR (?*=WR) gošnja pl 'đùbre, gnôj' Uhl 77, 85.

Similarly to *khoni f, which originally meant 'beeftallow', this word probably also meant 'merda bovīna', its etymology therefore is: OI *gōśaka- n SRSl 631b, 632a, T 4333 gōśakṛt- n, T 4333.2 gōśakara- n 'cowdung', cf. also T 12238 *śakana-. The -n- of *gošni is either a suffix, or, based on the variant gōśakṛt- n, we may explain it by the oblique cases, the word being heteroclitic its genitive is gōśaknás, cf. Gk σκῶρ, gen. σκατός. The Iranian equivalent of the word has an -n-stem in all Eastern Iranian languages and also in Western Iranian languages that got into the area, e.g. Bal saγan, Par saγồn, Orm ᵊskan (Oranskij 111). On the other hand, its fate was similar to that of Slavonic *govьno 'stercus', which was originally an adjective meaning 'bovīnus'. This ancient word is being replaced in Wallachian Romani gunùj m id. originating from Rumanian.

32 AR *haravli f '1) belt, 2) strap'

Its occurrences: WR haravljì f 'belt' Ve 68 w i t h o u t   a n   e t y m o l o g y, BR (?=WR) harvali, aravlji 'rĕmēn' Uhl 322, SkR (*=WR) haravlji f (with the note "reg") 'oprat' HŠŽ 119b.

The part har- may be identified as the following: OI vardhra- m 'belt', n 'strap' SRSl 568a. The second half could be a suffix, but I assume we are dealing with a compound here: OI āvali f '1) stripe, floss (i.e. Docke, écheveau), 2) row, 3) stroke' SRSl 101b. The starting point is thus *vardhrāvalikā-, hence Ḍ *ballhāvalī or due to dissimilation *bārhāvəlī, with aspiration shift *bhārāvəlī, and some way or other

the result of *bh- is *h- in this case, which again may be explained by dissimilation. For the possibility of -rdhr- > -lh- cf. AR *alo, WR jālò 'raw' < T 1340.1 ārdrá-, Pb allā, Hi ālā, Ne ālo, etc.

33 AR *həliel (*xind- perf.) 'cacat'

The verb has a variant with -l- in Hungary as well, besides WR xinjèl there exists xljel, too, with the imperatives xinjì and xlji, and the perfect xind-. Similarly BR xljav, xinav 'òbavīm nùždu vèliku' Uhl 209, SkR xinel 'srát' HŠŽ 127a, it is probably athematic, i.e. with a consonantal stem, cf. also Ve 67–70, where instead of the usual etymon — Gk χέζω — another one is proposed: Gk χύνω, the meanings of which are not suitable here, cf. also Pt II 166, Mi VII 23.

The word, however, as it is apparent from its -l- variant, is Indic in its origin. The details are complicated by the fact that two verbs have been mixed:
(1) T 13960 hádatē 1, hanná- pp id., √had — Turner without the Romani verb, SRSl 769a, which etymologically may be identified with the Gk χέζω mentioned above.
(2) T 3887 khindáti VI, khídyatē IV, khinttē VII, and even khídati I RV, khinná- pp 'pushes' SRSl 186b.

The Pali equivalent of the first verb, hadati ∼ hanati, haññati pass, received a new present stem, *hagg-, by analogy of bhajjaï :: bhagga- in other Modern Indic languages, hence Hi hagnā, Pb haggṇā id., etc. It may be assumed that in Romani the following has happened: *həliel > *hØliel > *xliel, since in preconsonantal position the buccalization of h is a n a t u r a l sound change, just like in the case of *čətar > *čØtar > *štar 'four' the spirantization č > š again in preconsonantal position. Still, the most likely explanation for the -i-vocalism of the word is the influence of (*)khin/dati.

34 AR *huliel (*hulist- perf.) 'gets off, alights

In Hungary the WR verb huljèl (huljil- perf.) is the equivalent of Hu száll, and imitates its prefixed forms: huljèl oprè 'gets on', huljèl tēlè 'gets off', huljèl andrè 'gets in', huljèl āvrì 'gets out'. In the Balkans, where the preverbal use of adverbials is unknown, huljel is only 'gets off', forming a pair with the verb inkljel 'gets on, starts off, goes out', unknown in Hungary, cf. Pt II 326, Mi VIII 25, Ve 74, 91, GR kḷiel 'rides (a horse)'. This is a true pair of verbs, both have a peculiar perfect in -ist-. Vekerdi (72) erroneously spells it with x-, and for some unclear reason gives xuttjel 'jumps' and xutjilel 'catches' as its synonyms, b u t d o e s n o t g i v e a n e t y m o l o g y .

A lexical archaism, absent in Turner, is the OI *huḍáti* VI '1) collects, 2) s u b - m e r g e s , dives, 3) sinks' SRSl 779a. The peculiar perfect suffix is perhaps the reanalysis of the ending of *ahuḍiṣṭa* 'he submerged (medial aorist)' into a tense suffix.

35 AR *xanḍuvəl (*xanḍud-* perf.) 'scratches, scrapes'

In Wallachian Romani (with simple *r* as well) *xaṙùndel* (*xaṙùnde* imperative, *xaṙundàs* perfect, *xaṙundinò* participle), its reflexive form is *xaṙùndelpe* 'scratches himself' Ve 68 w i t h o u t a n e t y m o l o g y . Likewise SkR *xaruvel pes* 'drbat (se), škrábat (se)' HŠŽ 125b, there is also BR *xanudav, xanrudav* 'češēm, grèbēm' Uhl 51, 91, 438. In German Romani *handrel* acquired the meaning 'combs', while *hanǯel* 'scratches' originally meant 'itches'. Vekerdi (67, 68) did not recognize the affiliation of *handrel,* therefore h i s e t y m o l o g y is: "Germ. *behandeln* ?" [sīc!]. The Ancient Romani form was *xanḍuvəl* (with *xanḍud-* pp). T 13645 *skar-* does not fit here, T 2689 *kaṇḍūyáti* on the other hand does, but he only mentions the Romani equivalents in the supplement under the number 14350, citing the forms *xanov-, xaruv-,* WeR *xatav-.* Pott II 167 correctly hit upon the exact origin of the word.

36 AR *xulai* m 'lord, master, peasant'

Wallachian Romani does not know the word, for other forms see Ve 72, as well as SkR *xulaj* m '1) hospodář, pán, 2) bohatý sedlák', *xulanji* f 'hospodině' HŠŽ 130a–b, BR *xulaj* 'gàzda, gospòdār, gospòdin', *xula(j)ni* f 'gàzdarica, gospodàrica, gòspoda' Uhl 82, 88. In its oblique cases the contraction must be rather early: **xula* nom pl, **xulas* obl sg, **xulan* obl pl, GR *xujlo* m has developed by metathesis. Ve-kerdi's etymology is: "Kurd. *xola* ?". What Kurdish word he was thinking of is not obvious, but the word is indeed Iranian and together with **ambrol* f 'pear' is o f s p e c i a l r e l e v a n c e . In the ancestor of European Romani (which I have labelled with the reconstructed **Ḍommānī (Jibbha)*), the reflex of intervocalic *-t(h)-,* *-d(h)-,* even *-ḍ(h)-* is *-l-,* as is in fact well-known. Two Iranian words have participated in this change, one is **ambrol,* mentioned above (Ve 14, its etymology is correct, cf. Pt II 57, Mi VII 6), the other is **xulai:* cf. MP /xwd'y/ *xwadāy,* then *xudāy* 'lord, master', cf. KAP 142, similarly in Parthian *xudāy,* /qdyxwd'y/ *kadexudāy* 'host, landlord' RM 199–120, ModP *xodâ* 'Lord, God' (the non-Arabic synonym of *ällâh, räbb*), as well as KKorm *xwadē, xway* m (KRSl 420a–b), etc. It is difficult to determine whether the borrowers have heard **xwadāy* or already **xudāy,* since **-ə-,* the regular replacement of Iranian *-a-* changed to *u* or *o* in a labial environment and it is not known when. E.g. **xumer* m 'pastry' < **xəmer* < Ir **xamēr,* **tovər* m 'axe' < **təvər* < Ir **tabar,* WR *vurdon* m 'cart' < AR **vərdon* < Ir **wardān* or **wr̥dān* (> CIP *gurdān,* Os *wærdōn*), **pošom* f 'wool' < AR **pəšəm* < MP **pašm,* **por* m

'feather' < AR *pər < MP *parr* < Ir *parna-, etc., as well as (W)R *dudum* m 'squash, pumpkin' < AR *dədum* < Arm *dᵊdum* id. When the latter was borrowed the (D) *-d-* vs. *-dd-* distinction was already lost and the changes (i) *-d- > -l-*, (ii) *-dd- > -d-*, as well as general loss of quantity had all been completed.

Let me note here that it is the I n d i c etymology of AR *rovli* f 'stick' that is correct: T 10875 *lakuṭa-* (> Hi *laurī* f 'penis', B /*larī*/ or /*lōrī*/ *lori* 'stick', etc.), while ModGk ραβδί n id. could only yield *ravdi* in Romani. Similarly Vekerdi's etymology for R *balan/ji* f 'tub' — "Serb. *bâdanj*" (the Slavonic equivalent of Hu *bödön*) is an absurd idea. The origin of WR *balaj, balaji* f id., SkR *balanji* f '1) koryto (i vymleté vodou, 2) okop)' HŠŽ 37b, BR *balanji* f, (=WR) *balaja* f, *balaj* m 'kòrito' (Uhl 137) is known: < Arm *baɫanik'/s*, ModArm *baɣnik'* 'bath(tub)' < Gk βαλα- νεῖον n, L *bal(i)neum* n, cf. Hübschmann 1895, 343 (No. 56).

37   AR **xuxur(mutro)* m 'poisonous mushroom, toadstool'

The word (*xuxùr* m) is rare in Wallachian Romani, *burăca* f (*burǎ/ĕci* pl < Rum *burète* m) is used instead, in Vekerdi HuR *huhur* (72) w i t h o u t   a n e t y m o l o g y, SkR *xuxur* m 'houba' HŠŽ 130, GR *xoxer* id. Sowa 100, etc., it has been known for long, cf. Pt I 69, II 160, Mi VII 65.

The key to an explanation is provided by Pb *kukkar-muttā* m (RPbSl 161a, PbRSl 351a), Hi *kukurmutā* m (Platts 839b) 'mushroom, toadstool'. The first part of the compound in these means 'dog', the second 'ūrīna', their etymons are T 3329 *kukkura-, kurkura-* m and T 10234 *mūtra-* n, more precisely **mūtraka-* m/n. There are similar phrases in other Indic languages as well: Te *kukka-goragu, puṭṭu-kukka-, puṭṭa-kokku, puṭṭa-goṛugu,* Ko *naay-kukka,* Ka *naayi-koṛeyu* (KaRSl 372b), Ta *naay- k-kuṛay* (TaRSl 807b), where *kukka* and *naay(i)* is 'dog', *kokku* is 'marsupial rat', *puṭṭa* is 'anthill', *koṛeyu, kuṛay* is 'shade'. Cf. also T 3211 *kukkuramaradaka-* m, Hi *kuk(u)rōdā* m 'Celsia coromandelina, a plant dogs are fond of smelling before uri- nating'. The semantic background of the denomination has been cleared earlier, see Lévi-Strauss 1970 and 1972 — the reference and the corroboration that I am seeking the etymon of *xuxur* in the right direction I owe to Sándor András Kicsi. B e s i d e s AR **žukəl* ~ **žuklo* m, **žukli* f the word for 'dog' was thus (originally) **xuxur* m, while '(poisonous) mushroom' was **xuxur-mutro* m < OI **kurkura-mūtraka-* m/n (T ∅), see above. Of the compound the meaning of *xuxur* 'dog' was forgotten and the second part became redundant/nonsensical. The meaning of the full compound was then shifted to the first part.

38   AR **-ičho* m (**-ičhi* f) diminutive for animals

Its meaning more precisely is the young of larger mammals. It is doubtlessly the suffixization of T 5026 **chāpa(ka)- >* AR **čhavo* m; Turner, however, does not

mention the suffix. We often find *-i-* as a linking sound (cf. AR **-ibən, *-ipən* <
OI *-tva(na)-* n), but it may also be the feminine *-i* ending of the dam. The feminine
form corresponding to *čhavo* originally also had a **-v-*: **čhavi* f '(Gypsy) girl'. Its
stem consonant (=[*w*]) was deleted (or had palatalized) in oblique and plural forms,
where the *i* was implemented in prevocalic position as a glide: **čhavia *[čha.wi̯a]* >
*[čhaβi̯a] (obl or plur) → **[čhai̯]* nom sg. Obviously AR **ḍoi* f 'spoon' comes from
an earlier **ḍovi* f, cf. T 5573.1 *ḍōva-*. The case of **goi* f 'sausage' may be similar,
but its origins are quite unknown.

The (HuR) words *čhavo* and *čhaj* have entered colloquial Hungarian: *csávó*
'(usually Gypsy) boy' and *csaj* 'girl' (now especially a teenager, without reference to
her ethnic background, thus the feminine equivalent of *srác* 'boy', which is of a Yid-
dish origin), respectively. The Indian background of these words is explained utterly
wrongly by MTESz I 466, since it claims them to be linked to PIE **yuwṇḱás*, which
is continued in AR **ʒuvli* f 'woman' < T 10504 *yuvatikā-* f.

39 WR *intja* 'in that direction, yonder'
In Wallachian Romani, similarly to Hungarian, the set of words for pointing is
not divided into three parts (according to the three persons, as in, say, Japanese), but
only into two, based on the relative distances near vs. far. Thus 'this here with you'
and 'that there with you' are both possible. This distinction, however, is comple-
mented by a contrastive cross classification, labelled demonstrative vs. remonstrative
by John Lotz: Hu *emez* 'this, and not that' vs. *amaz* 'that, and not this'.

The "demonstrative" pronouns and adverbs of Lovāri are the following:

	Near	Far		
	Near	**Far**		
Dem	*kado*	*kodo*	} 'this'	'that'
Rem	*kako*	*kuko*		
Dem	*kathe*	*kothe*	} 'here'	'there' Loc/Lat
Rem	*katka*	*kutka*		
Dem	*kathar*	*kothar*	} 'hence'	'thence' Abl
Rem	*katka(l/r)*	*kutka(l/r)*		

Approximate indication of direction is automatically [−Loc] (this would make it def-
inite). In this case in Wallachian Romani the Dem–Rem distinction also neutralizes:

	Near	Far		
	Near	**Far**		
(Rem)	*akorde*	*intja*	'this direction'	'that direction'

Vekerdi considers the latter to be Rumanian in its origin and proposes Rum *înaìntea* 'before' to be its source, with which it probably has nothing to do. Taking into consideration that in Wallachian Romani (i) -*tj*- could have been earlier *-*k(h)li*-, (ii) initial *in*- could result from the metathesis of *ni-, sometimes *le-, (iii) the -*l* of the adverbial ablative -*al* (sometimes -*ar* < OI -*ātah*) could have got lost (e.g. *te/ala* 'under' < **təlal, anda* 'from' < **andral, pa* 'from, about' < *upral)—the word can be internally reconstructed as deriving form an AWR adverb **niklial.* Bosnian Romani also has *oklja* 'ònámo' Uhl 227, thus I may be searching in the right direction. It is perhaps related to the AR verb *nikliel 'goes to, starts, mounts, rides', which Wallachian Romani does not have, but in Bosnian Romani is *niklivav, likav, intjav* 'ìzādēm, etc.' Uhl 109 < T 7478 *niṣkalati backformation to T 7484 *niṣkālayati,* which has a peculiar perfect stem (cf. AR *huliel), and is uncertain in every respect.

40 AR ? *kari f '*shot'

The derivates of this reconstructed word are known only in western and central dialects: HuR *karjadel,* SnR *kārdjālīnel,* GR *kārja del* 'shoots' (Ve 81, without an etymology), SkR *del karije* (with the note "reg. záp.") 'střílet' HŠŽ 145b, cf. Pt II 109. Also HuR *karjalo* m, GR *karmaskri* f 'gun' (Ve ibid.).

Romani has a nonproductive suffix *-*i* f that forms nouns out of verbs (nōmen actiōnis). There exist three certain examples containing it:

(i) WR *ćōràl,* HuR *čorjal* adv 'secretly, stealthily', SkR *čoral* 'potají' HŠŽ 70b are all archaic ablatives of *čori* f 'theft', which Slovakian Romani has in itself too: *čori* f '1) lup, kořist, 2) krádež' HŠŽ 71. (Similarly in Hindi: *cōrī* f 'theft' URSl 349a.) This is the derivate of the verb *čorəl 'steals' < T 4933 *cōrayati.*

(ii) Bosnian Romani has the word *čumi* f 'poljubac' Uhl 269, and it could have occurred in earlier Hungarian Romani as well, evidenced by Hu *csumi* 'kiss'. In Wallachian Romani it has been displaced by the form *čumid* f, which developed by backformation from the derivates *čumìdel* '(gives) kisses' and frequentative *čumidkerèn(pe)* 'they are kissing'. Although there are no traces of a verb **čuməl,* it is obvious that that there is the noun T 4868 *cumbā*- f, *cumbikā*- f and the verb T 4870 *cumbayati* behind them.

(iii) Only in the second part of a compound can the OI *mārikā- be spotted, which is cognate with the AR pair *marəl 'die' < T 9871 **marati, mriyātē* VI and *marəl 'beats to death (today: beats, hits)' < T 10066 *mārayati.* Even here it shows up in its lenited form, -*m*- > -*v*-: in AR *manušvari f 'gallows', originally literally 'homicide' < T 14746 *mānuṣamārikā- f.

The fourth is the hypothetical AR **kari,* which, I believe, comes from a derivate of T 2778, 13645 *kṛṇ́ŏti//kṛṇutě* V, *kṛṇ́ắti//kṛṇītě* IX, √*kṛ³* '1) wounds, 2) kills' (SRSl 150a): an OI *kārikā- f. Of the Romani words the adverb *karie contains the

Loc/Lat suffix -e, the variant karja- the ablative suffix. Ancient Romani might have had the verb *karəl 'shoot' for it, which is even more likely given GR karmaskri f 'gun' < AR *karibnàskəri.

Modern Greek has χαριοφίλι n (ÚMK 309a) '(kind of) gun', which is the borrowing of It Carlo e figlio (cf. BER II 241). The word cannot be derived from this (solely).

41 AWR *kekerȁška f 'magpie'

The WR Lovāri form is kekerȁška f '1) magpie, jay, 2) screech owl, 3) (in Vekerdi) vulva', it also occurs in Hungarian Romani, where it is probably a Wallachian Romani loanword. BR karakaška, karagačka, kakaraška 'svráka' Uhl 372. Bulgarian also has the word in the form κακαρὰcκα, καρακὰuκα f id. BER II 151. It is obviously Greek (ModGk χαραχάξα f), but the vocalism of the Wallachian Romani word requires some explanation. Now, it is needless to assume a quasi Rum *căcăràşcă f mediator — although the the existence of such a form cannot be excluded —, because, as Mohay notes, the tendency for the dissimilation α...ά > ε...ά is one of the characteristics of the Modern Greek vernacular (ÚMK 716). We may therefore start out of a dialectal Greek form *χερεχάξα as well.

42 AR *kərlo m 'sound, voice'

Its occurrences: WR kirlò, GR kurlo, SnR HuR kello m only with the meaning 'throat', the etymology of Vekerdi 90: "Serb. grlo". Cf. also SkR kirlo m '1) hrdlo, 2) hlas' HŠŽ 151b–152a, BR krlo m '1) gȑlo, 2) glâs' Uhl 83, 92, as well as "Krlo e Romengo" [The Voice of the Roma] — the title of a long-gone newspaper from Yugoslavia, also Pt II 96, Mi VII 89. The similarity of the word may indeed seduce one to derive it from Slavonic, but there is more than one reason why this cannot be so. We are dealing with a final-stressed thematic stem, which excludes its being a late loan word. On the other hand, nothing justifies the g- → k- "sound substitution" or "development" in it. Finally of its meanings 'voice' seems to be more original and only the meaning 'throat' could have developed under Southern Slavonic influence.

It is exactly because of the original meaning and the stem type that I consider it an Indic word. AR *kərlo < Ḍ *kərilau < *kəlilau < *kəlidau < OI kalitá- pp ← T 2914 kálatē I, kaláyati/ē X, √kal[1] and √kal[2] '1) gives a s o u n d, sounds, 2) counts, assumes', as well as '... 7) emits a s o u n d, etc.' SRSl 153. It is also probable that in the Ḍommānī word the -l- > -r- change took place a f t e r the lateralization of the -d-, therefore by dissimilation.

43 AR *kindo[2] 'wet, soaked'

This is a widely spread word together with its derivates: WR kindjārèl 'soaks', kìndjol 'gets (thoroughly) soaked', SkR cindo[2] 'mokrý', cindjol[2] 'namočit se,

zvíhnout, zmoknout, mokvat (o raně)', *cindjarel* 'namáčet, navlhčovat' HŠŽ 62a, BR *tjingo* 'mòkar' Uhl 173 — backformation with -*g*- instead of -*d*- from the verbs *tjindjarav, tjindjivav.* The e t y m o l o g y of Turner 5812 *tinta- (and following him that of Vekerdi 89) i s m i s t a k e n , the archetype had *k-: Pt II 103.

This adjective also derives from a participle, that of the verb *kinəl[2], the etymon of which is T 3620 *klídyati* IV, 3622 *klinná*- pp, √*klid* 'gets wet, gets soaked through' SRSl 178b. It is thus like *londo* 'salty' (Ve 102), behind which (today) there is no verb **lonel,* '(sprinkle with)' salt' is WR *londjarèl.* It is again the participle that provides the verb with a present stem, just like in the case of *pekəl 'bakes, roasts', *phagəl 'breaks', *čhinəl 'cuts', *mukhəl 'allows, yields' and many other verbs.

44 AR *kiravəl 'cooks'

Although German Romani also has this verb in the form *kiervel,* it is a typical Wallachian Romani word in place of the verb *tavəl id. of most other dialects, which Wallachian Romani does not use (Ve 80, 89–90). Although the usual etymology, T 3635 *kváthati* I 'boils' from the retroflexivized causative form *kvaṭhāpayati — in Middle Indic languages it did in fact retroflexivize: Pkr *kaḍhia*-, Ś *kaḍhida*- 'boiled' — , is possible, I propose that it is to be derived from something else. Regularly *kvaṭhāpayati would yield Ḍ *kəḍhāvēdi > AR **kəravəl,* merging thus with the causative form of *kərəl 'makes, does', which is T 2418 *karóti//kuruté* VIII, *kriyaté* pass, *kṛtá*- pp, √*kar*[1].

There exists another stem √*kar*[2], which is in fact √*kṛ,* 'throws, hurls, tosses': T 3172 *kiráti//kiraté* VI, *kiryáté* pass, *kīrṇá*- pp 'scatters, pours out, sows' SRSl 150a, and this is a perfect etymon for the verb in question. Semantically, we have to consider the analogy of It *butta la pasta adesso* 'cooks the pasta now' literally ' t h r o w s the pasta i n (the pot) now'. WR *kìrol* 'cooks (intr)' (< *kìriol) is, of course, an analogical form beside *kiravèl.*

45 AR **koḍ(i)* f 'neck'

'Neck' in western dialects is *men f < Ḍ **meññi/a* < T 9732, 9858 *mányē* du, *mányāh* pl SRSl 498a, with the palatalization *a > *e before **ññ.* In Wallachian Romani dialects it is replaced by the athematic *koŕ* f or thematic *koŕì* f (WRNE). SkR *kor* f 'hrdlo' HŠŽ 154b is provided with the note "reg". Uhlik's (430) forms *koŕ, kox* f 'vrât' are probably Wallachian Romani, while *men* f (ibid.) is Albanian Romani.

The etymon Vekerdi cites is T 3607 *kródá*- m 'breast' with a "?", which is in fact the etymon of R *kolin m/f 'chest'. Interestingly, the same word is given as the origin of **koḍ(i)* as well (Ve 93). The only source of Ancient Romani intervocalic *-*ḍ*- is OI -*ṭṭ*-, but I presume that in the antecedent dialect of Romani sometimes -*nṭh*-

denasalized to -*ṭṭh*- too, and thus also became AR *-*ḍ*-. I base my presumption on T 2680 *kaṇṭhá*- m 'neck, throat' SRS1 146. The corresponding Hi *kanṭh* m 'neck, throat, Adam's apple, voice' URS1 669a instead of the expected *kāṭh* does not seem to be a popular word, it obviously is Sanskrit.

46 WR *kropáco* m (pl -*uri/a*) 'pole (on a cart)'

Ve 95 w i t h o u t   a n   e t y m o l o g y. It is the borrowing of Rum *proṭàp* n (pl -*uri/-e*) id., whose source is Blg *пробпп*. In Uhlik (326) BR (=*WR) is also *kropaco* 'rúda (kôlā)'. I do not have data for Romani *procàpo*. Where did the metathesis happen? In Romani or already in the dialect from which it borrowed the word? The latter case, evidence for a dialectal Rum *cropàṭ* n, would help discovering the precise location of the ethnogenesis of the Wallachian Roma.

47 AR *kurko* m '1) Sunday, 2) week'

Of the two meanings of the word 'Sunday' is primary, the second meaning developed from 'the seven day period from Sunday to Sunday', which is connected to the similar meaning of the plural (τὰ σάββατα) of Gk σάββατον n, from the times when the week lasted "from Saturday to Saturday." It is a well-known word with a well-known origin (cf. Pt II 116, Mi VII 88–89, Ve 96). Although it ends in a stressed -*o*, it is a Greek loan. Its category, however, is not original, it probably used to be an adjective, similarly to its source. There must have been an AR phrase *o kurko divэs* m 'Sunday', the calque of Gk (ἡ) κυριακὴ ἡμέρα 'Diēs Dominica, the day of the Lord'. (The resurrection of Jesus fell on a Sunday, which was to be expected according to stories about dying and reviving gods, cf. L *Diēs Sōlis (Orientālis)*.) Greek also has ἡμέρα κυρίου with the same meaning, which justifies assuming an AR phrase *o divэs le r(ai)èskэro* or *o divэs le xulaièskэro* (or *xulàskэro*). It is not clear how κυριακή came to be stressed on its final syllable in Romani (the fact that it was so in Greek is not sufficient reason) and why it has lost its internal syllable. Furthermore, why was its υ substituted by *u* and not the usual *i*? (Was it still pronounced as *ü*?) It is clearly a very early borrowing and perhaps from Armenian, its source is then Arm *kiraki* (Grabar *kiwrakē*), in which there is no trace of the iota any more (cf. Hübschmann 1895, 357, No. 205).

48 AR *khangэri* f '(Christian) church'

This is a word primarily of eastern dialects, elsewhere it is a loan, WR *khangerì* f 'church', SkR *khangeri* f 'kostel' HŠŽ 159a, BR *kangeri, khandjiri* f 'cîrkva' Uhl 45. The often proposed Persian source cannot be correct: ModP *kongor, konge/ä/ore* 'loophole, crenellation (of a castle)' PRS1 II 362a, FK *kango/ura* id. DRS1 594b. The

word-initial *kh-* does not make this etymology probable. Therefore, the other etymology of Vekerdi 86—"*ghaṇṭāgāra* 'belltower' " is semantically a better proposal, yet it is not good because it would yield something like **khanɼar.*

Indic languages have countless onomatopoeic words for smaller and bigger versions of bells. It is usual to sew bells on the skirt of dancers, but even without this the association of bells and skirts is obvious by their shapes, which resulted in many words now meaning 'bell', now 'skirt'. The most probable source is T 4444 *ghargharikā* f 'bell' SRSl 201a, which absolutely plausibly yields **ghaggərī* first, then or immediately **ghaṅgərī,* and finally AR **khangəri* f, cf. Pkr *ghagghara-* n 'string of bells', Hi *gāghrī* f, Pb *ghaggrī* f 'skirt'. Obviously, when the Roma reached the Christian world, the most charasteristic novelty of churches was the bell, which gives the reason for the semantic shift 'bell' → 'church'.

49 AR **khan(i)* f 'pēditum inaudītum'

R **khand* m '(bad) smell, stink' is etymologically all right: T 4114 *gandhá-* m/n SRSl 189a, which rather is 'fragrance'. In many dialects it has a variant *khan* m, without the *-d* by reanalysis of *khàndel* 'stinks, smells badly, is putrid': *khan+del* ← *khand+del*. There is, however, another word with which this has apparently been confused in Bosnian Romani: *khan(d)* f(!) 'smrâd, smrâdež' Uhl 347, elsewhere the two words are kept apart: SkR *khand* m '1) puch, 2) plyn'; *khan* f and *khanji* f 'prd' HŠŽ 158b, 159a and in Hungary WR *khāj* f 'pēditum inaudītum', which is n o t r e c o r d e d by Ve 68, but appears in DF 32 with the gloss 'szellentés'. A derived verb also belongs to the latter: HuR *khanjarel* (hence Hu *kanyerászik, kanyerál*), SnR *khanjārel* 'breaks wind' Ve 86 w i t h o u t a n e t y m o l o g y and the Wallachian Romani equivalent, *khajārel* id.

Its etymon is easily identifiable: T 4531 *ghrāṇā-* f 'smell', *ghrāṇa-* m/n '1) smell, olfaction, 2) nose' SRSl 203b, AR **khan(i)* f and Gu Ma *ghāṇ(i)* f 'stink' are both from the f i r s t variant, Turner does not cite the Romani equivalents.

50 AR **khoni* f 'tallow'

The Wallachian Romani form of the word is *khōjì* f (*khōnji* is not a Wallachian Romani record) Ve 88 w i t h o u t a n e t y m o l o g y. Elsewhere: SkR *khonji* f 'lůj' HŠŽ 161a, BR *koni* (ɔ: **khoni*), **=WR khoj* f 'lôj' Uhl 156.

Its origin is the following: the first part is T 4255 **gɔ̄-* m/f '(of the) cattle', the second part is T 13802 *snēha-* m '1) stickiness, oiliness, 2) glue, 3) grease, 4) etc.' SRSl 757a. Of the conjectural **gɔ̄snēha-* the Ḍommānī form is regularly **gonnhī* f, as if its antecedent were **gɔ̄snikā-*. Then by the leftward shift of the voiced aspiration **gonnhī* > **ghonnī,* finally AR **khoni.*

The leftward shift of the voiced aspiration can be witnessed in the following etymologies:

— AR *čhib* f 'tongue, language' < Ḍ *jhibba* < *jibbha* < T 5228 *jihvā-* f
— AR *čhon* m 'month' < Ḍ *jhonno* < *jonnho* < T 5301 *jyautsna-* m
— AR *khabni* f 'pregnant woman' < Ḍ *ghabbiṇī* < *gabbhiṇī* < T 4062 *garbhiṇī-*
— AR *khand* m '(bad) smell' < Ḍ *ghando* < *gandho* < T 4014 *gandhá-* m/n
— AR *khul* m 'fæcēs' < Ḍ *ghūdo* < *gūdho* < T 4225 *gūtha-*
— AR *khuvəl* 'weaves' < Ḍ *ghuvədi* < *guvhədi* < T 4205.1 *guphóti*
— AR *phandəl* 'ties' < Ḍ *bhandədi* < *bandhədi* < T 9139 *bandhati* ← *badhnātī/badhnīté* IX, √*bandh*
— AR *phivlo* 'widow, unmarried' < Ḍ *bhivəlau* < *bidhəvī* T 11752 *vidhávā-* f, *vidhutikā-* f
— AR *phuro* 'old' < Ḍ *bhuḍḍau* < *buḍḍhau* < T 9271 *vṛḍḍhaka-* < *vṛddhá-*, √*vardh*
— AR *thar* f 'molar tooth' < Ḍ *dhaḍḍi* < *daḍḍhi* < T 6250 *dāḍḍhā-* < *dáṃṣṭrā-* f
— AR *thud* m 'milk' < Ḍ *dhuddo* < *duddho* < T 6391 *dugdhá-* n ← *dṓgdhi* ll, etc. √*duh*
— AR *thužəl* 'milks' < Ḍ *dhujjədi* < *dujjhədi* < T 14613 *dúhyati* IV, etc. √*duh*

On aspiration shift see Turner 1959b, partly erroneously, and Aichele 1957.

51 AR *lakhliarəl* '1) perceives, feels, smells, 2) understands'

The verb 'understands' is a simplified form in most dialects: HuR SnR *hajol,* GR *hajel* (Ve 69), a more original form is represented by Slovakian Romani: *axaljol* '1) rozumět, chápat, 2) mínit, myslet', *sar oda axaljos?* 'jak to myslíš?', *sar tut axaljos?* 'jak se cítíš?' HŠŽ 25a. The etymology Vekerdi gives is correct (T 1040 *ākhyāta-* pp ← *ākhyáti* II, cf. SRSl 88b–89a, 187a), but does not refer to the WR verb *hatjārèl* usually considered to have the same origin, meaning 'understands' as an interjection: . . . *hatjārès(?)* 'you know(?), you see(?)', its more fundamental meaning is 'perceives': *hatjāràv hodj šil si* 'it is cold, I feel it', *hatjārèlpe zurálèske* 'feels himself strong' (Vekerdi's sample sentences).

The word-internal Wallachian Romani -*tj*- cluster can result from the palatalization of five different consonants or consonant clusters: (i) *t*, (ii) *k*, (iii) *kl*, (iv) *tl*, (v) *khl*, as is copiously exemplified. I suspect that in this case we are also dealing with the derivate of a participle: while *axàliol* contains the *+iol* med suffix, this has *-iarəl* tr. Both are logical, since perception is an inactive process on the one hand, and it can have an object on the other. That we have the suffixation of a *. . . khlo* pp is apparent from the Bosnian Romani equivalent. Besides *hatjarav* there also exists

atjharav 'prìmētīm, zàpazīm, slûtīm' Uhl 345, 444. The participle which is the input to the derivation is obviously that of AR **lakhəl,* which I have mentioned in connection with **arakhəl.* It seems probable, however, that first the lateral dissimilation **lakhlo* > **ǿakhlo* took place, after the productive verb **lakhəl* had ceased to exist, and **-iarəl* has been added to this. Therefore it appears to be more appropriate to hypothesize the verb **akhliarəl* instead.

52 AR **leǯal* (**legl-* perf) 'takes, leads'
    Its origin can be considered known (in Vekerdi it is without an etymology): it is a compound of the stem of AR **lel* (**lino* pp) 'takes, grasps' (T 10948) and **ǯal* 'goes', the perfect stem of which is suppletive (T 10452, T 4008). A similar structure is well-known from other Modern Indic languages: Pb *lai jānā* (*lainṇa* + *jānā,* RPbSl 976b), Hi *lē jānā* (*lēnā* + *jānā,* URSl 735a), Ne *laijānu* (*linu* + *jānu,* NRSl 1003). Its paradigm in Ancient Romani could have been the following:

'to go'		'to take, lead'		WRLo 'to carry'	
*ǯav	**ǯas*	**leǯav*	**leǯas*	*ingràv*	*ingràs*
*ǯas	**ǯan*	**leǯas*	**leǯan*	*ingrès*	*ingrèn*
*ǯal	**ǯan*	**leǯal*	**leǯan*	*ingrèl*	*ingrèn*
*gəliom	**gəliam*	**legliom*	**legliam*	*ingerdèm*	*ingerdàm*
*gəlial	**gəlian*	**leglial*	**leglian*	*ingerdàn*	*ingerdàn*
*gəlo/i	**gəle*	**leglias*	**legle*	*ingerdàs*	*ingerdè*

In other dialects: WRNE *ligrel* (*ligerd-* perf), HuR *liǯel,* SnR *lēǯel,* SkR *ljiǯal* (*ljigend-* perf) 'odnést, odvézt' HŠŽ 168a, GR *hiǯrel,* BR (partly WR) *li/edjarav, (l)indjarav, ingalav, ingerav,* etc. 'nòsīm, donèsēm, odnèsēm, vòdīm, odvèdēm' Uhl 67, 206, 219, 429. During its history in several dialects it has by analogy entered the group of old ... +*karõti* compound verbs, in which **-kərədi* has become voiced initially because of the preceding *n,* cf. e.g. AR **čhingrəl* (**čhingərd-* perf) 'tears, slashes' (T 5046). In WR *inkrèl,* BR *intjarel* 'holds' the change *-nk-* > *-ng-* has already become inactive by the time of the prefix metathesis. Its etymon could not have been T 7474 *níṣkarõti* because of its meaning ('expels, breaks into pieces' SRSl 347a), instead it is also a compound with the stem of *lel:* AR **lekrel.*
    HM talk about the verb "*irigrel*" with a similar meaning (76a), which they even conjugate (40), and which Vekerdi records in his dictionary, with the note "rare" to be sure. There exists no such or similar verb in Lovāri or other Wallachian Romani dialects. The most likely story of this verb sheds light on the nature of Gypsy philology. It was solved by László Szegő some 25 years ago. He claims that Hutterer

and/or Mészáros have read an old handwritten glossary, in which a f l y s p e c k had got above the word *ingrel resulting in the -n- > -ri- "sound change".

53 AWR *lunǯisarəl 'hands (over) sthing to sbody'

In Wallachian Romani dialects the following words are in general use: i/unzārèl 'hands (over)' and i/ùnzol 'reaches out for sthing', I have not met a word for 'touches'. Vekerdi records only the latter, intransitive form without an etymology.

Formerly I have suspected it to be of an Indic origin. I have thought of the derivate of OI nihsárati I, nihsísarti III, nihsrtá- pp '1) flows out, 2) goes out, 3) steps up, 4) appears' SRSl 350, something like: *nihsārati, *nihsārāpayati. Via AR *nisarəl this could only get to AWR *insarəl, thus besides obvious semantic problems there is no explanation for the voicing in Wallachian Romani -ns- > -nz-. The fact that this word is not known in non-Wallachian Romani dialects suggests that we are dealing with a word of Rumanian origin. It is probable that in this case it was an initial Rumanian lu- that fell victim to Wallachian Romani prefix metathesis. The word is the borrowing of Rum a lungì (lungèsc, lungìt pp) '1) lengthens, 2) stretches, 3) o f f e r s, 4) stretches out, 5) pulls (an ear), 6) extends, 7) dilute (water)' DRM 1 741b. In Romani this regularly becomes *lunǯisarəl (long, Greek-style form) or *lunǯi(nə)l (short form). The first developed into *lunzarəl with the e x c e p t i o n a l d e l e t i o n of -i-. The intransitive a se lungì (with the adequate intransitive and medial meanings) has accordingly become *lùnzol. This was followed by *lun- > *in-. WR lunžìj 'lengthens' and inzārèl thus have the same source. The corresponding forms in Uhlik are:
  (i) in(d)zarav, anzarav, unzarav, izarav 'prûžīm, pròtēgnēm, pròstrēm' (303, 305),
  (ii) lundjarav, lungarav, inzarav 'nàstavīm, pròdūžīm, razvúčēm' (299),
  (iii) lundjivav, lundjarav ma 'pròdūžīm se' (299), lundjivav 'ràstēgnēm se' (314),
        inzivav, unzivav, izivav, anzivav 'pròtēgnēm se, òpružīm se' (306).

54 AR *mangən m '1) dowry, heritage, 2) property'

In Hungary WR mandjìn m 'treasure' Ve 105 w i t h o u t a n e t y m o l o g y. Its earlier meaning is preserved probably by BR (*=WR) mandjin m '1) blâgo, bògatstvo, imánje, imétak, imòvina, 2) mìrāz' Uhl 33, 102, 170.

It is because of its meaning that it may be assumed to be connected to the verb *mangəl 'ask for', which < T 10074 mắrgati/ē 1, mārgáyati/ē X, mārgita- pp '..., 3) demand sthing of sbody' SRSl ∅, SRSl D 930a. Its etymon is: T 10073 mārgaṇa-n 'search, investigation' SRSl 510b, but obviously also '*dowry', cf. Hi magnī f 'asking in marriage', Pb maṅgnī f RPbSl 634a. In Romani *-ən may have changed to -in (by analogy), e.g. AR *sostən f > HuR sosten f (Ve 153), WR sostjin f 'pants'

< T 13468.2 *sumstanā-, but it could also have remained -en: AR *čikən m 'grease'
(< T 4782 cikkaṇa- SRS1 210b) > HuR čiken, WR ćikèn m id. Ve 42.

55 AR *miàsekos m 'month'

Turner has this word among those corresponding to 10104 OI mãsa- m in the
form másek with the note "lit. one month", that is, as if the -ek part was the AR nu-
meral *ek. The order of the elements in the compound makes this unlikely. Based on
Vekerdi's HuR SnR masek 'month' and trin masekengero 'three-month-long' (107)
this is a substantīvum palæocliticum, thus it could even be an original word. Uhlik's
(167) masek djive 'mĕsēc dâna' also shows the word to be athematic (with a con-
sonantal stem). However, the following is read in Jašari's (Albanian Romani) play:
"Kurkenca thaj masikonca (instr plur) na dikhljum tut, sar te nakhavelahine i kalji
phuv" (6) 'I haven't seen you for weeks, for months as if you were devoured by the
black earth', thus here it is a substantīvum xenocliticum.

This is probably not a part of the Indic heritage, but the borrowing of Blg месец,
in fact, *месек m, and it has the following history:

S i n g	P l u r	Changes in rules and paradigms
*miàsecos	*miàsecia	
*miàsekos	—"—	*k replaces *c on analogy of stems in the alterna- tion Sing k :: Plur c — perhaps in the Bulgarian source already
—"—	*miàsekia	*k replaces *c on analogy of the non-alternating stems
*miasekò	*miasekà/è	Sing -os :: Plur (')ia is replaced by Sing -ò :: Plur -à (and perhaps later -è)
*miasèk	*miasèk(à)	Sing -ò :: Plur -à (later -è) is replaced by Sing -∅ :: Plur -à (later -∅)

The Hungarian and Slovenian Romani equivalents of the last phase are examples of
the total Romanization of the word. We probably have here an ancient borrowing
from Bulgarian, since the reflex of Sl *ĕ is always i in them, suggesting a Serbo-
Croatian dialect with this feature as a source, cf. HuR svito m 'world' (Ve 157) <
SC (*)svît < Sl *svĕtъ, etc. I can only assume that in Albanian Romani the word in
singular is *masiko, its plural I do not even dare make a guess about. Since there
exists a variant Blg заек 'rabbit', while Macedonian has зајак m sg, зајаци pl, the
form *месек as input is well imaginable (cf. Vasmer II 84).

56 AWR *mòsura f '(hideous, ugly) frightening mug'

In connection with WR mòsura f Vekerdi gives simply the gloss 'face' (111, without an etymology), but it is not a synonym of the word muj m (mos/n- obl) '1) mouth, 2) face'. Its use is extremely offensive, its inappropriate use must be avoided.

I have tracked it down in Bulgarian: мѝсурѣ f 'муцуна на говедо' Ralev 1977, 147a and мѝтра f 'грозно неприятно лице' Šklifov 1977, 267.

57 AR *musarəl 'wastes'

The Wallachian Romani form is musārèl '1) spoils, ruins, 2) wastes, spends (money)' Ve 113. Its connection to the verbs HuR ro/uminel, WR ro/umìj, rumusārèl, WRNE rumisārèl, GR rumierel and WRNE rumisàjvel is unclear. Ve 145 links musārèl as well to the Gk verb ῥίμαζω m 'I get worse'.

Its fate is probably parallel to that of *hamisiarəl: it is an original word, which got into the class of loan verbs (verba xenoclitica) by analogy: T 10298 mr̥sā adv 'un-necessarily, in vain', hence *mr̥sā karði 'does sthing in vain → *wastes → *spoils', similarly to OI mōghī karði id. SRSl 522a, which this verb cannot be derived from.

58 AR *musi f '(upper) arm, biceps'

A word of the W and C dialects: HuR SnR GR musi (probably *mūsi) f id., SkR musi f 'paže' HŠŽ 185b, BR musi f 'rúka' Uhl 326, interestingly in the form musikh f 'arm' in Kosovo Albanian Romani: "Djala oljeste thaj doljelalje andar i musikh" Hilmi Jašari 5 'goes up to him and takes him by the arm', the idea is clear, the following can also be read in the same place: "i djuvlji doljela e murše andar i baj e vastesiri thaj na mekhelelje te djal olatar" 'the woman takes the man by his sleeve and does not let him leave her' with the usual ablative of the locus tactī. Vekerdi 113 repeats the bad etymology of Pott II 457–458: he identifies the etymon as T 10221 musṭí- m/f, which means '1) fist, 2) hand(ful)', cf. Hi mūṭh f '1) fist, 2) hand, 3) hilt, 4) handle' URSl 793a, which is indeed of this. In Romani the Old Indic word could only yield a word like *mušt(i) f.

This word is a feminine form derived from AR *muso m 'mouse' < T 10258.1 mūṣaka- m, obviously on the analogy of Blg мѝшка. The view on which the deriva-tion is based is an etymological commonplace, cf. L mūsculus, Gk μυών, ModGk μυς, ποντίχι, OHG mūs, Ru мѝшка, as well as T 10261 *mūṣala- 'mus-cle'. Namely, when bending the arm the biceps runs upwards like a mouse. This is similar to comparing the lower leg of humans to a fish: the calf ends at the "cloaca of the fish", where the disembowelling of a fish is begun. The most obvious lin-guistic reflection of this are the Slavonic words *jьkro/a[1] and *jьkro/a[2]. Arabic

also provides evidence: *baṭṭaTU ssaql* (or *rriğll*) 'calf' is a derivate of *baṭṭa* 'disembowels' (ArRSl 90a). Also Persian *måhiče* 'biceps < *little fish', *måhičeye på* 'calf' PRSl II 447a, Afghani *māhigák* 'calf', *māháy* 'fish' RASl 240a, 643a. And in B o s n i a n   R o m a n i: *prnehko mačho* 'lîst (na nòzi)' Uhl 560 — SC *lîst* m 'Pleuronectes flesus' — literally 'the fish of the leg'.

59 AR **nasul* adj 'bad, evil', m 'evil (man), the Evil One'

The word *nāsùl* is primarily Wallachian Romani, cf. also Mi VIII 23, Vekerdi's etymology "Gk **ἀνάφελος* ~ *ἀνώφελος*?" c a n n o t   b e   c o r r e c t. Cf. also BR (?=WR) *nasul* 'zào, ṙđav' Uhl 443, WRNE and SkR *nafel* '1) zlý, špatný, 2) "rare" nečistý', m/f 'd'ábel, nečistá síla' HŠŽ 186.

The solution is provided by the origins of the adjective *nasvalo* 'ill', which Ve 116 records w i t h o u t   a n   e t y m o l o g y, cf. also Pt II 323, Mi VIII 23, (T ∅) *na/ā sábala-* 'weak, powerless', OI *sábala-* 'strong, powerful' ← *sa-+bála-* n 'power' SRSl 462b. The Romani adjective has developed by analogy under the influence of the suffix *-(v)alo*. Without this influence the expected form of the word would be **nasvəl* with the oblique forms **nas∅vəle(s)* or **nasəv∅le(s)*, the latter becoming **nasule(s)* by the vocalization of the *-v-*. The meaning of Ancient Romani **nasul* has developed like that of **nasvalo*: the author of the illness, that is 'the Foul one' = **o Bišuǯo* m, τό πνεῦμα Ἀχάθαρτον n, and 'the Evil (spirit)', since "ἡ ἀσθένεια ἐχ τοῦ Πονηροῦ ἐστιν." Since the version *nafel* is unlikely to be independent of *nasul*, one is bound to think that its *-f-*, which is untypical of Romani, is a r a r e reflex of the **-sv-* cluster in the oblique case **nas∅vəle(s)*. In this light, the etymology of, for example, **fèdər* 'better' may also be reconsidered: it is not the reanalysis of T 9377 *bhadrám* 'well' into a comparative, but it may be, for example, the comparative of T 12473 *śivám* 'auspiciously (acc)', *śivēna* ins, Ḍ **sivēdər(o)* > AR **s∅vedər* > **fèdər*.

60 AR **niprəl* (**nipərd-* perf.) 'keeps mentioning'

It is from Mihály Rostás's dialect (WRNE) that I know the verb *liprèl* (*liperd-* perf.) 'keeps mentioning' (cf. Tálos 1988), non-pseudo-Wallachian Romani does not have it (Ve ∅), the expected form would be WRLo **imprèl*, cf. also BR *liparav* (also with initial *ri-, re-, le-*) '(s)pòmēnēm' Uhl 270, 352, elsewhere 'štùcām, I hiccup' (!) 380. (It is well-known that in Bosnian Romani the verbs in **-Crəl* became invariable *-C-arel* due to the influence of the suffix **-arəl*. WRNE *phabrèl* (*phaberd-* perfect) 'sets on fire', the equivalent of WRLo *phabārel* id., has developed by reverse analogy.)

The word is the equivalent of T 7214 *nípatati,* that is **ní-paṭati* I 'falls in ruin, settles down, falls at the feet of', f u r t h e r   m e a n i n g s of which are: '1) flies away, 2) falls out, upon, 3) meets, 4) c o m e s   t o   s b o d y ' s   m i n d ' (!) SRSl 329b.

61 AR **nizdral* 'trembles'

WR *izdràl* 'trembles, quivers', BR (=WR) *izdrav* 'dr̂htīm' Uhl 70, *izdrano* 'dr̂htav', BR (=AlbR) *lezdru* 'dr̂šćēm' Uhl 71 probably contained a **ni-* suffix earlier, but the conjectural transitional form, **inzdral* perhaps never existed:

```
*n  i   z   d   r   a   l
    ╳
*i  Ø   z   d   r   a   l
```

Perhaps it is of Indic origin and is doubly prefixed: in its *-z-* we may assume the trace of OI *saṃ-,* while its stem is partly T 6006 *trásati* I, T 6006.3 *trásyati* IV (with the participle T 6008 *trastá-*) 'trembles, is afraid'. There also exists *santrásati* with an inchoative meaning (SRSl 249b, 685b). Now its meaning is imperfective, therefore it has an *-a-*stem and as a result it has lost its sibilant. It may have got contaminated with some **-t/drắti* II or **-t/drāyati* (**t/drāṇá-* pp). The OI form could have been something like **niḥsaṃtrásyati.* (Ve 75 "Pers. *larzīdan*? Slav?")

62 AR **patr(i)adi* f/?m 'Easter'

Vekerdi records the word w i t h o u t   a n   e t y m o l o g y and as masculine (125), which is probably mistaken, HuR *patradi,* SnR *patrādja,* SkR *patradji* f 'velikonoce' HŠŽ 205b, BR *patradji(n),* *patragi* f 'Úskrs, Vàskrs' Uhl 411, 421, cf. Pt II 397, Mi VIII 35.

I know from Aleksej D. Belugin (personal communication, 1986) that this is a compound of T 7733 *páttra-* n or *pattrikā-* f 'leaf' and T 6333 *divasá-* m 'day', thus "Leaf(y)day", which is obviously a mirror translation of Hu *Virágvasárnap,* Blg *Цветница,* SC *Cvětna Nèdelja, Cvêti,* etc. all meaning 'Palm Sunday', literally 'flower Sunday'. Palm Sunday is the Sunday b e f o r e Easter. Since **kurko* m is not only 'Sunday' but also 'week', WR *bàrò kurkò* means both 'great week', i.e. 'Holy Week' (← Gk ἡ μεγάλη ἑβδομάς) and 'great Sunday', i.e. 'Easter Sunday', the name of which is up to the present day WR *bāro djēs* m literally 'great day' (← Blg *Великден*) or *bàri rātjì* 'great night', even *sùnto rātjì* f 'holy night' — the latter may also mean Christmas Eve. Actually, the meaning of WR *patrādjì* is literally 'Easter holidays' including Easter Monday, too. The " r i g h t w a r d   s h i f t " of the holiday may have been brought about by the time lag between the Julian and the

Gregorian calendars, which is 13 days, the latter being so much ahead. This explanation works only for the word for 'Easter' in the dialects of Central Europe. 'Cvĕtna Nèdelja' BR *luludjengo kurko* Uhl 48 itself is the equivalent of Gk ἡ Κυριαχὴ τῶν βαΐων. It can also be added that in the frayed suffix we may suspect either T 6328 *dína-* n 'day' or T 6331 *divá-* m id., and in Ancient Romani the word may indeed have been masculine with the oblique form **patr(i)adies/n*.

63 AR **peskərodo* 'alone'

Wallachian Romani has structures like *дéвушка больнá лежúт дóма*, adverbs may agree in gender: *i ralkjì nasvāljì pàšljol khērè* 'the girl lies at home ill'. WR *korkorò* (or *korkōrò*) 'alone' also behaves like this: *o sāvò korkorò sas* 'the boy was alone', *i šĕj korkorì gēlàstar kaj o orvòši* 'the girl went to the doctor alone', *mìndig korkorè sàmas* 'we were always alone'. However, there also occurs an invariant, gender insensitive adverb *korkorò* or *korkorì* (and also with stress on the first syllable) too, as well as *korkorès* with the suffix *-es* forming an adverb of adjectives. In Western dialects its first *-r-* may be missing, as in GR *kokres* (Ve 93, without an etymology), such *r*-less form can also be found in Uhlik (330).

I assume that the word contains a diminutive, thus SkR *korkororo* 'samotinky' HŠŽ 155 is doubly diminutive. In other languages of the area 'alone' is a diminutive form: Blg *самúчек*, Rum *singurèl,* but this is also so in Hungarian, though it is not really a Balkanic language: *egyedül* (*egy* 'one', *-(e)d* obsolete diminutive suffix). Its stem is not the numeral 'one', but the reflexive pronoun, *pes-* (the obl stem) → **pèskəro* 'his own ...', **pèskərodo* (diminutive possessive adjective) 'himself, alone'. It has probably lost its stem syllable by h a p l o l o g y in phrases like the following:

AR **nangiardiàs pes₁ pès₂ kərodo* 'he undressed himself/alone'

↘↓

*nangiardiàs pes₁ ∅₂ *kərodò* id.

The first *pes* is the reflexive pronoun. The resulting clipped word has been reanalysed as a reduplicated form: → *kərkodo.*

The following solution is less likely: in Balkanic languages (in Bulgarian, in Rumanian) the word for 'cuckoo' often expresses loneliness. If there existed an AR word **koko* m 'Cuculus' — and why not? — its diminutive was **kokodo* m, which would mean that the forms considered dissimilated above are primary.

64 AR **p(ə)rati* f 'band, belt'
Ve 136 "p r ā t i Mé: belt" w i t h o u t   a n   e t y m o l o g y .

Pt II 345 links it to Hu *párta* 'girl's headdress' and Hi *barhā* m, after the latter mentioning "Skr *waratra*" in parentheses. The latter, T 11320 *varatrā̆-* f, **varatrikā-* f '1) strap, 2) (saddle) girth (for elephants)' is a c o r r e c t etymology. There are two ways to hypothesize the prefix OI *pari-*, AR **pər-*. One is that it has radiated into AR **bərati* f id., yielding **pərati*. The other is to start out of OI **pari-varatrikā-* in the first instance, which became Ḍ **pərivərattī* > AR **pərvərati,* then by haplology **pərati.* The word has also suffered *-r-*dissimilation: the Ancient Romani descendant of word-internal *-t(t)r-* > Ḍ **-ttr-* is always **-t(ə)r-* except when preceded by another *r*, that is:

T 10702 *rātrikā-* f 'night' > Ḍ **rattᶖī* > AR **rati* f id.

T 11320 **-varatrikā-* f 'belt' > Ḍ **-vərattᶖī* > AR **-vərati* f.

65 AR **pəršuk* f '(bread)crumb(s)'

The Wallachian Romani word *prušùk* f is often used in the plural: *prušukà* plf '(bread)crumb(s)'. On the basis of BR (*=WR) *prušuk, brušuk, prušik* and *puršuk* f 'mṛva' Uhl 177 the archetype may have been **pəršuk.* Vekerdi 139 follows the etymological tradition in claiming it to be of an Armenian origin: "Arm *phšrankh*".

The word is a lexical a r c h a i s m , other Modern Indic languages have not preserved it, for example B /*pariśuṣka*/ *poriśuško* BRSl 541b is a direct loan from Sanskrit, a so-called "tatsama". The source of the word is thus OI (T Ø) *pariśuṣka-* '1) very dry, 2) drained, waterless' SRSl 380a. AR **pəršuk* is a r e g u l a r equivalent of this form, cf. OI *pari-* > AR **pər-* in other words as well, OI *śúṣka-,* **śúṣkaka-* (T 12548) SRSl 651b > AR **śuko* '1) dry, arid, parched, 2) thirsty', cf. also AR **ašukiarəl.*

66 AR **sapano* and **sapo* 'wet'

The Wallachian Romani dialects of Hungary do not know this word, it cannot be found in Uhlik's dictionary either, but GR *sapeno* 'wet' Ve 147 w i t h o u t a n e t y m o l o g y , SkR (with the note "reg") *sapano* 'vlhký' HŠŽ 243b.

As the adjective T 1340 *ārdrá-* 'wet' SRSl 99b (> AR **alo*) had a pair prefixed with *sa-,* T 1368 *sārdra-* id. SRSl 728, the adjective T 1208 *ā́pya-* 'wet, water-, aquatic' SRSl 95b could also have a variant (T Ø) **sā́pya-* id., which, or rather, the extended form **sā́pyaka-* of which became Ḍ **sappau-* > AR **sapo,* to which the pleonastic suffix *-ano* is added: **sapano.* OI *ā́pya-* is the derivate of *āpa-* n 'water' SRSl 94b.

67 AR **sovlia(l) xal* 'swears'

The origins of AR **sovəl* or **sovli* f 'oath' (Pt II 228, Mi VIII 67) are known: T 12290 *śapátha-* m, **śapathā-* f, SRSl 635b, cf. Hi *sā̃* f, Pb *sahũ* f id., moreover,

GR *sōvel* f, WR *s/colàx* f id., similarly *solax* in Sovakian Romani HŠŽ 249b 'přísaha (i manželská)'. Clearly these latter forms have shortened from **sōlax* or **sovlax,* but what is the ending *-ax* in them? A suffix?

The explanation is provided by the expression of 'taking an oath' in the languages of the Middle East and of India. In these areas the person taking the oath "e a t s" the o a t h : KKorm *sōnd xwārén* KRSl 338a, ModP *sowgänd xordän* PRSl II 71a, OsI *ārd (bā)xæryn* ROSl 212a, Pb *sahū khāṇā* RPbSl 309b, Hi *sū khānā* URSl 526b, Ne *kiriyā khānu* RNSl 211b, NRSl 217b, and so forth. Based on this, 1 suggested some while ago that the etymon of present-day WRLo *colaxārèl* 'swears' may be some expression like **sovlia xal* 'eats oaths', or **sovl(i)al xal* 'eats of the oath', and the initial *x-* of *xal* stuck to the end of the preceding word as the meaning of the expression became obscured. The assumption has proven right, in Uhlik under 'zàklinjēm se' 439 one can find besides *dav sovel* or *sovli* 'I give oath' the following: *solax xav* that is 'I e a t oath'. Besides BR *sovl(j)i, sovlax* f 'zákletva' Uhl 439, the variant *sovlal* f may also be found, this probably still merits the label "Abl(adv)" instead of "f", and perhaps there also exists the BR expression **sovlal xav.*

68 AR **suləm* m 'straw'

The word *sulum* m is a typical Wallachian Romani word, Vekerdi 154 *suluma* f appears to be a mistake, WRNE *sulma* is plural and suggests that originally this is a stem with vowel–zero alternation with *-ə-* in its second syllable, cf. also BR (?=WR) *sulum* m 'sláma' Uhl 342, *tjirvehko sulam* 'kûmova sláma' Uhl 148. Vekerdi's e t y m o l o g y : "Slav *slama* ?" is a m i s t a k e, which is not extenuated even by the "?".

This is also a lexical archaism (T ∅): OI *śumbala-* pln SRSl 651a 'straw, chaff', of which Ḍ **summəlo,* by metathesis **sulləmo,* then AR **suləm.*

69 AR **sumnəl* 'saint, holy'

The word is present in Bosnian Romani in the form: *sumnal, somnal(o)* 'svêt, svéti, svéta/o', HŠŽ 251 records it with an *: *sumnal* 'svĕt' (!), *sumnalu(t)no* 'svĕtový' (!), more on the meaning below. There is no data of the word in Hungary (Ve ∅), it is replaced in Wallachian Romani by *sùnto* (< Rum *sfînt* < Sl **svętъ.* The use of the word *sumnal* in the meaning 'world' is exemplified by the Skopje monthly paper entitled Романи Сумнал//Ромски Свет (since 1993), which is the result of the confusion of two homophonous Macedonian words: *ceem*[1] < **svĕtъ* 'world', *ceem*[2] < **svętъ* 'saint, holy'.

This is an Indic lexical archaism (T ∅), < OI *sunirmala-* 'very clear, absolutely transparent (water)' SRSl 737a, without the prefix *sú-* 'good, beautiful' (∼ Gk εὐ- and ὐ-): T 7366 *nirmala-* 'spotless, clean', of which Ḍ **sunimməlo,* then with the

metathesis of *-nimm- it became AR *sumnəl. The cause of the variant *sumnalo* is the infiltration of the suffix -alo, but the -a- of the variant *sumnal* still wants an explanation.

One possibility is that the -ə- > -a- change took place in order to get rid of the inconvenient alternation of the stem. The word *sumnəl belonged to the *nasvəl type, in which the internal -Cə- sequence changed place before a vowel suffix:

Nom	Obl		
*nasvəl	*nasəvl-e-	but later > WR nasul > C dialects nafel
*sumnəl	*sumən!l-e-	but later > BR sumnal

The variant with -a- ceases to be alternating, similarly to Hungarian, where -o/ö/e- is epenthetic, schwa-like:

	Nom	Acc	
	vacok	vack-ot	'den'
but	vacak	vacak-ot	'measly'

It could even be the case that the metathesis of *-nimm- happened later for the same reason:

*sun!məl   *sunəm!l-e-

In this both consonant clusters are uneasy, because -nm- should regularly become -mm-, while -ml- should become -mbl-, but an alternating *summəl :: *sunəmbl-e- would be too opaque. The case of *suləm (cf. above) could be cited here. Its oblique cases, *sulmes/n-, do not cause any problem, but a form like *suməl would require oblique cases of the form *sumbles/n-, which is again not transparent. One of the ways of getting rid of this problem is by metathesis — this is what had happened —, the other is the creation of *sumbəl nom — of which there is no evidence —, analogously to the following:

*sarəl	'remembers'	*sərdo pp
*bisrəl	'forgets'	*bisərdo pp
↓		↓
*bistrəl	id.	*bistərdo pp

It is interesting to note that this verb has again become opaque in some Wallachian Romani dialects acquiring the form *bristèl* :: *bisterdò*.

70 AR *šax m 'cabbage'

Similarly in most dialects, but in the Western dialects ModGk ἄρμα f 'brine' (cf. also BER I 15) has totally supplanted it (Ve 155 without indication of gender, but with a correct etymology): T 12370 śāka- m 'vegetable' (SRSl 640a). Elsewhere (1981, 410–411) he comments on this the following way: "It is remarkable that the names of cultivated plants and domesticated animals has in most cases developed by a narrowing of the meaning in Romani: 'vegetable > cabbage', 'young animal > swine', 'kid, lamb > sheep'. These changes in meaning make the assumption probable that these notions did not play a significant role in their everyday activity (e.g. they did not grow these and breed those) and they were indifferent to a more precise definition of these notions: any type of edible plant was simply 'vegetable'. It is possible that the differentiation of these notions took place only after leaving India (but in this case very early, still on Iranian territories, because the usage and meaning of these words is uniform in the different Romani languages)." Vekerdi is again trying to "reveal" the "primitiveness" of Romani thought, even though the semantic change at issue is well-known from a number of other languages, especially of the area concerned, cf.:

	Modern Greek	Bulgarian	Rumanian
'vegetable'	λαχανικά pln	зеленчук m	(L *viridia* pln)
'cabbage'	λάχανο n	зеле n	vàrză f

(cf. BER I 631, 633). After all the etymon of the word for 'cabbage' in Slavic languages is also L *composita* pln 'vegetable' (cf. Vasmer II 188).

71 AR *šelo m 'rope'

Vekerdi's etymology, "Skr śṛṅkhala ? /T 12544: sulva (ɔ: śulv/ba- m)" c a n n o t b e t a k e n s e r i o u s l y . The former could only yield something like *šingəl, the latter *šub(o).

Its etymon is obviously the following: T 13591 *sēlli-, of which Pb sēlhī f, Hi sēlī f, Ne seli, Gu selɔ m, all meaning '(thin) rope, string'. This would be the only example for the OI s > R š reflex, but more probably Turner's form correctly is *šēlli- or *sēlli-, since the languages from which the input was reconstructed all merge s(h)ibilants, while Romani does not.

72 AR *šol f 'whistle'

Its variants are: WR šol f 'whistle', SkR šol m 'písknutí', del šol 'písknout, zapískat', del šola 'pískat', šoljarel 'pískat, hvízdat' HŠŽ 257a, BR šo/il f 'zvížduk', šoljarav, šoldav, dav šo/il 'zvíždím' Uhl 460.

We are undoubtedly dealing with an onomatopoeic, expressive word, therefore it may even be considered as internally created. However, it may also be the case that it

is related to *šunəl* 'hears, listens to' < T 12598, 14289 *śṛṅóti//śṛṇuté* V, *śrūyáte* pass, *śrutá-* pp. Its source may be the feminine form of the participle T 12744 *śrutā-* f, or the well-known *śruti-* f, and because of the word's vowel we may also think of the adjective *śrautā-* f exhibiting the vṛddhi degree; all of which are connected to the notion of hearing and thus perhaps to that of sounding as well. This is still better than Vekerdi's etymology, which is "Pers. *šur* 'shout' ?". The Bengali equivalent of 'whistle' is worthy of notice: /*śõ*/ *śõ* BRSl 808a. The corroborative power of this fact is decreased by the dozens of other Bengali words meaning 'whistle' (BRSl passim).

73 AR **šuǯo* 'clear, pure'

The word for 'clear' is WR *(v)ūžo*, HuR *šūžo* (the etymology of Ve 159, 171 — "Skr. **śucya? *śudhya?*" — is correct after all, but he only mentions it in the entry *šūžo*), SkR *žužo* 'čistý' HŠŽ 313a, BR (=WR) *(v)užo* 'čìst' Uhl 53. The variants are explained by an original **šuǯo*:

(i)	0	š	u	ǯ	o
	1	š	u	ž	o

assimilation

(ii)	1	š	u	ž	o
	2	ž	u	ž	o

complete assimilation

(iii)	1	š	u	ž	o
	2	∅	u	ž	o
	3	v	u	ž	o

complete dissimilation

prothesis

For the variant *užo* Turner (T 2448) proposes OI *ṛjú-*, Pkr *ujju-*, which is a bad idea in any case, since it is obviously related to the others. 'Clear' in Old Indic is T 12520 *śuddhá-* adj/pp ← T 12524 *śúdhati/ē* I, T 12525 *śúdhyai/ē* IV, *śudhyáte* passive, √*śudh* SRSl 649b, 650a. Of this and of its causative two Ḍommānī verbs were derived: AR **šulavəl* 'sweeps, cleans' < Ḍ **śō/udhāvēdi* < T 12630 *śōdh(āp)ayati* and AR **šùǯiol* < Ḍ **śujjhədi*. The verb **šuǯiarəl* was derived of this and finally the adjective **šuǯo* was backformed from it.

WR *šudrò* 'cold', as well as HuR *šudro*, SkR *šudro* 'chladný' is primarily an attribute of *pājì* m 'water', the examples of Vekerdi and HŠŽ all relate to it (Ve 158, HŠŽ 258b), since its meaning is 'fresh' and it is probably the extension of T 12520

śuddhá-*-ra-. Vekerdi does not give an etymology and Turner does not know it. The 'weather, wind' and 'ice' are always connoted by *śilālò* 'cold, freezing', the etymology of which is known (T 12487).

74 AR **tromal* 'dare'

Its forms are: HuR *tromal*, WR *tromàl* 'dare' (Ve 167 without an etymology), SkR *tromal* '1) odvážit se, troufnout si, opovážit se, 2) "reg." smět, mít (ve významu smět)' HŠŽ 267a, BR *tromav/m*, (*=AlbR) *tromu* 'smêm, ùsūdīm se' Uhl 346, 414.

The word looks suspiciously Greek, but while ModGk τρομάζω 'I frighten' does not suit it, the meaning of τολμῶ (ÚMK 630a) is exactly the same. In Dhimotiki λ > ρ is frequent before a labial consonant, e.g. ἀδελφός > ἀδερφός, ἄλμη > ἄρμη, Ἀλβανίτης > Ἀρβανίτης > Turkish *Arnavut* (cf. BER I 15), etc. The *VR > RV* metathesis is common in Greek, e.g. AR **tetràdi* f 'Wednesday' < Gk Τετράδη > Τετάρτη f. I am certain that the source of *tromal* is a Greek dialectal *τρομῶ < *τορμῶ < τολμῶ.

75 AR **them* m 'country'

This is a widespread and well-known word. Vekerdi's etymology (63 and 1981, 415) is "Gk θέμα", Turner is on a better track: 13761 **sthāmya-*, since this is most probably an ancient word. The hypothesized **sthāmya-* and the AR **than* m 'place' < T 13753 *sthǎna-* n are both derivates of √*sthā* 'to stand'. I suppose that semantically 'country' has more to do with 'land' than with 'place'. To put it differently, AR **vlàxiko them* 'Wallachia, Βλαχία, Ţàra rumîneàscă' is the calque of Blg *Влашка зėм(л)я*. I assume that **them* has resulted of earlier **khem* by metathesis in the following way:

$$
\begin{bmatrix} +\text{grave} \\ -\text{anter} \end{bmatrix} \rightarrow \begin{bmatrix} -\text{grave} \\ +\text{anter} \end{bmatrix} \Big/ \underline{\quad} \mathrm{V} \begin{bmatrix} +\text{grave} \\ +\text{anter} \end{bmatrix}
$$

$$
kh \quad \rightarrow \quad th \quad \Big/ \underline{\quad} e \quad m
$$

Its etymon is OI (T ∅) *ksamya-* adj 'earthy' SRSl 180a > Ḍ **khemmo* (or **themmo* already), the *-e-* (instead of **-a-*) is explained by the following *-m(m)y-*. This is the type of consonantal metaphony also found in AR **men* f 'neck' < Ḍ **meñña* < T 9732 *mányā-* f. The masculine oblique ending can be explained similarly:

(i) AR **ḍomes* 'Rom-(acc)' < Ḍ **ḍommesso* < OI *ḍōmbasya* gen, cf. OI *ḍōm-bah* nom

(ii) AR **les* 'that, him (acc)' < Ḍ **desso* < OI *tásya* gen, cf. OI *sa(h)* m, *tat* n nom

In the latter the lenition *t- > d- > l-* is parallelled by English *þæt > that* [ðæt], in the AR **o* m definite article the result of lenition is ∅-. Still, Vekerdi (119) gives "Skr

asau ?" as a source b e s i d e s Gk ό, even though Turner 1928 treats the fate of *tásya* in Romani convincingly.

76 AR *thirəl* > *thìrdel* 'pulls, sucks (milk)'

Vekerdi 35–36 discusses the word, cf. also Pt II 290, Mi VIII 86, and considers ModGk σέρνω to be its source. For a Greek etymology it would be more reasonable to cite the verb σύρω (< AGk σύρω), which has the same meaning but is more adequate as for its forms.

The verbs for 'pull' and 'suck (milk)' are related in many languages. I used to think that WR *voj cìrdel ek thuvāljì* 'he smokes a cigarette' is the mirror translation of Rum *el tràge o ṭigàrǎ*. It is not impossible that 'sucks (milk)' was the p r i m a r y meaning of WR *cirdel,* which may then be connected to the following participle: OI T 6738 *dhītá-* pp ← *dháyati* 1, *dhīyátē* pass (SRSl 308b) '1) sucks (milk), 2) drinks', to which we have to assume the retroflexivization of *-t-,* which does sporadically occur (e.g. AR **pərəl* vi 'falls' < Ḍ **pəḍədi* < T 7722 *pátati* I, *patitá-* pp) and the infiltration of the participial form into the present stem, which is rather common. The initial AR **thi-,* like apparently always in all dialects, has changed to *ci-,* the *-rd-* cluster has simplified to *-ɸd-* in many W and C and in Albanian Romani dialects.

77 AR **thiro* m 'time'

GR *cīro* m 'time' has been known for a long time, see Pt II 200, Mi VII 34, Vekerdi (34) follows the etymological tradition in identifying ModGk χαιρός as its etymon, which, after all, is not impossible, but the R *ci-* ~ ModGk *ke-* correspondence is somewhat spurious. Considering that the consonant **c,* which is otherwise rather rare in Romani, is most common before **i,* before which on the other hand **th* does not occur, we may hypothesize a **/thi/* → **ci* variant rule or sound change in early Ancient Romani, supported by a number of etymologies, especially the interdialectal *tikno* ~ WR *ci(g)no* 'small'. T 5839 *tīkṣṇá-* > **tīkṣiṇa-* (SRSl 243a) > Ḍ **tikkhiṇau,* becoming (early) AR **tikhno.* In most Romani dialects this word underwent the simplification *-khn-* > *-kn-* before the leftward shift of aspiration. This failed to happen in Wallachian Romani, the emerging **thikno* thus became **cikno.* In German Romani the difference between nominals of the old and new declensions has ceased to exist, *cīro* could have earlier been finally stressed thematic, that is, not a loan.

Therefore at least two Indic etymons can be suggested for it: (i) T 6817 *dhīra-* '1) strong, 2) solid, 3) resolute, 4) p e r m a n e n t, u n c h a n g i n g, ... 7) s l o w' SRSl 306a, via Ḍ **dhīrau,* but the following is more probable: (ii) T 13771 *sthira-* '1) strong, hard, 2) p e r m a n e n t, firm, d u r a b l e, 3) obstinate, tough,

4) u n c h a n g i n g ...' SRSl 755a–b. Its connection to the notion of 'time' is more obvious in its derivates: *sthira-yauvana-* n 'eternal youth', etc. via some Ḍ **thirau.* Two more examples as semantic evidence: MP /y'wl/, Man /j'r/ *jār* 'time' < Ir **yāwa-war-,* Av *yav-* is 'durability, continuity' (RM 34); OI *lagná-* m/n 'convenient time' < *lagná-* pp ← *lágati* I '1) stays, resides, 2) owes sbody, 3) p a s s e s (t i m e)' (SRSl 551b).

78 AR **thivo* 'clever, able'

It is from J. D. Taikon's dialect (WRKld) that I know the following word: *c̓vo* (or *civo*) 'clever, able' Gjerdmann–Ljungberg 1963. Perhaps it was stressed finally originally and is of an Indic origin (see below).

In Wallachian Romani, when **kh* turns into the affricate *c̓* and **th* into c before **/i/* [i] ~ [i̭], they simultaneously also lose their aspiration, which is not really natural, unless the following steps are assumed:

(i) $h \rightarrow \acute{s} / k \underline{\quad} i$
     $\rightarrow s / t \underline{\quad} i$

(ii) $k \rightarrow \acute{t} / \underline{\quad} \acute{s}$

(iii) $\acute{t}\acute{s} \rightarrow \acute{c}$
      $ts \rightarrow c$

The fact that in many western dialects original *c̓h* loses its aspiration, while *ph, th* and *kh* retain it is in accordance with this segmentalized explanation. This is natural:

(i) $h \rightarrow \check{s} / \check{c} \underline{\quad}$

(ii) $\check{s} \rightarrow \emptyset / \check{c} \underline{\quad}$

or $\check{c}\check{s} \rightarrow \check{c}$

It is to be noted that the original Wallachian Romani triplet **c̓h* :: **c̓* :: **ǯ* is *ś* :: *c̓* :: *ź* in most dialects. If the palatalization of **kh* is early, then *c̓, t́ś* and *c̓h* were contrastive in early Wallachian Romani. It is also notable that **khli* yields unaspirated *(t́)t́* too, that is, *(t)tj* in a simpler notation.

Examples for the cases mentioned are as follows: WR *dic̓ol* 'seems' is not the direct continuation of an OI verb **dr̥śyate* IV/pass, but a late derivate of **dikhǝl* (< T 6507.2 **dēkṣati*) from AWR **dikhiol.* The relationship of WR *nǎc̓ol* 'passes (of time)' to the verb *nakhèl* 'runs' is similar—the latter is rare in Hungary, *nāšel* id. is used instead (< T 7087 *nāśyati* IV, *naṣṭá-* pp). Furthermore:

WR *mac̓* f 'fly' < AR **makhi* < D **makkhi* < T 9696 *mákṣikā-* f
WR *c̓il* m 'butter' < AR **khil* < Ḍ **ghido* < T 4501 *ghr̥ta-* n/pp

WR *či* '1) (do) not, 2) neither, nor' < Arm *k'i*, if it is not a continuation of T 14386 *cid*.

The change turning *th* into an affricate before *i* has taken place in other dialects as well: WR *ci(n)nò, cignò* 'small' and SkR *cikno* 'malý' HŠŽ 61a–b < *thikno* < *tikhno* < T 5839 *tīkṣinaka-* < *tīkṣná-*. Elsewhere the aspiration has been lost in the -*khn*- cluster before its shift leftwards: in other cases all dialects uniformly have *c*-. Thus *thirəl* later *cìrdel* 'pulls', *thiro* > *ciro* 'time', see above, and perhaps in the word discussed here as well, to which I propose the following source: T ∅ *dhīmant*- 'thoughtful' and 'clever, wise' SRSl 306a and T ∅ *dhìvan(t)*- 'skilled, competent, sensible, pious' SRSl ibid., either yields *thivo* via Ḍ *dhìvau*.

79 AR *uǔhai* f and *uǔhalin* f 'shadow'

Its forms are: WR *uśaljin* f 'shadow, cool place' Ve 170, BR (partly =WR) *uǔhalin, vušalin* f, *uǔhalipe* m, etc. 'sênka' Uhl 355. In Slovakian Romani the word *uǔhal, uǔhaljin* f 'stín' HŠŽ 271a, *uǔhaljiben* m 'chládek, stinné místo' HŠŽ 271b also has the variant *uǔhaj* f, which will turn out to be important for us. Vekerdi cites the stem T 763 *avachāda-* as its etymon, but Turner has other ideas too, in T 2542 he relates it to Ne *ojhel* 'shadow, half shade' NRSl 188b. This can in no way be correct, there is no natural explanation for the word-internal *-jjh-* becoming voiceless.

The simplest explanation happens to be the correct one: T 5027 *chāyằ-* f 'shadow' SRSl 215b (cf. ModP *sằye*, CIP *sāya*). The initial *u*- (< Ol *ava*- or *ud*-) is a suffix, the basic variant of the word did not have -*l*-, this is retained by Slovakian Romani. The form *uǔhalin* contains -*in* f, a suffix forming tree names. We may assume the following in Ancient Romani:

akhor m 'walnut' → *akhorin* f 'walnut tree' (HŠŽ 27a, Uhl 232)
khiliav f 'plum' → *khiliavin* f 'plum tree' (HŠŽ 160b, Uhl 378)
pəndəx f 'hazelnut' → *pəndəxin* f 'hazel' (HŠŽ 206, Uhl 154)
ambrol m/f 'pear' → *ambrolin* f 'pear tree' (HŠŽ 27b, Uhl 145)

The -*l*- of the latter has intruded in the following two, obviously because the otherwise resulting *... ai-in* would have created the nonexistent *ii* cluster, that is, hiatus:

phabai f 'apple' → *phabalin* f 'apple tree' (HŠŽ 223a–b, Uhl 116)
uǔhai f 'shadow' → *uǔhalin* f '*shadow casting tree, the shadow of the tree' (cf. Vekerdi above: 'cool place'). While *uśaljin* is 'shade', *ulùv* (see below) is '(artificial) shadow'.

80 AR *udit* m 'light'

The word 'light' is reported by Vekerdi (168 u d u d) as rare with the indication of Szt and Mü as sources, cf. also Pt II 310, Mi VII 46. Dialects of Wallachian Romani seem indeed not to have the word, I have only found it in the North Eastern,

pseudo-Wallachian Romani dialect: my informant, József Rézműves did not know the word itself, only thus (1977): *jakhèngi vudùt* (f) 'pupil (of the eye)'. No doubt this is the word discussed here. Uhl 454 BR *dud* m, *duti* m 'zĕnica', the Romani expression is the equivalent of Rum *lùmea òchiului* (or *ochilor*) and Hu *szeme fénye* (or *világa*). SkR *(v)udud* m HŠŽ 272a, 282b, '1) svítidlo z vydlabaného bramboru, ve kterém se pálí lůj, 2) "reg." svĕtlo', *ududoro* m 'svĕtélko, bludička', *udu(d)del* "reg." 'svítit'.

Its etymon may be found in Turner but without the Romani word: T 1994 *uddīptá*- n/pp ← *úddīpyatē* IV 'blazes up', as well as Pkr *udditta*- n 'flaming'. This might have been the Ḍommānī form as well: **udditto* then > AR **udit* m 'light', of which *udud* has developed by double assimilation. The form with -*t*- and the -*i*- of BR *duti* is the evidence for the original and expected **udit*. Vekerdi provides the word with the etymology "T 6606 Skr *dyuti* (?)", which is not correct since it cannot yield but something like **žul*.

81 AR **uluv* m 'shade'

It is a Wallachian Romani word, Vekerdi (169) quotes it from Horváth's glossary in the form u l u v 'shade' w i t h o u t indication of gender or a n e t y m o l o g y. The word is quite widely known, but because of its meaning it is difficult to elicit. DF's dictionary gives its meaning precisely but with a Hungarian provincialism: "szárnyék". In any case we know that while WR *uśaljìn* f is 'natural shadow, for example, of a tree', *ulùv* is 'atrificial shadow or its cause', as is obvious from Uhlik's dictionary: BR (?=*WR) *luv, uluv, vuluv* m 'hl'ădnjăk, záklon, zâštita' Uhl 97, 447.

It evidently shares its origin with the following Panjabi word: /*uhalā*/ pronounced **[ō/lā]*, with the spelling /*ōhalā*/ in RPbSl 235b, 681b–682a, 971b and 1042b, according to PbRSl 146a: '1) занавес, ширма, 2) укрытие, прикрытие'. It does not occur in Turner. Its etymon is obviously: OI *ullŏca*- m 'an awning, canopy' Monier Williams 219c, SRSl ∅.

82 AR **umblal* m 'a piece of glowing embers'

It is known in the form *umblàl* m in Wallachian Romani, but is not a well-known word (Ve ∅), WRNE (Mihály Rostás) *unglàl* m, SkR *umblal/v, unglal* m 'žhavý uhlík, oharek' HŠŽ 273a, BR (?=WR) *um(b)lal, umbral* m 'glávnja, ùgarak' Uhl 83, 400. It only accidentally resembles the verb **umblaval* 'hangs' and the Slavonic word **ǫgъlъ* m 'coal'.

For the notion of 'piece of glowing embers' there exist at least three words in Old Indic: 1) *alāta*- n, 2) *ulkắ*- f, 3) *úlmuka*- n SRSl 73a, 131b, 132a. I assume that the Romani word is a compound of 3) and 1) in this order: **ulm(uka)alāta*- > Ḍ **ummǝlādo* > AR **um(b)lal*. Its internal -*b*- is the same type of excrescent sound

as that of the verb *um(b)lavǝl: T 2230 avalamb(h)ayati, or rather *ullamb(h)ayati >
Ḍ *ummallāvēdi > *ummǝlāvǝdi > AR *umblavǝl, or that in the word *am(b)rol m/f
'pear', which is the borrowing of CIP amrūd. The following is obviously some-
how related to, but cannot directly be derived from, the word *ulmalāta-: T 2342
*umbāḍa- > Si umaṛu m 'lighted stick', umāṛī f 'half-burnt log, firebrand', Gu umāṛ,
umāṛɔ m, umāṛiyū n 'firebrand', ūbāṛiyū n 'piece of wood lighted at one end'.

83 AR *uǯ(i)le adv 'loan'

WR (v)u(n)źi/ulè 'loan', uźlipè m 'loan', BR (*=WR) uǯilipe m, uǯilimos m 'dûg,
dugòvanje, krèdīt, zájam' Uhl 73, 142, 439 is a rather enigmatic word, in Vekerdi
(169, 171) "Burushaski uš 'debt' ", which is a b s o l u t e l y i m p r o b a b l e, see also
Pt II 76, Mi VIII 92.

It is an original Indic word, T 1674 ujjháti VI, ujjhitá- pp '1) leaves, escapes,
2) sends back, g i v e s   u p, 3) r e n o u n c e s' SRSl 113a, we may hypothesize a
noun Ḍ *ujjhido m/n or a participle *ujjhidau. The word under discussion may be
the regular local adverb AR *uǯil(o) or the adverb of the participle *uǯilo.

84 AWR *zenia pl 'back(bone)'

From the form of WR zejà pl 'back' (according to Vekerdi "backside, hip" 177)
it seems evident that it is a word in plural, to which belongs a feminine singular base
form. Vekerdi's etymology is interesting because it is p a r t l y   c o r r e c t. He gives
Rum șàle plf 'the hip of humans, the back of animals' as a source, which is out of the
question p h o n e t i c a l l y. The Rumanian word, however, is the long variant of the
plural of șa f (șei pl) 'saddle, seat, dorsal bone' (from L sella).

This word is the m i r r o r   t r a n s l a t i o n of the Rumanian denomination. It
is nothing but the plural of zen f 'saddle', whose origin is well-known (Pt II 253,
Mi VIII 98 < CIP zēn, ModP zin). In Lovāri 'saddles' is zenjà today. The original
*zenia meaning 'back' has departed from the base form and became monomorphe-
matic, but at least the *n in it was not recoverable, therefore the typical denasalization
under Rumanian influence, WR -nj- > -j- became permanent. The Bosnian Romani
data are instructive, among them one can find zïn f 'gr̀ba', as well as zeja pl 'léđa,
gr̀baca, hr̀bat' — and, of course, zen f 'sèdlo' Uhl 91, 152, 333.

This semantic shift has interestingly taken the opposite direction is Persian:
MP /kwpk'/ kōfag 'saddle' is the derivate of /kwp/ kōf > CIP kūh > ModP kuh 'moun-
tain (≅ back)' (RM 69).

85 AWR *zǝlag f 'earring'

WR zlag f 'earring' is a typical dialectal word for the following word of other di-
alects: HuR čeni f, čenji f, SkR čen f 'náušnice', BR čen and čej f 'nàušnica' Uhl 191,

the latter can only be Wallachian Romani. The monosyllabic athematic variants are all backformations from the plural AR *čənia (HŠŽ 68b, Ve 38, 178, in both places without an etymology). Although we would expect word-initial *čh-, its etymology will be correct: T 5333 jhaṇikā- f, a word rhyming with kaṇikā- f and the like. The words that could be mentioned now mean 'earring', now 'drop'. Eng eardrop helps understanding the origins of the denomination. I assume that whatever it is derived from the strange WR zlag originally meant 'drop'. Starting out of the words connected to those meaning 'drop', I have arrived at the Wallachian Romani word for 'saliva'. It is sàlja f sg or saljà pl Ve 146, coming from Greek: σάλιο n. To the plural singulative forms *saljìn or *saljìk f 'drop of saliva' may also belong, similar words can be found in Slovakian Romani: saljikh, saljig 'trocha, trochu', saljigori 'trošičku' HŠŽ 243a. This takes us to one of the sources of HuR zalog '(a) little': SC zălogáj 'a bite (of food)', but this is suspiciously partly Greek, which has also στάλα f 'drop'. (Cf. also BR zalo/a, zaloga 'málko, málčice' Uhl 163.) Based on the few data available I claim, despite the missing links in the etymology, that the origins of zlag are to be searched in this direction.

86 AR *zəmavəl 'tries, tastes'

Wallachian Romani zumavel (Ve 178) w i t h o u t a n e t y m o l o g y. It is certainly connected to ClP āzmūdan, āzmāyīdan (āzmā(y)- pres) 'tries, attempts', as well as ModP âzmudän, âzmâyidän (âzmâ(y)- pres) id. PRSl 70, similarly in Fārsī Kābulī, but Afgh āzmūyəl id. seems to be a Persian loanword (RASl 578a).

The Ancient Romani form was probably *azmavəl originally, which after the clipping of the *a- considered a prefix automatically yielded *zəmavəl, while the ə labialized to u under the influence of the -m- (see the entry *xulai). Uhlik's (267, 297) datum is probably Wallachian Romani: zumavav 'iskùšävām, pòkušām, prôbām'.

87 AR *zəveli f '1) scrambled eggs, 2) crackling (?)'

Its occurrences: WR zevelji f 'crackling' Ve 177 w i t h o u t a n e t y m o l o g y, SkR ʒevelji f 'míchaná vajíčka' HŠŽ 99b, BR (?=WR) zevelin, zevel[j]i f 'čvárak' Uhl 55.

The word is usually derived from Armenian, provided that its original meaning is 'scrambled eggs', cf. Arm ʒvacel/t id. (which is a derivate of ModArm ʒu (ʒvi gen, in Grabar ʒu (ʒuoy gen) 'egg' (from a well-known stem), cf. Pokorny 784), Hübschmann 1895, 40, No. 179—it has found its way into Georgian, where it is ṭapamcvali id. KRGSl 821b.

I doubt this etymology because of phonological difficulties. It must be connected to the following Persian word: ModP zävâle (< ClP *zawāla) 'pastry shaped round (for baking bread)' PRSl 770a. That is, the earliest (interpolated) meaning of the

Romani word must have been '[egg]pie'. The Armenian word may be the cause of
the initial *ʒ-, but it may also be secondary.

Abbreviations of languages and dialects

Afgh	Afghan (i.e. Pashtu)
AGk	Ancient Greek
AlbR	Albanian Romani
Ar	Arabic
AR	(reconstructed) Ancient Romani
Arm	Armenian
Av	Avestan
AWR	(reconstructed) Ancient Wallachian Romani
B	Bengali
Bal	Baluchi
Blg	Bulgarian
BR	Bosnian Romani
C	the central dialect group of Romani
ClP	Classical Persian
Ḍ	Ḍommānī, the reconstructed early period of Ancient Romani, a kind of Prakrit
E	the eastern dialect group of Romani
Eng	English
FK	Fārsī Kābulī
Gk	Greek
GR	German Romani
Gra	Grabar (i.e. Classical Old Armenian)
Gu	Gujarati
Hi	Hindi
Hu	Hungarian
HuR	Hungarian Romani
IE	Indo-European
Ir	Iranian
It	Italian
Jud	Judío (i.e. spoken Judeo-Spanish)
Ka	Kannada
KKorm	Kurdish, Kurmanji dialect
Ko	Koya
L	Latin

Ma	Marathi
ModArm	Modern Armenian
ModGk	Modern Greek
ModP	Modern Persian
MP	Middle Persian
Ne	Nepali
OE	Old English
OF	Old French
OHG	Old High German
OI	Old Indic
Orm	Ormuri
Os	Ossetic
OsD	Ossetic, Digoron dialect (in the south)
OsI	Ossetic, Iron dialect (in the north)
Pa	Pali
Par	Parachi
Pb	Panjabi
PIE	Proto-Indo-European
Pkr	Prakrit
Port	Portuguese
Py	Parya
R	Romani (i.e. European Gypsy)
Ru	Russian
Rum	Rumanian
Ś	Śaurasēnī
SC	Serbo-Croatian
Si	Sindhi
Skr	Sanskrit
SkR	Slovakian Romani
Sl	Slavonic
SnR	Slovenian Romani
So	Sogdian
Span	Spanish
Ta	Tamil
Te	Telugu
W	the western dialect group of Romani
WeR	Welsh Romani
WR	Wallachian (or Vlach) Romani
WRKld	Wallachian Romani, Kelderāši dialect

WRLo   Wallachian Romani, Lovári dialect

WRNE   a pseudo-Wallachian dialect of Romani, spoken in the northeast
         of Hungary

References and the abbreviation of the works cited

Abaev, V.I. 1970. Russko-osetinskij slovar' (2nd ed.). Sovetskaja énciklopedija, Moskva. 583pp.

Aichele, W. 1957. Die ursprünglichen aspirierten Medien der Zigeunersprache. Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft 10.2, 101–2.

Andronov, M.S. et al. 1979. Kannada-russkij slovar'. Russkij jazyk, Moskva. 762pp.

Anufriev, F.F. 1979. Russko-pandžabskij slovar'. Russkij jazyk, Moskva. 1126pp.

ArRSl = Baranov 1962

Bakaev, Č.J. 1957. Kurdsko-russkij slovar'. GIINSl, Moskva. 618pp.

Baranov, H.K. 1962 Arabsko-russkij slovar'. GIINSl, Moskva. 1187pp.

Benkő L. (ed.) 1967–1976. A magyar nyelv történeti etimológiai szótára [A historical etymological dictionary of the Hungarian language]. Akadémiai Kiadó, Budapest. I: A–Gy 1142pp., II: H–Ó 1108pp., III: Ö–Zs 1230pp.

BER = Georgiev 1971/1979

Birjulev, S.V. et al. 1964. Urdu-russkij slovar'. Sovetskaja énciklopedija, Moskva. 890pp.

BRSl = Bykova et al. 1957

BSO(A)S = Bulletin of the School of Oriental (and African) Studies

Bykova, E.M. – Kolobkov, I.S. 1957. Kratkie svedenija po fonetike i slovoobrazovaniju bengal'skogo jazyka. Priloženie k bengal'skomu slovarju.

Bykova, E.M. et al. 1957. Bengal'sko-russkij slovar'. GIINSl, Moskva. 908pp.

CP = Turner 1957

Čunakova, O.M. (ed.) 1987. Kārnāmag-ī Ardašīr-ī Pābagān [Kniga dejanij Ardašira syna Papaka]. Pamjatniki pis'mennosti Vostoka LXXVIII. Nauka, Moskva. 163pp.

DF = Daróczy J. – Feyér L. 1984. Cigány–magyar kisszótár [Romani–Hungarian pocket dictionary]. TIT, Budapest. 114pp.

DRM = Kelemen 1964

DRSl = Kiseleva – Mikolajčik 1978

Efimov, V.A. 1965. Jazyk afganskih hazara. Jakaulangskij dialekt. Nauka, Jazyki narodov Azii i Afriki, Moskva. 98pp.

ÉSlSl = Trubačev 1974

Georgiev, V. et al. 1971/1979. Bylgarski etimologičen rečnik. Izdatelstvo na Bylgarskata Akademija na Naukite, Sofija. I: A–Z 672pp., II: I–Krepja 736pp.

Gjerdman, O. – Ljungberg, E. 1963. The language of the Swedish coppersmith Johan Dimitri Taikon. Falköping.

Hajdú P. 1969. A szamojéd konnektív-reciprok képző genetikai és arcális-tipológiai összefüggései [The genetic and areal–typological correlations of the Samoyedic connective-reciprocal suffix]. Nyelvtudományi Közlemények 71.1, 61–78.

Hajdú, P. 1975. Samojedologische Schriften. Studia Uralo-Altaica 6. Szeged. 145pp.

HM = Hutterer – Mészáros 1967

HŠŽ = Hübschmannová et al. 1991

Hübschmann, H. 1895. Armenische Grammatik. I. Die persischen un aramäischen Lehnwörter in Altarmenischen. Leipzig.

Hübschmannová, M. – Šebková, H. – Žigová, A. 1991. Romsko-český a česko-romský kapesní slovník [Romani–Czech and Czech–Romani pocket dictionary]. Státní pedagogické nakladatelství, Praha. 651pp.

Hutterer M. – Mészáros Gy. 1967. A lovári cigány dialektus leíró nyelvtana [A descriptive grammar of the Lovāri dialect of Romani]. MNyTK 117, Budapest. 95pp.

Jašari, H. 1985. Ano drom [On the road]. Drama. Prosveta, Gilana. 37pp.

KAP = Čunakova 1987

KaRSl = Andronov 1979

Kelemen, B. (ed.) 1964. Dicţionar romîn–maghiar I–II. Editura Academiei, s.l. 743+847pp.

Kiseleva, L.I. – Mikolajčik, V.I. 1978. Dari-russkij slovar'. IRJa, Moskva. 744pp.

Kočergina, V.A. 1978a. Sanskritsko-russkij slovar'. Russkij jazyk, Moskva. 896pp.

Kočergina, V.A. 1978b. Dopolnenie k Sanskritsko-russkomu slovarju. In: Sanskritsko-russkij slovar' izd. 2-e. Russkij jazyk, Moskva. 896–943.

Korolev, N.I. – Kudrjavcev, B.V. 1975. Russko-nepal'skij slovar'. Russkij jazyk, Moskva. 728pp.

KRGSl = Torotadze – Torotadze 1959

KRSl = Bakaev 1957

Lacková, E. 1992. Rómske rozprávky [Romani tales]. Východo-slovenské Vydavateľstvo, Košice. 95pp.

Lebedev, K.A. et al. 1973. Russko-afganskij slovar' (puštu). Sovetskaja ènciklopedija, Moskva. 872pp.

Lévi-Strauss, C. 1970, 1972. La champignon dans la culture. 1) L'homme 10.1, 5–16; 2) Antropologie structurale, II, 263–79, especially 270f.

LH = Livšic, V.A. – Hromov, A.L. 1981. Sogdijskij jazyk. In: OIJa. Sredneiranskie jazyki 347–514.

Maslov, Ju.S. 1981. Grammatika bolgarskogo jazyka. Vysšaja škola, Moskva. 408pp.

Mayrhofer, M. 1951. Handbuch des Pāli I–II. Carl Winter, Heidelberg.

Mi = Miklosich, F. X. 1872–1881. Beiträge zur Kenntnis der Zigeunermundarten, Über die Mundarten und die Wanderungen der Zigeuner Europas I–XII. Vienna.

Mohay A. 1988. Újgörög–magyar kéziszótár [Modern Greek–Hungarian concise dictionary]. Akadémiai Kiadó, Budapest. 773pp.

Monier-Williams, M. 1956. A Sanskrit–English dictionary. Oxford UP, Oxford.

Morgenstierne, G. 1929–1938. Indoiranian frontier languages I–II. Oslo.

MTESz = Benkő 1967–1976

NRSl = Rabinovič et al. 1968

OIJa = Rastorgueva et al. 1979

Oranskij, I.M. 1977. Fol'klor i jazyk gissarskih par'ja. (Srednjaja Azija). Vvedenie, teksty, slovar'. Nauka, Moskva. 448pp.

Oranskij, I.M. 1979. Vvedenie. OIJa. Drevneiranskie jazyki 10–128.

Osmanov, M.N. et al. 1985. Persidsko-russkij slovar' (3rd ed.). Russkij jazyk, Moskva. I: alef–že 800pp., II: sin–yå 864pp.

PbRSl = Rabinovič – Serebrjakov 1961

Pjatigorskij, A.M. – Rudin, S.G. 1960. Tamil'sko-russkij slovar'. GIINSl, Moskva. 1384pp.

Platts, J.T. 1930. A dictionary of Urdu, Classical Hindi and English (5th ed.).

Pokorny, J. 1947–1959. Indogermanisches etymologisches Wörterbuch. Bern.

Pott, A.F. 1844–1845. Die Zigeuner in Europa und Asien I–II, Halle.

PRSl = Osmanov et al. 1985

Pt = Pott 1844–1845

Rabinovič, N.S. – Serebrjakov, I.D. 1961. Pandžabsko-russkij slovar'. GIINSI. Moskva.

Rabinovič, N.S. et al. 1968. Nepal'sko-russkij slovar'. Sovetskaja ènciklopedija, Moskva. 1328pp.

Ralev, L. 1977. Govoryt na selo Vojnjagovo, Karlovsko. Bylgarska dialektologija. Kniga VIII, 3–199. Izdatelstvo na Bylgarskata Akademija na Naukite, Sofija.

RASl = Lebedev et al. 1973

Rastorgueva, V.S. et al. 1979–. Osnovy iranskogo jazykoznanija. Nauka, Moskva. I–V.

Rastorgueva, V.S. – Molčanova, E.K. 1981ab. Srednepersidskij jazyk. 6–146, Parfianskij jazyk. 147–232. OIJa. Sredneiranskie jazyki.

Ray, P. S. et al. 1966. Bengali language handbook. Center for Applied Linguistics, Washington, D.C. 137pp.

RM = Rastorgueva – Molčanova 1981ab

RNSl = Korolev – Kudrjavcev 1975

ROSl = Abaev 1970

RPbSl = Anufriev 1979

ShRsl = Tolstoj 1976

Šklifov, B. 1977. Rečnik na kosturskija govor. Bylgarska dialektologija. Kniga VIII, 201–327. Izdatelstvo na Bylgarskata Akademija na Naukite, Sofija.

Sowa, R. 1898. Wörterbuch des Dialekts der deutschen Zigeuner. Brockhaus, Leipzig. 128pp.

SRSl = Kočergina 1978a

SRSl D = Kočergina 1978b

T = Turner 1959a

Tálos, E. 1980–. Ḍōmbabhāṣayardhvanītihāsaḥ. Ms.

Tálos, E. 1988. The language of Mihály Rostás. In: Kovalcsik, K. (ed.): Mihály Rostás, a Gypsy storyteller. Hungarian Gypsy Studies 5. MTA Néprajzi Kutatócsoport, Budapest. 197–201.

Tálos, E. 1993. Máskép(p) (Kép szöveg nélkül II) [(An)OtherWise (Picture without caption II)]. In: Hajdú Péter 70 éves [Péter Hajdú is 70 years old]. Linguistica, Series A, Studia et Dissertationes 15, Budapest. 391–4.

TaRSl = Pjatigorskij – Rudin 1960

Tolstoj, I.I. 1976. Serbskohorvatsko-russkij slovar' (izd. 4.). IRJa, Moskva. 735pp.

Torotadze, A.G. – Torotadze, È.A. 1959. Kratkij russko-gruzinskij slovar'. GI Sabčota Sakartvelo, Tbilisi.

Trubačev, O.N. (ed.) 1974–. Ètimologičeskij slovar' slavjanskih jazykov. Praslavjanskij leksičeskih fond. Nauka, Moskva.

Turner, R. L. 1926a. The position of Romani in Indo-Aryan. Journal of the Gypsy Lore Society 3rd S V.4, 145–89. (= CP 251–90)

Turner, R. L. 1926b. Middle Indian -ḍ- and -ḍḍ-. In: Beiträge zur Wissenschaft und Geistesgeschichte Indiens. Festgabe Jacobi. Klopp. (= CP 239–50)

Turner, R. L. 1928. Romani les and Sanskrit tásya. BSOS 5.1, 43–51. (= CP 310–8)

Turner, R. L. 1957. Collected papers. Oxford UP, London. xvi+435pp.

Turner, R. L. 1959a. Comparative dictionary of the Indo-Aryan languages. London. Indexes: D.R. Turner.

Turner, R. L. 1959b. Transference of aspiration in European Gypsy. BSOAS 22.3, 491–8. (= CP 381–9)

Turner, R. L. 1970. Early shortening of geminates with compensatory lengthening in Indo-Aryan. BSOAS 33, 111–21.

Uhl = Uhlik, R. 1983. Srpskohrvatsko-romsko-engelski rečnik. Romengo Alavari. Svjetlost, Sarajevo. 473pp.

ÚMK = Mohay 1988

URSl = Birjulev et al. 1964

Vasmer, M. 1964. Ètimologičeskij slovar' russkogo jazyka I–IV. (In O.N. Trubačev's Russian translation.) Progress, Moskva.

Ve = Vekerdi 1983

Vekerdi J. 1974. A cigány népmese [The Romani folk tale]. Kőrösi Csoma Kiskönyvtár 13. Akadémiai Kiadó, Budapest. 280pp.

Vekerdi J. 1981. Nyelvészeti adalékok a cigányság őstörténetéhez [Linguistic contributions to the prehistory of the Roma]. Nyelvtudományi Közlemények 83.2, 409–21.

Vekerdi J. 1983. A magyarországi cigány nyelvjárások szótára [A dictionary of the Romani dialects of Hungary]. Pécs. 246pp.

Ventcel', T.V. – Čerenkov, L.M. 1976. Dialekty cyganskogo jazyka. In: N.I. Konrad (ed.): Indoevropejskie jazyki. Nauka, Moskva. 283–338.

Address of the author: Endre Tálos
                          Research Institute for Linguistics
                          Hungarian Academy of Sciences
                          P.O.Box 19
                          1250 Budapest
                          Hungary

ASPECTS OF LANGUAGE IDEOLOGY
IN A TRANSYLVANIAN VLACH GYPSY COMMUNITY*

KATALIN KOVALCSIK

Abstract

The article presents a Kelderash Vlach Gypsy song from Transylvania, which the performer improvised about his ideology of the local variant of the Romani language. In Vlach Gypsy communities, formal speech is differentiated from everyday informal speech. The frames of formal speech are the narratives including songs introduced by formulae of asking permission and greeting at communal events. In the paper, the form, content and music of the song are analyzed, together with some peculiarities of language ideology revealed in the song. In terms of dichotomies the singer separates the language and speakers of his community from other variants of Romani and from other languages and their speakers. The components of speech events have their constant epithets. These are "true", "pure" and "Romani". Language use also implies a value order, and the speech of Gypsy communities whose value system is identical with the singer's is considered "correct". At the same time, the "purely", that is, "correctly" performed narrative is the expression of the ritual purity of the community and the "correctness" of the given value order—the social–economic–cultural construction of Vlach Gypsies.

Popular ideologies of native speakers in different linguistic communities about the particular characteristics and functions of their vernacular have recently raised much interest among linguistic anthropologists. On the basis of Michael Silverstein and Alan Rumsey, Kathryn A. Woolard highlights aspects of language ideologies as follows: Linguistic ideologies are shared bodies of commonsense notions about the nature of language in the world. The notion of the linguistic ideology includes cultural conceptions not only of language and language variation, but of the nature and purpose of communication, and of communicative behaviour as an enactment of a collective order (Woolard 1992, 235). At the same time, relying on the findings of Alvin W. Gouldner and John B. Thompson, Woolard also points out that ideology is, among other things, a conscious public discourse, that part of consciousness which can be said. But in many other uses, the claim is not necessarily one of conscious, deliberate, or systematically organized thought (ibid. 238).

From among various ideologies, emphasis is to be laid on a particularly typical ideology of everyday thought, purism, which includes ideas about the purity of a language. In terms of language, linguistic purity means the "correct" use of the

vocabulary and various grammatical categories as compared to a certain norm which, in turn, is derived from a comparison with an unchanging, idealized state of the language. Taking a look at the informal and formal variants of language use also reveals that formal speech, which changes more slowly than the colloquial, is the main carrier of the linguistic traditions that generate the notion of correctness, that is, "purity" in the speakers. This aspect already belongs to the social function of language: the "correctly" used idioms are symbols of the "correctness" of social behaviour as accepted in the given community.

In the following, I attempt to examine some aspects of a purist folk ideology of the language as reflected by a Kelderash Vlach Gypsy song from Transylvania (Romania). In section 1, I am analyzing the peculiarities of formal speech in the communicative economy of the Vlach Gypsy communities. Section 2 carries the analysis of the song in terms of form, content and musical characteristics, while in section 3, I briefly summarize some of the specific features of the Vlach Gypsy linguistic ideology as it appears in the song.[1]

1. Daily speech–formal speech in the traditional Vlach Gypsy community

Research carried out in various Vlach Gypsy communities in Hungary have revealed that Vlach Gypsies differentiate two levels of language use. On the basis of his experiences gained in a Mashari Vlach Gypsy community, Michael Stewart has found that informal daily speech is called *duma*. Another type of speech is *vorba* that "takes on the special meaning of formalised speech preceded by greetings which indicate that the speech is for the benefit of all, not part of a private conversation. 'Be lucky!' (*T'aven baxtale!*) and 'I find you with God!' (*Devlesa rakhav tumen!*) are the most common forms. The content of *vorba* may be a tale, a joke or a riddle, but most often it is a song" (Stewart 1989, 85). Besides *vorba*, another term, that of *čači vorba* 'true speech' is also used to denote formalized speech and narrative genres. As against songs sung at *mulatšago*s, or community events (Hung. *mulatság* 'amusement'), at least three conventions are to be observed:

– they are to be in Romani (*romanes*);

– their style must meet the norms of the community, that is, they must be 'Gypsy' (*romano*);

– and their content must be 'true' (*čačo*), that is, revelatory of their real life (Stewart 1989, 86–7).

Vlach Gypsies differentiate two kinds of songs: slow, lyrical song (*loki djili*) and dance song (*khelimaski djili*). It is the former of the two that is called "speech" or "true speech". Extensive collections have revealed that this designation is used by the Vlach Gypsies all over Central Europe. The designation "speech" or "true speech" is uttered during community events. The person wishing to sing asks permission to take the floor. The formulae of taking leave may include: 'I beg permission from everyone, that you may forgive me if I offer you a true speech' (*Engedelmo mangav savora żenendar, ša jerton ma, te phenav tumenge jekh čači vorba*) (Stewart 1989, 85); or in a Lovari community in Slovakia: 'I beg permission, Gypsies, I have a little speech, let me tell you!' (*Engedelmo, romale, engedinen, si ma skurto vorba, phenav la tumenge!*) (Kovalcsik 1985, 46). The same designation may appear in the reply of the audience, for example in a Lovari community of Hungary: 'Tell your speech!' (*Phen tji vorba!*) (Bari 1996, 1/4), etc. The nature of the "truth" content of the songs depends on the context. Talking of the texts sung at a male gathering of the community, Stewart states: "The truth of these songs is stereotypical and general" (ibid. 91). Personal "truths" cannot be uttered on such occasions, since the aim of singing together is to express the brotherhood, the unity of Vlach Gypsy men. The leading singer picks from general textual elements referring to the entire community. These consist of pairs of six- or eight-syllabic lines that can be combined in various ways.

Personal "truths" are usually narrated in a narrow family circle. There is a shift of phase among various Vlach Gypsy communities in Hungary as to whether the personal "truth" is clad in newly invented text or in a slightly modified textual element commonly used in the community. Personal "truths" normally channel the performer's own problems, but this "truth" cannot sharply deviate from the norms of the community, and the song very often warns of the violation of collective norms. In a Lovari song I recorded in Voivodina (Yugoslavia) the male singer complained—to an embarrassed audience of his male relatives—that while he was in prison, his brothers neglected him, and did not visit him regularly. A woman in a north-eastern Tserhari-Churari community in Hungary designed the words of the song she was to sing at a family gathering to reconcile her sons (Kovalcsik 1993, 16–7). She calculated that uttering her message in a formalized language would produce the solemn atmosphere that, in turn, would force her sons to take their mother's advice and restore the unity of the family. In a greater part of such personal songs, when performed by men, there is a point close to the end where the singer declares that the content of his song is "true", and strengthens it with an oath.

It is noteworthy that there are several transitional stations between the two kinds of speech and types of truths which may be selected from according to a given set of parameters (e.g. gender, context, communicative intention, etc.). Irén

Kertész Wilkinson (1997) observed in another Mashari community in south-eastern Hungary not related to the one investigated by Stewart that the "truths" of a song are constantly commented on in prose, often in every strophe, and that this belongs to performing that song. On the other hand, Vlach Gypsies may perceive a song either as an emotionally distanced and general or as a closely personal message, this being further complicated by the fact that a performer may express his or her personal feelings or someone else's emotions. During a vigil, for example, the male singer's line "You too had a good mother" was corrected by a woman saying that his mother was still alive, though it was not she who had brought him up. When repeating the line, the singer already sang the particular line corrected ("You too have good mothers", ibid. 154).

Another group of formulae introducing or terminating a narrative reveals that the triple "truth" of language use, form, content, which characterizes a narrative presented in the right way is meant to epitomize the right social behaviour as the symbol of the ritual purity (see e.g. Sutherland 1975; Salo 1979; Okely 1983; Formoso 1988; Stewart 1997; Kertész Wilkinson 1997) of the community.[2] In male gatherings studied by Stewart women were not allowed to take part in singing at all (Stewart 1997, 186–7). In other communities, e.g. in the mentioned Tserhari-Churari group, women accompany their husbands' singing. Sometimes they can also take the lead, provided that their husband, brother or father has asked permission from the rest of the men. For example: 'I'm asking for permission, Gypsies, let us be pure and lucky, let my wife entertain us!' (*Engedelmo mangav, romale, t avas vuže taj baxtale, šaj kerel muŕi romnji amari voja!*) (Kovalcsik 1993, 16). Yet this community also observes events in which only men can be the actors: namely, storytelling. The ideal of the triple "truth" formulated on these occasions refers to the participants, the language and the genre. The participants must be "true" (Tserhari-Churari) Vlach Gypsies, their language must be Romani, for the "truth" can only be uttered in this tongue (Kovalcsik ibid. 4). The tale text related at a communal event must be "pure" in the sense that it contains no words or phrases alluding to sexuality. Should one occur in a text, the narrator has to resolve the taboo with the formula 'Forgive me, boys, let us be pure and lucky, that's how it is in the story!' (*Jertjinen, šavale, t avas vuže taj baxtale, ande paramičate-j!*) (Kovalcsik ibid. 16).

A part of the taboos relating to the content are undoubtedly linked up with the ideas of purity. Death, for example, is an impure event (Stewart 1994) therefore

---

[2] Vlach Gypsies conceive of the human body as divided into two parts by the waist: the upper, pure, and the lower, impure parts. Both parts can further be subdivided into internal and external pure and impure parts. There are strict rules to preserve purity, and only those who adhere to these rules can be lucky and healthy. The main source of impurity is believed to be female sexuality.

funeral songs are not allowed at the *mulatšago* of men. "If anyone sings songs of death and mourning it is typically the women (*romni*-s) at funerals" (Stewart 1989, 86). Illness is also a taboo, since it is an impure state which can also be caused by the violation of communical norms, that is, by some impure deed. A man who had fallen gravely ill in the mentioned Tserhari-Churari community composed a long song of the presumed causes of his illness. Fearing, however, that by making the illness public he would bring even greater disgrace upon his family he did not sing the song to anyone but recorded it on tape and listened to it for weeks, until he had psychologically elaborated the grave case (Kovalcsik 1991).

The following formulae of greetings may also appear at the beginning or end of lyrical songs: 'For your [fine] honour!' (*Anda tumari [šukar] patjiv!*), 'To your [fine] health!' (*Pe [šukare] sastjimas[te]!*), 'Be healthy and lucky!' (*T aven saste taj baxtale!*) (see examples in Kovalcsik 1985). Summarizing the attributes used so far, one finds that "Gypsy" and "true" refer to the genres, while "true", "pure", "lucky", "honourable" and "healthy" refer to the performers. The attributes are mutually interdependent, even if they are not both included in a formula of greeting. Presumably, when one condition is not fulfilled, then the narrative cannot be called "true speech" and the narrator and listeners are no longer "true Gypsies".

An interesting analogy is offered by the Chamula Indian community studied by Gossen. They regard their language, Tzotzil, the best of all the surrounding tongues or the "true language" (Gossen 1977, 86) and call the most formal genres of their narratives "pure" or "true speech" (Gossen ibid. 89). Although researches in Hungary so far have not found the adjective "pure" with reference to formal speech in the formulae, it can rightly be presumed that the "purity" of performers and that of the genre are mutually dependent, hence "purity" also applies to the three conventions determining the ways of speaking.

The clarification of the problem of purity is to be approached from the linguistic aspect by the analysis of a slow, lyrical song I recorded in a Kelderash Vlach Gypsy community in Transylvania—dedicated to the problem of *čači vorba*. It was improvised for me by the performer to show what he meant by the purity of his mother tongue, a local variant of Romani. Prior to that, the singer told me that non-Vlach Gypsies living in neighbouring villages did not speak as purely as they. "They speak broken Romani", he said. Wanting to pacify him, I set against his everyday opinion another everyday—relativistic—opinion prevalent in my culture, namely that their language is nice, too. My host pointed at the tape recorder and said: "C'mon, switch it on", and he began to sing.

On the one hand, the song is a metalinguistic product (Jakobson 1960), in that the singer resorts to the formal language to present his ideas about both types of—formal and informal—speech. On the other hand, it is a metanarrative (Babcock

1977), a "song about the song" since he characterizes the same type of formal speech, "Romani song" (*ʒili romani*) with the help of formal speech. The formal traits suggest that the personal message took the channel of formal speech because the performer deemed it generally valid to the entire community.

2. Analysis of the song

2.1. Form

In terms of form, the text conforms to the following criteria:

2.1.1. Formulae used in the song

Some formulae of formal speech are only used in prose, some only in singing, and again some in both. As for their place, there are some which have fixed places and some that can be shifted freely. Opening and closing formulae have fixed places. Among the Kelderash, asking for permission to sing, as the introduction to the linguistic event, and the good wishes at the end addressed to the audience are usually said in prose. Our performer regarded the circumstances of recording (only his wife being present) as a private context in which these formulae were not obligatory. That he presented both the opening and the closing formulae singing indicates that the highly important message of the text turned the context public for him.

Freely used formulae are oaths. Some types are part of the generally used textual elements, others are attached to personal, improvised texts. Since the statements of improvised texts are not necessarily valid generally, the singer ensures with the following formulae that the audience shall believe the "truth" of the statements: "May my head die", "May my children die" [i.e. if I lie, or, if I don't tell the truth]. These formulae are used in colloquial speech, when people want to stress something they are saying. In song texts, the formulae of swearing also serve as thematic transition, for the formula is followed by a new statement.

Addresses belonging to the interactive construction of slow lyrical songs constitute a separate type of formulae among the freely placed ones. The singer keeps in constant touch with the audience, addressing by the name or the designation of the status of the person(s) she or he is singing to.

In the studied text (Appendix 1) the opening formula was improvised, because the situation was extraordinary. The singer asked my husband for permission to sing to me. The reason is that Vlach Gypsy men may not sing to strange women except their immediate kin (mother, wife, sister). The singer reassured my husband that his intention was not courting me: "Don't fear, brother, don't fear, / We don't take it 'like that'". My husband did not have to have worries, for the singer's inten-

tions were honourable. Therefore, he could safely allow him to sing ("Leave it to me, brother"). To give weight to his honourable intentions, he ended the phase of asking for permission with an oath ("May my head die").

The closing formula of the song was directly addressed to me, but the good wishes forwarded to my children and husband also prove the honourable intention ("Be all your three children / And your man healthy").

Within the framework of interactive construction, the singer attached status names both to my husband and to me, so we became included among the Vlach Gypsy "brothers". It proves the gesture of admission that my husband was called *phrala* ('brother', vocative) and I was called *pheno* ('sister', vocative), at times replaced by the even more intimate or indulging term *dile* ('silly', vocative).

2.1.2. Rules of formal text-construction

The singer performs the text in a language used for Kelderash slow lyrical songs whose rules slightly deviate from daily speech. The detailed description of these rules is beyond the purview of this paper. The basic number of syllables in the Kelderash slow song is eight, with minor deviations. Shorter lines can be extended with meaningful (for, then, alas) or meaningless padding words or syllables, and the longer ones can be cut back by omitting certain less important words. Constant formal elements are the interjections, conjunctions and vocatives inserted before the lines, or at times after them, not belonging to the main text. At these places, one or two-syllabic extensions are also possible. In our case, the interjections *haj* ('ay') and *no* ('well'), the conjunctions *de* ('but') and *kë* ('for') and some additions to the content were inserted at the head of lines. In sung Kelderash, the text is arranged in rhyming lines, and the rhyming syllables (*-re*, *-u*, *-ju*, etc.) may change the phonological structure of the word it is attached to. (E.g. the word *djeh* 'day' may be sung as *djehu* or *djeju*.)

The studied song has six-line strophes throughout, except for one with seven lines. The song of sixteen strophes consists of six-line strophes throughout, except for one that has seven lines. In six strophes, only four different lines of text are sung to six musical lines, so lines one and three are repeated. In another six strophes, there are five lines of words. In five of these, the singer omitted the repetition of line three. In one case, the rule of extension was preserved, so the musical strophe was enlarged to contain seven lines. In three strophes, the lines of different text number six, so there are no repetitions. The last strophe is a half-strophe of three lines. It is typical of individual male singing that the last strophe ends earlier than the last musical line (here, as can be seen, on the third). (Perhaps, the singers thereby suggest that the termination is forced, they would still have much to say.)

2.2. Major elements of content

As it was described above, the first and last strophes of the song are reserved for asking permission and saying farewell, respectively. Strophe 2 carries the central statement, namely: that the language of the singer is 'pure Romani tongue' (*romani śib uži*), which is 'not broken' (*naj phadji*). In strophe 3, a long, three-line swearing formula is attached to underscore the "truth" of this statement. Strophes 4–15 contain the arguments supporting the same statement. They include the following motives:

(a) The carrier of the Romani language is the Romani song. Two examples are listed to illustrate that the type of speech of Romani song contains the "truths" formulated about the life and values of the Kelderash Vlach Gypsy community. The two stereotypical "truths" or values are:

the migrating way of life (strophes 4–5);

and the family of many children (strophe 7).

The first one, the itinerant way of living is due to the traditional occupation of the community. The singer and his relatives live by making, repairing and selling metal utensils. Their work and hawking activity is accepted in the Romanian shortage economy, for few would undertake such long absences from home touring villages far and near.

The second-named value is the child. The singer would always want children by his side, for them he would risk death. The two examples show that these are fragile values, as they are endangered. The neighbouring people mock them for their migrating way of life. The performer has no other means to take revenge on the taunters than cursing. So as to preserve the purity of formal speech, the singer utters an apologizing and well-wishing phrase before the curse-words ("Forgive me, be lucky"). (It is analogous with the formula that serves to ensure the purity of the audience of the Tserhari-Churari tales.)

A greater danger than derision is the threat of the outside world rending a Vlach Gypsy from his family. The mournful days (lit. black days) may refer to military service or imprisonment. The feeling of endangerment prompts the singer to state that he would sacrifice himself for his children, or he would not go on living without them.

After these two examples, the singer draws the conclusion: Man's fate of a "whole of life" is determined by God, and the songs are the reflection of this life. The Romani songs mirror the life of the community, carried by the Romani speech (strophes 8–9), which is not broken (strophes 9–10), he repeats.

(b) The presentation of the speakers of the pure language. This tongue is spoken by the *drotoša*, as the Kelderash are called by the Hungarians, for they repair metal vessels by wiring them (Hung. *drót* 'wire'). These tinkers are the "true" Gypsies as they sing in Romani "in the true manner" (strophes 11–12).

(c) Comparison of languages. The singer communicates his everyday experience that while you speak your mother tongue without difficulty, you "break your tongue twenty-five times" before you utter other languages (strophe 13).

(d) Aesthetic evaluation of the Romani language, and the singer's *ars poetica*. On the climax of the song, the singer illustrates with a metaphor the place of his own tongue among the languages he knows. Romani is the highest (as the rising Sun in the sky). That is why his aim is to sing purely in this language. The notion of purity also implies correct usage, the right content and the well-learnt form. The singer can be a fully-fledged member of the community if he has learnt the rules of "true speech" (strophe 14).

(e) The penultimate strophe (15) before saying farewell reinforces the existence of Romani as a separate language.

It is also part of the performance to wish to enhance the good reputation of the community. He mentions in three strophes (2, 11 and 12) that we must show the Gypsies in Budapest the recording which proves that we have visited "true Gypsies". Thus, the singer also accepted his "sister and brother" from a faraway place as mediators of his message.

2.3. Song

The tune (the model see in Appendix 2) is a peculiar variant of a style known as psalmodizing in the Hungarian folk music stock (Dobszay–Szendrei 1992). Though the style has many international ties, in the Hungarian language territory its use is localized to the Székely counties (Rom. Secuime) in Romania. This is supported by the fact that this Kelderash group had moved to their present whereabouts, the Mezőség (Rom. Cîmpia Ardealului), from the Székely region. The tune is however also known by Kelderash people living elsewhere in Transylvania. In Hungary, its four-line six-syllable variants can be found among both Romungros and Vlach Gypsies all over the country. The three- and four-line variants with eight-syllabic lines are known among the Romanian speaking Boyash Gypsies in Hungary, but they are prevalent among Boyashes in Voivodina (Yugoslavia). A representative of Boyash variants in popular art music is the anthem of the Hungarian Rom (composed by Gusztáv Varga, Kalyi Jag, 1989, B/9). Psalmodizing dance tunes are rarer in the repertory of Gypsy communities.[3]

---

[3] Some published variants of Gypsy tunes in psalmodizing style: Víg, 1984, A/1, six-syllabic Romungro tune. Kovalcsik 1988, 188, the Vlach Gypsy variant of the former. Kovalcsik 1994, 21, no. 2, four-line Boyash dance tune, ibid. 64–5, nos. 37–8, three-line remote variants. Kovalcsik 1988, 187, Boyash tune from Voivodina, ibid. 186, variant of the tune discussed in this paper by the same performer. Further variants: Bari 1996, 1/16, 2/4, 4/29. Ibid. 2/20 and 4/3 are dance tunes.

This is one of the most important tunes of the Transylvanian Kelderash. With a Hungarian word, they term it *keserves* (lit. plaintive song or lament), which is the central tune, among the collective events of the community, of the vigil sung to special words. Both the variability of the syllable scheme and the recitative character indicate a state somewhere between the free dirges and the set laments.

The middle cadence divides the tune into two part, the *b flat-c-d* range being extended to the *f-f'* octave and the upper *g'*. The arch declining from *g'* already appears in the first part of the first strophe (line 3), and in the second part begun unusually low, on *f*, it gets complemented with an arch rising high up. The first part of the second strophe remains in the *b flat-c-d* range, whereas the second part reaches the high register with a ninth leap, quite unexpectedly. This vertical change of focus between the two parts also prevails in the other strophes. If you take a look at the scheme of melody line variation, you will find that the singer conceives of the tune in pairs of tones. In the *b flat-c-d* range, he alternates the seconds of note pairs (see the list of the variations).[4] The large leaps may seem idiosyncrasies, but they are not: the principle of a new melodic construction is at work here. As against the former melodic ideal: the descending line from high, a new pattern of melody starting low and reaching the high ranges has been gaining ground in recent decades, resulting in the transformation of nearly the entire stock of Gypsy slow songs in Hungary. One of the first signs of this transformation can be discerned in the more archaic Kelderash songs.

3. Discussion and conclusion

From the above analysis, the following can be concluded as to the linguistic ideology of the community:

Firstly, the devices of representing the ideal of linguistic purity are dichotomy and the use of constant epithets. The singer mentions the following dichotomies concerning the language, the speech and the content:

– language (*śib*): pure–broken (*uži–phadji*)
– speech (*vorba*): Romani–others (*romani* – 'German' *njamcicko*, 'Jewish' *židovicko*, etc.)
– content: true–false ('I'm telling the truth' *čače phenav* – 'I am lying' *xoxavav*)

The pure–broken and Romani–others dichotomies apply to both the informal and the formal levels of language. The informal language is pure when it is spoken as

---

[4] The list only shows the more important variations without the complements of the pairs of tones.

the mother tongue, and 'broken' otherwise. The token of the purity of a formal language is its 'true' content. Romani is the language of both the informal and the formal spheres. Content only comes in when the formal sphere of Romani is concerned.

On the positive (Romani) side, the components of the speech event have constant epithets:

	true	Romani	pure
speakers	+	+	
language	+	+	+
speech	+	+	+
content	+		

– the speakers of Romani are: true Gypsies (*čače/adevër rom[a]*)
– the language is: true (*adevër*), Gypsy (*romani*) and pure (*uži*)
– the speech is: true (*čači*), Gypsy (*romani*) and pure ('[I speak] purely' *tistan*)
– the content is: true ('I tell the truth' *čače phenav*).

The speakers of the language are the Rom, the adjective Romani denotes both language and speech, while the attribute 'true' applies to all four categories. The adjective 'pure' only refers to the Romani language and speech.

Secondly, the singer regards a given linguistic situation as proper. Two of the attributes are expressed with words taken from two different languages. Besides Romani *čači* 'true', *adevër* 'idem' of Romanian origin (Rom. *adevăr[at]*) and besides *uži* 'pure', *tistan* of Hungarian origin (Hung. *tisztán* 'purely') are used as synonyms. Thus in the given linguistic situation, they are not averse to incorporating the lexicons of the two surrounding elaborated languages, Romanian and Hungarian, in the Romani language.

It is true that the pretext for singing the song was a conversation in which the singer commented, at least in my interpretation, on the informal use of the language by neighbouring Gypsy communities. Yet in the song he first addressed the theme of formal language use. Thus, our **third point** may be that, talking first about formal language means that for him, that is the 'true' language and the vehicle of the 'truths' important for the life of the community. For him, language use is primarily the conveyor of a value system. Those whose values are identical with his speak "pure, true Romani". Those whose values are different, speak badly, incorrectly, they simply speak a "broken" Romani language.

A **following point** in the correctly performed narrative is the expression and promotion of the community's ritual purity. The guarantee for the 'purity' of the

text is the realization of the ideal of 'pure' utterance devoid of any implications of sexuality. By turning us all into relatives for the duration of the singing who can enjoy the 'pure language' performed with a rich set of formulae, the singer ensured the ritual purity of all three of us. Returning now to the adjectives used in the greetings of performers and audience, one finds the attributes of ritually pure people. These people are 'true', 'pure', 'healthy, 'honourable' and 'lucky'.

Finally, let us try to answer the question why he conveyed his message through the Romani song. What is more, he even picked the most central tune of the community in view of the importance of the theme. Anthropological literature generally claims that the prejudice against the Gypsies derives from the difference between sedentary and nomadic ways of life, between productive and improductive activities (see e.g. Formoso 1988). Investigations in a few Gypsy communities, however, have shown that in actual fact ethic oppositions have arisen in which both parties regard their own ethic as correct. While peasants idealize productive work and the "toil" it implies, some Gypsy groups esteem a good business sense and its instrument, a glib tongue highest (see e.g. Stewart 1997, 237). At the same time, these groups live their lives as if in a constant state of siege, defending themselves with their egalitarian society against the outside world. It is via their formal behaviour at the communal events that they express their conviction which is the opposite of what the external world says: they are 'true' and genuine, pure and honourable.

They idealize the formal speech that carries these characteristics. As Stewart writes: "The Rom said that men became Rom in 'speech' (*vorba*). One answer to the question 'What does a true Gypsy do?/How can I become a Rom?' was '[if you learn] to speak in a refined way in company'. The answer to a common riddle 'When is a man happy?' was 'in *vorba*'. There was the straight meaning here, but perhaps a joke was also intended: All people, the Rom included, are always 'better' in speech than in practice. But by saying this, I do not wish to revive the tired distinction between the 'ideal' and the 'real', or—to use the sociological jargon— social structure and social action. Speech (including what we distinguish as song) had a significance for the Rom that it does not have for us since it was through speech that the Gypsies represented the reproduction of their society" (Stewart 1997, 202–3).

Romani song is the elevated form of speech by which it can be expressed that the mocked, disdained and endangered social–economic–cultural construction that provides the frame of Vlach Gypsy existence is correct by their standards. Anyone therefore who may question the unique purity of the speech of Vlach Gypsies living within these frames, expresses doubt about the correctness of this order.

References

Babcock, B.A. 1977. The story in the story: Metanarration in folk narrative. In: Bauman, R. (ed.): Verbal art as performance, 62–79. Waveland Press, Prospect Heights IL.

Bari, K. (collector) 1996. Anthology of Gypsy folk songs I–IV. Hungary–Rumania. EMI-QUINT Records Ltd.

Dobszay, L.–Szendrei, J. 1992. Catalogue of Hungarian folksong types. Arranged according to styles I. Institute for Musicology of the Hungarian Academy of Sciences, Budapest.

Gossen, G.H. 1977. Chamula genres of verbal behavior. In: Bauman, R. (ed.): Verbal art as performance, 81–115. Waveland Press, Prospect Heights IL.

Formoso, B. 1988. Tsiganes et Sedentaries: La reproduction culturelle d'une société. L'Harmattan, Paris.

Jacobson, R. 1960. Linguistics and poetics. In: Sebeok, T.A. (ed.): Style in language, 350–77. The MIT Press, Cambridge MA.

Kertész Wilkinson, I. 1997. The fair is ahead of me. Individual creativity and social contexts in the performances of a southern Hungarian Vlach Gypsy slow song. In: Gypsy folk music of Europe 4. Institute for Musicology of the Hungarian Academy for Sciences, Budapest.

Kalyi Jag, 1989. Lungoj o drom angla mande. Gipsy folk songs from Hungary. Hungaroton SLPX 18179.

Kovalcsik, K. 1985. Vlach Gypsy folk songs in Slovakia. In: Gypsy folk music in Europe I. Institute for Musicology of the Hungarian Academy for Sciences, Budapest.

Kovalcsik, K. 1988. On the historical folk music layers of the Gypsies in Hungary. In: Salo, M.T. (ed.): 100 years of Gypsy studies, 178–93. The Gypsy Lore Society.

Kovalcsik, K. 1991. Chansons tsiganes lentes sur l'expérience personnelle. In: Oralité Tzigane. Cahier de Littérature Orale 30: 45–64.

Kovalcsik, K. 1993. Men's and women's storytelling in a Hungarian Vlach Gypsy community. In: Journal of the Gypsy Lore Society 5, Vol. 3, No. 1: 1–20.

Kovalcsik, K. 1994. Florilyé dă primăvără 1–2. Tavaszi virágok 1–2. Beás cigány iskolai énekeskönyv. [Spring flowers 1–2. School song books for Boyash children.] Gandhi Középiskola, Pécs.

Okely, J. 1983. The traveller Gypsies. Cambridge University Press, Cambridge.

Salo, M.T. 1979. Gypsy ethnicity: Implications of native categories and interaction for ethnic classification. In: Ethnicity 6: 73–96.

Stewart, M. 1989. 'True speech': Song and the moral order of a Hungarian Vlach Gypsy community. In: Man (N.S.) 24: 79–102.

Stewart, M. 1994. Mauvaises morts, pretres impurs et le pouvoir récupérateur du chant, les rituels mortuaires chez les Tsiganes de Hongrie. In: Terrain 20: 21–36.

Stewart, M. 1997. The time of the Gypsies. Westview Press, Boulder CO.

Sutherland, A. 1975. Gypsies: The hidden Americans. Tavistock Publications, London.

Vig, R. 1984. Folk songs by Hungarian Gipsies. Hungaroton SLPX 18082.

Woolard, K.A. 1992. Language ideology: Issues and approaches. In: Pragmatics 2–3: 235–49.

Appendix 1

1. *kë*   /: Na dara, phrala, na na*ju*, :/   1. Don't fear brother, don't fear,
          Na lah kade, sa kadej*u*,   We don't take it "like that",
haj kë   /: Mukh la, phrala, pala ma*re*, :/   Leave it to me, brother,
          Te merel muŕo šoro*re*.   May my head die.

2. *kë*   /: Me ŕomane horbima*re*, :/   2. I speak Romani,
          Kade-j e ŕomani šib uži*re*,   That's the pure Romani tongue,
kade   /: Te phenen la vi khere*re*, :/   Tell them all at home,
          Kë mïŕi šib naj phadji*re*.   We don't speak a broken language.

3.   /: Ke mïŕi šib ŕomani-j*re*, :/   3. For my language is Romani,
          Te merel muŕo šoro*re*,   May my head die,
          /: Pheno, me te xoxavav*u*, :/   Sister, if I'm telling a lie,
          Pheno, me čęče te n av*u*.   Sister, if I'm not telling the truth.

4. *kë*   /: Dile, naj ma so kërav*u*, :/   4. Silly, I can't do any better
          Sajek phirav ol droma*ju*,   But roam the roads,
da   /: Kana phirav ol droma*re*, :/   And because I keep roaming the roads,
          Malan man ol phirijasa*re*.   I am the target of mockery.

5. *k*   /: Adevër ŕomane phenav, :/   5. I tell you in genuine Romani,
          Kake vorba ŕomani*re*,   For this is true Romani speech,
kë   /: Kade-j vorba ŕomani*re*, :/   That's what Romani is like:
          Te jertis, t ah baxtali*re*,   Forgive me, be lucky,
          Te dabulje la mama*re*.   "Give it to your mother."

6.   /: Dakë na čęče phenav*u*, :/   6. If I'm not telling the truth,
          Opral kade vorba ŕo[mani] na ma *jalu*,   Nothing is higher than [Romani],
kë   /: Numa d ek kocom phenav*u*, :/   Therefore, I only say,
          Pheno, me čęče phenav*u*.   Sister: I'm telling the truth.

7. *d*   /: Ane mïŕe kale djej*u*, :/   7. In my mournful days,
          Niči jekhar na na dikhljom*u*,   I wasn't allowed to see anyone,
          Numa sa le xurden*u*,   If only I'd been allowed to see my children,
          T ęvav lenge me merav*u*,   I would die for them,
          Merav lenca so *d* alaj*u*.   If it comes to that, I'll die with them.

8. *kë*	Żęneh, pheno, so *d* ala*re*,	8. You know, sister, how it will be,
	Żęneh, phrala, so *d* ala*re*,	You know, brother, how it will be,
	So kamel o lašo Del*re*,	What the Lord God grants me
k	Ane mïŕe saste zile*re*,	For the whole of my life.
	So ʒilabadjom tuke*re*,	What I sang to you
	Kade-j e ʒili ŕomani*re*.	Is the Romani song.

9.	Sar phenen l ane ʒili*re*,	9. As it is said in the song,
	Th ane lęte, joj, gëla*re*,	And what there is in the song,
	Kade-j vorba ŕomani*re*,	That's the Romani speech.
	Sa te merel muŕo šoro*re*,	May my head die,
	Pheno me te xoxavav*re*,	Sister, if I'm telling a lie,
	K amari šib but phadjol*re*.	Our language is not broken.

10.	I šib żanel te phagëla*re*,	10. You can speak a broken language,
	Sar tomnatu vaka *mere*,	As you can break ... in the autumn,
	Da kade-j adevër, joj*re*,	But this is true language,
	Haj te meren mïŕe šave*re*,	May my children die,
	Pheno, me te xoxavav*re*,	Sister, if I'm telling a lie,
	Kade-j amari šib maškar ame*re*.	We speak this language among ourselves.

11.	/: Maškar ol drotoša*re*, :/	11. We the tinkers,
	Tumende kade phenen*re*,	As you call us,
	Kadala-j o ŕom adevajre,	They are the true Rom,
	Save ʒiljaben čęče*re*,	Who sing in the true way,
	Ta dela vęste po Peśće*re*.	And news of them reach Budapest.

12. *no*	/: Kadala-j o ŕom adevër*u*, :/	12. They are the true Rom,
	Save ʒiljaben ŕomane*ju*,	Who sing in Romani.
	Anen tjo sasto sastjo*re*,	They wish you good health,
	Phen la *d* avri tu, pheno*re*,	Tell it to everyone, sister,
	De drom, te żal tjo kastatefono*re*.	Switch it on, let your tape-recorder work.

13. *the vi*	/: Kade-j vorba ŕomani*ju*, :/	13. This is the Romani speech,
	So na maj pharola khonig*re*,	It's not hard for anyone,
	/: Te kamo te horbima*ju*, :/	But if I want to speak (another language),
	Bišthępanżvar phagama*ju*.	My tongue breaks into it twenty-five times.

14.   /: Numa kocom kamlu, joj*re*, :/   14. All I want is
       Tistan ŕomane te ʒiljabav*re*,   To sing clearly in Romani,
       Kë de kathar avil o Kham te żal*u*,   For as the rising sun goes in the sky,
       Kade-j vorba ŕomani*re*,   Such is the Romani speech:
       S opral pe kade na maj naj*re*.   Nothing is higher than that.

15. *kë*   Kade-j vorba ŕomanij*re*,   15. For this is Romani speech,
       Kade-j naj či njamcicko, či   This is not German or Jewish
                   židovicko*re*,   language,
       Oda, pheno, ŕomani*re*,   This, sister, is the Romani language,
       /: Kë tu żęneh, so phënav*u*, :/   For you know what I am saying,
       Dile, me na xoxavav*u*.   Silly, I'm not lying.

16.   Sa de trin tje śęve*re*,   16. Be all your three children
       Sa vi tjo manuš bax-sastju*re*,   And your man healthy.
       Te trajil tja baxtali*re*!   Be your luck happy!

Appendix 2

The variations of the

1st line

a_1 b_1 c_1 d_1 = 1-2. 8. 13.
a_1 b_1 c_3 d_2 = 14. 16.
a_2 b_1 c_1 d_1 = 3-7. 9-10.
a_2 b_1 c_2 d_2 = 11.
a_3 b_2 c_1 d_1 = 12.
a_3 b_1 c_1 d_1 = 15.

2nd line

a_1 b_1 c_1 d_1 = 1-2. 4-7. 9. 12-14.
a_1 b_1 c_3 d_1 = 9. 15-16.
a_1 b_2 c_1 d_1 = 3.
a_1 b_2 c_2 d_2 = 8.
a_2 b_3 c_3 d_1 = 10.
a_3 b_2 c_1 d_1 = 11.

3rd line

a_1 b_1 c_1 d_1 = 1.
a_2 b_2 c_2 d_1 = 3-4. 6-8. 11-14.
a_2 b_1 c_1 d_1 = 5. 9-10. 15-16.

4th line

a_1 b_1 c_1 d_1 = 1.
a_1 b_1 c_2 d_2 = 2. 12-13.
a_1 b_1 c_2 d_3 = 14-15.
a_1 b_2 c_3 d_3 = 7.
a_2 b_2 c_2 d_2 = 4-6. 8-11.

5th line a_1 b_1 c_1 d_1 a_2 a_3 b_2 b_3 b_4 b_5 c_2

$a_1\ b_1\ c_1\ d_1 = 1.$
$a_1\ b_5\ c_2\ d_1 = 14.$
$a_2\ b_2\ c_2\ d_1 = 2\text{-}3.\ 12\text{-}13.$
$a_2\ b_3\ c_2\ d_1 = 4\text{-}6.\ 8\text{-}11.$
$a_3\ b_4\ c_1\ d_1 = 15.$
$a_3\ b_4\ c_2\ d_1 = 7.$

6th line a_1 b_1 c_1 d_1 a_2 b_2

$a_1\ b_1\ c_1\ d_1 = 1\text{-}4.\ [5.]\ 6\text{-}7.\ 9.\ 13\text{-}15.$
$a_2\ b_1\ c_1\ d_1 = 2\text{-}3.\ 11.$
$a_1\ b_2\ c_1\ d_1 = 8.\ 10.\ 12.$

The 6th line in the 5th strophe:

Address of the author: Katalin Kovalcsik
                            Institute for Musicology
                            Hungarian Academy of Sciences
                            H–1250 Budapest, P.O.Box 28
                            e-mail: Kovalcsik@zti.hu

TEASING IN THE LINGUISTIC SOCIALIZATION
OF GYPSY CHILDREN IN HUNGARY*

ZITA RÉGER

Abstract

Teasing has been demonstrated to be a potentially important, culture-specific means of of linguistic socialization in different linguistic communities and social groups. Structural and pragmatic characteristics of early teasing have been examined in a traditional Romani linguistic community in Hungary. Teasing proved to be widespread from the earliest age on in the speech addressed to Gypsy babies and children. Culturally specific features of early teasing and age-related changes in its patterns are analyzed. As to the acquisition of this discourse skill, Gypsy children seem to recognize and use very early some of the specific "contextualization cues" necessary for the idenfication of the underlying intention of teasing behind the surface form.

1. Introduction

Teasing has been demonstrated to be a potentially important, culture-specific means of linguistic socialization[1] in different linguistic communities and social groups (e.g., Heath 1983; Schieffelin 1986; Eisenberg 1986). The use of teasing, and variation in its structural and pragmatic characteristics is related to factors such as typical ways of achieving social control, beliefs about child rearing, and types of speech genres used in the community. Accordingly, teasing may serve a variety of pragmatic ends (e.g. social play and control of the behaviour of children). Teasing has also been found to be a source of language learning and affective socialization (Miller 1986).

According to earlier linguistic and anthropological research (Kaprow 1989; Réger–Gleason 1991) teasing is extremely widespread in traditional Gypsy communities, in child–adult interaction as well as among adults, and it is used in a variety of situations and genres. According to these data, teasing may be a source of amuse-

---

[1] Linguistic socialization is meant here as a term covering the processes by which children learn culturally and socially appropriate ways of speaking in their community (Ely – Gleason 1995).

ment, a means of scaffolding narratives (see the use of test questions in adult narratives described in Réger–Gleason 1991) or a technique of memorization in these nonliterate communities. It is also a means of verbal manipulation and social control in these basically egalitarian societies.[2] Furthermore, teasing is an extremely important communicative means in **inter-ethnic communication** as a way of manipulating the *gazhe* ('non-Gypsies') from a socially more vulnerable position, in order to reach various goals.

The aim of the present paper is to shed light on the use of these skills in Gypsy children's early linguistic and social development. More specifically, the paper will try to highlight some structural and pragmatic characteristics of teasing in the early interactions of children growing up in a traditional Gypsy community in Hungary: how adults tease young children, what children's earliest teasing skills look like, and what children may learn through teasing.

The analysis presented here is a preliminary one, and its primary goal is taxonomic and methodological. That is, it will focus mainly on identifying basic topics and discourse categories and their rough frequency of occurrence in early teasing among traditional Gypsies in Hungary.[3] (More detailed developmental analysis including further aspects, among others the examination of gender differences in early socialization to teasing, will be included in other papers.)

1.1. Some structural and pragmatic properties of teasing in earlier descriptions

As a multifaceted phenomenon showing substantial variation in its structural features and pragmatic ends, teasing has been differently defined in the available descriptions (Schieffelin 1986; Eisenberg 1986). Categories used in the analysis that follows will draw mostly on Eisenberg's analysis of the use of teasing in Mexican American families (Eisenberg 1986), as the nature of her data seemed to be closest to those presented in this paper.

---

[2] The importance of teasing in this respect is that it offers face saving possibilities and helps to maintain mutual respect. (That is, after a face threatening utterance one can always add: 'I have just been joking'.)

[3] Hungary's Gypsy population has recently been estimated at one half million (Havas – Kemény 1995). About one third of them are native speakers of one of the Romani dialects (Vekerdi 1977). Gypsies thus comprise the largest group of bilingual speakers in Hungary: their bilingualism is for the most part of the classic diglossic variety (see Fishman 1967) in which Romani is reserved for intra-group or "home" use, and Hungarian is the language used in all dealings with outsiders. Romani itself is an Indo-European language of Indian origin.

1.1.1. What is teasing and how does it "work"?

Eisenberg defined teasing as "any conversational sequence that opened with a mock challenge, insult, or threat" (Eisenberg 1986, 183–4). As a key feature of the teasing sequence she assumed that "the teaser did not intend the recipient to continue to believe the utterance was true, although he or she might intend the recipient to believe it initially" (Eisenberg 1986, 184). Thus, teasing sequences have been considered to be inherently ambiguous and intended to produce uncertainty: in each case, the addressee of the tease must decide whether the sequence produced by the speaker was serious or whether he/she was only joking.

However, the success of the tease may not depend only on the receiver's understanding. A teasing sequence opened by the speaker may move further successfully in two ways:

- The recipient immediately recognizes the tease and gives appropriate reply;
- The recipient fails to recognize the teasing intent, and thus becomes the "butt" of the tease.

In the first case, teasing "works" because all the participants play together. The second "works" because someone's vulnerability has been revealed. Thus, play and amusement could be shared with the receiver, or be at his/her expense (Eisenberg 1986, 186). "Contextualization cues" (Gumperz 1982) or "key" features (Hymes 1972) help the receiver in framing the utterance as teasing or play. These cues may be for example:

- discourse-related (e.g., the use of a disclaimer after the challenge, for example *I am beating you! No, I am not beating you*);
- suprasegmental (e.g., the use of sing-song intonation);
- paralinguistic (e.g., volume, way of delivery of the utterance); or
- nonverbal (smile/laugh/winks accompanying the teasing utterance).

Eisenberg has mentioned two further aspects of communication that help the child receiver to identify teasing:

- Repetitiveness: in child–adult discourse, teasing sequences often become routinized. This provides a frame for interpreting the challenge as non-serious.
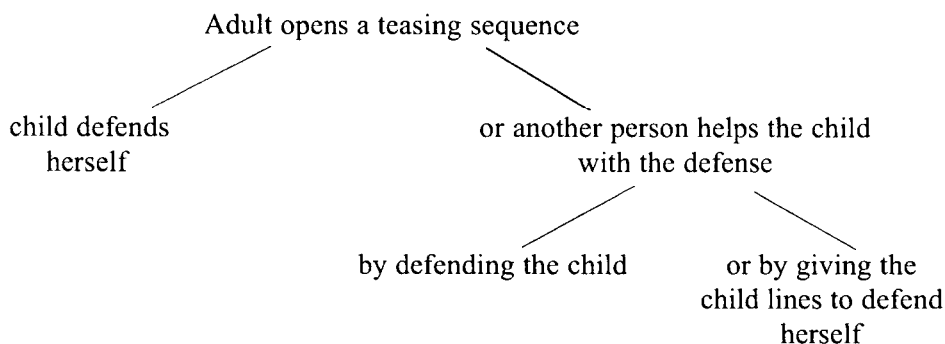- Safe context: the teaser is well known to the child.

Repetitiveness and routinization of teasing result in restriction of topic choice. Some prototypical mock threats and challenges identified in teasing young children in different cultures are for example:

– inflicting bodily harm (*I am going to beat you*);
– disrupting important relationships (*I am going to take your mum away*);
– withholding affection (*I don't like you*);
– attacking valued attribute (*You are ugly!*), etc.

(Eisenberg 1986; Miller 1986)

1.1.2. Some typical discourse strategies used in teasing

There are a number of typical discourse strategies found in adult–child teasing sequences. Some of the most frequent are the following:

Adult opens a teasing sequence

child defends   or another person helps the child
herself   with the defense

by defending the child   or by giving the
child lines to defend
herself

(Eisenberg 1986)

Other forms of teasing worth mentioning are:

– issuing a teasing statement addressed to a third party **about** the child (with the child expected to overhear, e.g., *She is ugly, isn't she?*); or
– inviting the child to tease another person together with the speaker.

Thus, pragmatic and discourse properties of teasing as well as related grammatical structures show rather great variability. Teasing may be:

– direct or indirect;
– dyadic, triadic (with a co-teaser) or multiparty;
– teasing may have different grammatical forms (declarative or imperative sentences, questions of different types) with different illocutionary force (see for example the use of rhetorical questions with the illocutionary force of shaming among the Kaluli: *Is it yours?* meaning: 'It isn't yours' in Schieffelin 1986).

2. Teasing in adult–child interaction in traditional Gypsy communities

2.1. Methods

2.1.1. Subjects and data

The present paper is based on longitudinal naturalistic data on two girls living in the same community. The community belongs to the group speaking the Mashari subdialect of the Vlach dialect.[4] One of the girls, Pitjinka has been audio-recorded on various occasions between 1 month and 3 years, the other, nicknamed Bici has been videotaped from 1;0 to 6;0, at regular intervals, in an ongoing project. Two half-hour records have been selected for more detailed analysis here. The first one (recorded when Pitjinka was one month old) will allow illustration of some culturally specific ways of teasing preverbal babies. The second one (a recording made with Bici at 2 years 1 month of age) will highlight some adult strategies in teasing children at the early phases of language development as well as the young child's earliest attempts to use this culturally important genre of speech.

Recordings were transcribed in CHAT format, and all instances of teasing were coded and analyzed using the CHILDES system (MacWhinney 1994).

2.1.2. The definition of teasing used in this paper

For the purposes of the present paper, Eisenberg's definition (see above) has been expanded in two respects:

– In addition to mock challenge, insult or threat, any other type of social play involving children that contained ambiguity and served as a potential source of joking and amusement has been included in this analysis.

– In addition to sequences of mock challenge, insult or threat, this analysis also included any single utterance representing some of these categories. (Thus, for example, teasing names or initiations of teasing that have been subsequently dropped or remained unanswered have also been included in the analysis.)

2.2. Results and discussion

In the two half-hour records analyzed, 60.1 and 38.0% of the total number of utterances, respectively, belonged to this category. (Record 1 contained 520 and Record 2 781 utterances addressed to the child, produced overwhelmingly by the respective

---

[4] The Mashari subdialect of the Vlach dialect of Romani, is characterized among others by the following features:
– at the end of the syllable *s* is replaced by *h*;
– some old Vlach Romani sounds related to Romanian ones (open *e* [ę] and velar *i* [ï]) may be found;
– *k* may be replaced by *g* in certain lexical items, e.g., *kodo* becomes *gado* (see Stewart 1987, 187).

mothers. In parallel Hungarian data, teasing occurred only occasionally, see Réger 1986.) Thus, teasing seemed to be pervasive in the speech addressed to Gypsy babies and children from the earliest age on. (Observational data also indicated that it was practised by everyone in the extended family and neighbourhood surrounding these children.) Preverbal children seemed to be even more frequently teased than older ones.

Items of teasing in the two samples have been analysed concerning topics and discourse strategies used by the respective mothers, and the basic speech act types identifiable within the teasing sequences. The levels of analysis proposed here can partially be matched to those proposed by the most elaborated analytical system in developmental pragmatics, the Ninio–Wheeler model (Ninio–Snow 1996). In terms of this model, teasing is to be considered as the type of interchange selected for study, and "basic speech acts" serve to categorize segments of speech at the utterance level (see Ninio–Snow 1996, 40).

Results indicated important qualitative and quantitative change in teasing, occurring as a function of the child's age, growing communicative ability, and the age-related transformation of her culturally defined attributes (see e.g. footnote 10).

2.2.1. Teasing a preverbal baby

As to the first topic, teasing preverbal children, the most characteristic topics and discourse strategies may be illustrated by the following longer sequence from Record 1.[5]

---

[5] Record 1 was made by Judit Szegő during her stay in the community while involved in a project of social anthropology. Record 2 was made by the author.

The text was segmented according to the principles of segmentation recommended in Fletcher and Garman (1988).

In the transcription of text, I used the following notation:

Aspirated sounds are denoted by an *h* after the respective consonant: *th kh ph.*

Palatal sounds are denoted by a *j* after the sound: *tj dj nj lj.*

Alveopalatal sounds: *č š ž*

Palatalized sibilants: *ś ź*

The /r̄/ sound is a trill, usually uvular or long *rr.*

x is a uvular fricative.

Velar *e* and *i* are denoted as *ë* and *ï.*

Stress is indicated by a grave accent.

Length is indicated by a bar on the respective vowel.

A short pause is denoted by /; a longer pause is shown as //.

Hungarian words in the text are underlined.

EXAMPLE 1

Cica, mother, 23 years old
Kincho, brother, 5 years old
Pitjinka, daughter, one month old

Mother:

1   Sòste roves de aś ta te dikhàv?
     Why are you crying, let me see!

2   Sò kerden, hã!
     What have you done, hmm?

3   Soj?
     What?

4   Čò kerden <u>édeske</u>?
     What have you done my dear?

5   Kon?
     Who?

6   Jãj! (kisses the baby)
     Oh!

7   Sopij āri tiře miźè!
     I'll suck out your pussy!

8   Mittõj?
     All right?

9   Cãrel e mama tji mìź!
     Mum is licking out your pussy!

     (the baby is crying)
10   Kòn mardẹh tu?
     Who has beaten you?

11   Mištoj aba,
     All right,

12   Mittoj de na rov!
     All right, don't cry!

13   Arak / de!
     Wait a little!

14   Aś ta te dikhàv la!
     Wait, so I can see it!

15 Aś ta!
    Wait a little!

16 Jàj, dè dulmut dikhlem tu, <u>édeske</u>!
    I haven't seen you for ages, sweetie!

17 Kàj sanas?
    Where have you been?

18 Hā!
    Hmm?

19 Kāj phīres mange, he?
    Where have you been walking, hmm?

20 De!
    Hmm!

21 Phèn abā!
    Tell me!

22 Phen aba kaj sanah mange <u>édeske</u>!
    Tell me where have you been, honey!

23 Phèn aba!
    Tell me!

    (addressing Kincho)
24 Kāj sah tji phej, Kinčo?
    Where has your sister been, Kincho?

    (giving lines Kincho to repeat)
25 "Sanas te r̄odes tuke r̄omès?"
    "Have you been looking for a man for yourself?"

26 Ha!
    Hmm!

27 Phèn aba lake!
    Tell her already!

28 "Sanas te r̄odeh r̄omès?"
    "Have you been looking for a man for yourself?"

29 Ha?
    Hmm?

(speaking to Pitjinka again)

30   De phèn hāt!
      Tell me!

31   Phen ke mindjār māràv tu!
      Tell me because I am going to beat you!

32   De, gyorsan!
      Promptly!

33   De!
      Why?

34   Cì pheneh?
      Won't you tell me?

35   Aś tà!
      Wait a little,

36   Atta, att- att- att(a)!
      Wait, wait, wait, wait!

37   Aś ta te maràv tut!
      Wait a little, I am beating you!

38   Aś tà!
      Wait!

      (the baby starts crying)

39   Mittoj, nem!
      All right, no!

40   Či mārel tu e mama!
      Your mother doesn't beat you!

41   Či mārel tut!
      She doesn't beat you!

42   Mm!
      Hmm!

43   "Jaj de bokhāli sim mange édesanya!
      "I am so hungry, mother!

44   De m aba k čerra te xav!
      Give me a little to eat!

45   De bokhāli čim mange!
      I am so hungry!

46   De m aba k čerra te xǎv!
      Give me a little to eat!

47   Me nàdon bokhāli sim!"
      I am very hungry!"

48   Nàdjon bokhāli, udje <u>anyu</u>?
      Very hungry, aren't you, mother?

49   Hā!
      Hmm!

50   (offering her breast)
      // Ehe dik so dav tu!
      Look what I am giving to you!

51   Le!
      Take it!

Topics and discourse strategies characteristic in the earliest form of teasing among traditional Gypsies are almost all represented in the passage above.

Topics of teasing in Record 1

The main topics of teasing in the record were the following:

— Sexual teasing of the type demonstrated by lines 7 and 9 in Example 1. Further ways of sexual teasing in Record 1 were calling the baby *kurva* 'whore', *büdös kurva* 'damned whore' or *inci-punci* 'teeny pussy', and the corresponding teasing statements and questions (e.g., *Jaj kurva j la!* 'Oh, she is a whore'; *Kāj la goda kurva Pitjinka?* 'Where is that Pitjinka whore?'). These items were also included in this category. Sexual teasing including these types represented almost one quarter (24.3%) of the teasing utterances addressed to one-month old Pitjinka by her mother and brother;

— Mock threat and mock scolding. Mock threats of inflicting bodily harm (see lines 31–37), mock threats and mock scolding realized through specific intonation patterns, as well as other types of scolding (for example, attributing negative qualities to the baby, e.g., *Jaj, dik de bużengli la!* 'Oh, look how she is cunning!') were included into this category. They represented 12.1% of the utterances related to teasing in Record 1;

— and the uniquely Gypsy type of teasing called in this paper "Scenes of future life" where the mother and older brother address mock-scolding, challenges, or threats to baby Pitjinka in her imagined form of a young girl to be married (see lines 16–28 in Example 1. In some cases this category was overlapping with that of sex-

ual teasing and mock threat. These items were double-coded.) This "future-oriented" way of teasing was represented by almost the half (44.9%!) of the teasing utterances in Record 1 (see also Fig. 1 for the relative proportion of these categories).
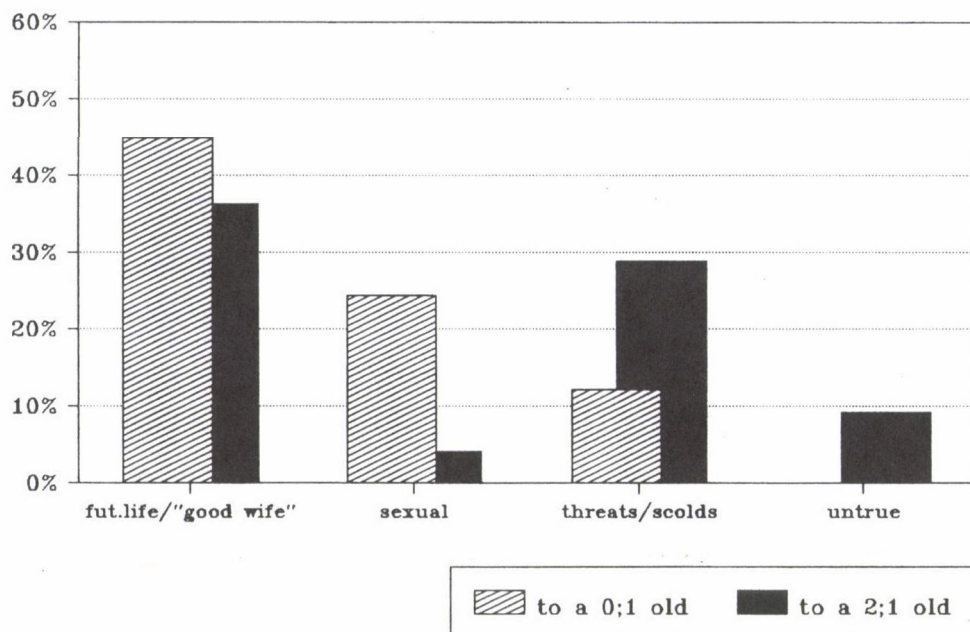

Fig. 1
Topics of teasing in Record 1 and Record 2

In analyzing the **discourse strategies** observed in Records 1 and 2, teasing was first classified as either direct or indirect. In addition, on the basis of the grammatical form and illocutionary force of utterances constituting teasing sequences, basic speech acts were identified in both direct and indirect teasing.

Discourse strategies in Record 1: speech to a preverbal baby

Direct and indirect teasing

In direct teasing, the utterance was addressed directly to the target child (in 2nd or 3rd person singular), and to her only. In indirect teasing, the teaser involved another person (or persons) into the tease.
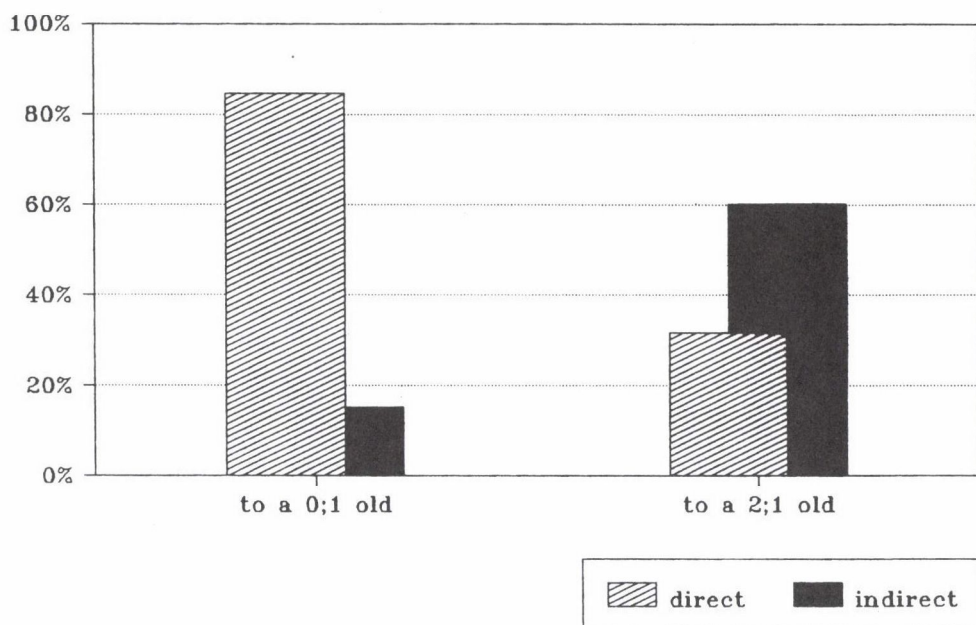
Fig. 2
Discourse strategies in teasing: direct and indirect

Indirect teasing types used in the teasing sequences of Record 1 were the following:

– Inviting a co-teaser (an older participant, see lines 24–28 in Example 1 where the young mother involved 5-year-old brother Kincho as co-teaser, by giving him lines to repeat to baby sister Pitjinka). In this case, "speaker(s) to hearer alignment" (Goffman 1981) may schematically be represented as **Adult and Other > Child.**

– Improvising mock dialogue with shifts between the roles of speaker and addressee (see for example lines 37–48 in Example 1. Here the mother addressed mock scolding to Pitjinka, and after several disclaimers, she shifted to the baby's role and started to speak on her behalf). The discourse strategy of 'filling in' both speakers' roles may be called indirect in a special sense. Namely, in such cases, teasing is performed through modelling the appropriate interaction itself to the target child as audience. (Schematically, this could be represented in the following way: **Adult and "Child" > Child.)**
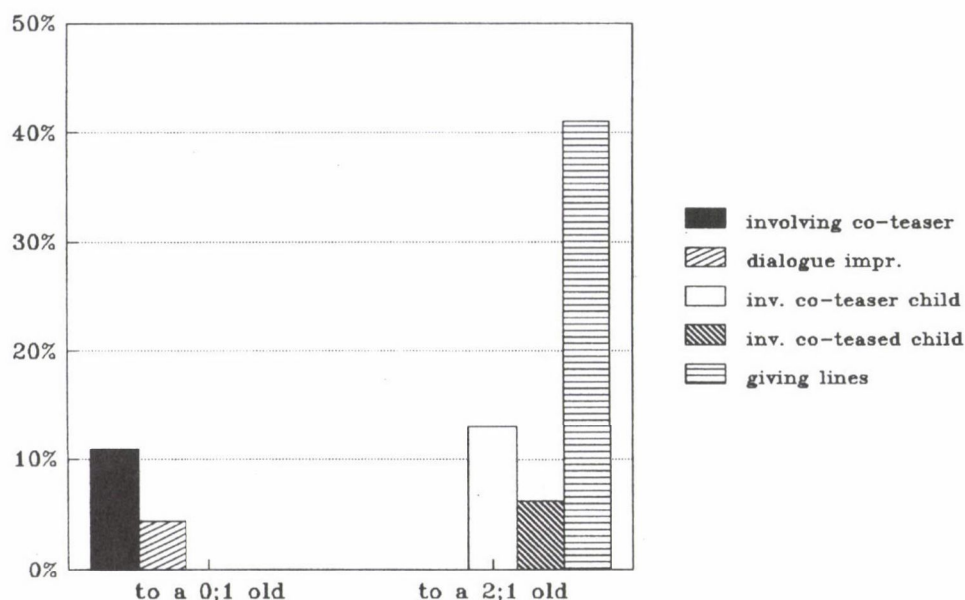
Fig. 3
Types of indirect teasing in Record 1 and Record 2

Basic speech acts

Some of the most important speech act categories were the following:

– Directives (direct and indirect), mostly aimed at eliciting some sort of response (nonverbal, vocalization) from the baby (see for example lines 12, 21–23, 30–31, and 34 in Example 1);

– Teasing formulated as statements (see for example lines 7 and 9). (Emphatic statements formulated as exclamatory sentences were also included in this category.)

– Naming/calling routines (see the examples of sexual teasing cited above on p. 298);

– Explicit disclaimers after threats (formulated as statements) (see lines 39–41 in Example 1 above).

– Mock information requests (illustrated by lines 17 and 19 in Example 1).

Frequency of direct and indirect teasing, and basic speech act categories occurring in the two records are given in Figures 2 and 4. They clearly show that the baby at this early age was mostly directly teased, with directives and statements as dominating speech act types. Figure 3 shows that when using indirect discourse strategies, the mother of one-month old Pitjinka preferred to involve a co-teaser, and constructed mock dialogues less often.
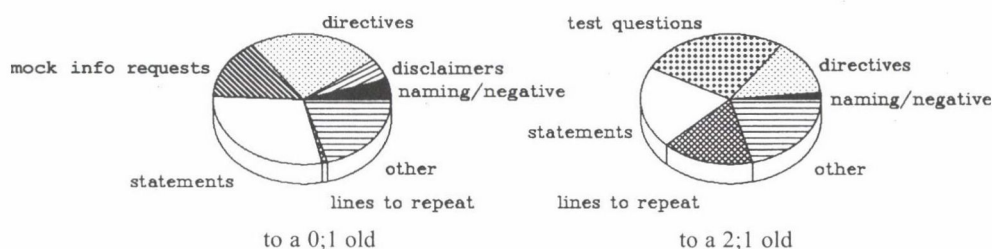
Fig. 4
Basic speech acts within teasing sequences

Even this preliminary analysis may shed light on the highly sophisticated nature of the early forms of teasing in traditional Gypsy communities. Among others, this complexity is highlighted by the fact that teasing in this record involved at least four types of shifts:

– shifts in topic
– shifts in time (between present and—expected—future)
– shifts in roles (speaker – addressee)
– and shifts in alignment of speaker(s) to hearer (see Goffman 1981)

In addition to characteristic topics and discourse strategies, contextualization cues such as sing-song intonation and way of delivery (e.g., an extremely affectionate tone of voice) as well as nonverbal cues unequivocally indicated to the audience the teasing force of these utterances.

2.2.2. Teasing a toddler

Examination of the types of teasing in Record 2 has shown a partially different range of topics and discourse strategies used with older children at the early phases of language development. Teasing at this age has already been strongly routinized and particular teasing sequences frequently repeated seemed to serve as building blocks in adult–child interaction.

Topics of teasing in Record 2

Examples 2 and 3 demonstrate some of the most typical topics of teasing addressed to a girl of toddler age: making her speak and act like a loving and caring wife. (These routines are called "good wife routines" in this paper.) This category of teasing in which the child was called on to act out gender-specific behaviour expected from girls at a later age seems to be the equivalent for this age of the earlier form of teasing "scenes of future life" performed with infants. In addition, Example 2 also illustrates

the widespread habit of Gypsy adults of linking the young child with a mate of the opposite sex by declaring them "lovers" (*piramnji* and *piramno*). This social play allows Gypsy adults and older children to perform a kind of "double teasing" (i.e. teasing both of the children at the same time or getting one child teased by the other).

EXAMPLE 2

> Maron, mother, 37 years old
> Bici, daughter, 2;1 years old, "girlfriend"
> Dombi, playmate, 2;1 years old, "boyfriend"
> Dragica, sister, 12 years old

1 Mother: Jāj de, "figyelem" phen!
               Oh, "attention", say!

2 Child: (trying to imitate) [se:da:m]!

3 Mother: Phen, "figyelem"!
               Say: "attention"!

4 Child: Dombi! (name of playmate and "mock-boyfriend")

5 Mother: "Av khĕrè", phèn!
               "Come home!", say!

6 Child: "Av khĕrè!"
               "Come home!"

7 Mother: "Av xà"!
               "Come, eat!"

8 Child: (trying to repeat) [xakka]!

9 Mother: So?
               What?

10   Mas.
               Meat.

11 Child: Miż.
               Pussy.

(laughing: the mother playfully hits Bici)

12 Mother: Muri śej!
               My daughter!

13   Astār ta lehki kor̃!
      Grab him (Dombi) by the neck!

14   Astár lehki kor̃!
      Grab him (Dombi) by the neck!

15   Astàr lehki kor̃!
      Grab him (Dombi) by the neck!

16 Sister: Šukārès!
             Gently!

17 Child: Ihi / [jo:]!
             Oh, well!

18 Mother: Phen, "Mātò san, mò"?
             Say, "Are you drunk, boy"?

19 Child: Māsò san?
             Are you drunk?

20 Mother: De čumide- sār čumideh e Dombes?
             Well, kiss- how are you kissing Dombi?

21   Sikàv ta sar kamèh les!
             Show me, how you love him!

22 Child: (with playful intonation)
             Dombi māto / māto / mato!
             Dombi drunk / drunk / drunk!

23 Mother: Phen "Àv čak kathe Dombi pàša ma!
             Say, "Come here, Dombi, by my side!

24   Beš tēle"!
             Sit down!"

25 Child: (looking at the video camera)
             Mamẹ, mame čok ko(do)?
             Mum, mum, what is this?

26 Mother: O repülōvo.
             The airplane.

27 Child: (trying to imitate) [pe:je].

28 Mother: Pe god avèl tjo dad!
             It's with that your father is coming.

Mock-display of emotions and prototypical (gender appropriate) teasing utterances performed by the child seemed to be an inexhaustible source of amusement for the audience. "Good wife routines" organized in this frame were the most frequent topic in Record 2 (representing more than one third of the total number of teasing utterances, see Fig. 1 on p. 299).

The second group of teasing (about one quarter of the utterances coded as teasing in Record 2) included mock threats, mock insults and mock scolding. This type of teasing occurred more than twice as frequently as in Record 1 (see Fig. 1).

A further type of teasing (illustrated by lines 26 and 28 in Example 2) consisted in making the child believe facts that were evidently untrue. In such cases subsequent amusement was at the expense of the child. Teasing through misleading utterances occurred in 9.1% of the total number of utterances related to teasing in Record 2.

Sexual teasing, a pervasive topic in Record 1 occurred only occasionally in Record 2 (see Fig. 1), but the proficient use of this type of teasing by the child has given evidence of its earlier importance in child–adult interaction (see line 11 in Example 2). Sexual teasing included also the same type of naming/calling and "qualifying" routines as those found in Record 1 (e.g., *Kurv avesa, Bici!* 'You are going to be a whore!').

Discourse strategies in Record 2

Direct and indirect teasing

With respect to discourse types, one of the major changes in Record 2 when compared to Record 1 was the significant increase in indirect teasing (60.2% in Record 2 as compared to 15.2% in Record 1, see Fig. 2 on p. 300). The explanation for this change may be that the growing linguistic and social competence of the child allowed a greater variation of types of shifts in speaker(s) to hearer(s) roles and alignment. There were three major types of indirect teasing, all three representing different types of speaker(s) to hearer(s) alignment:

– Involving the child as co-teaser in teasing a third person in the family or the community who becomes the "butt" of the tease. This latter may be sometimes present and expected to overhear, or absent at the time of the interaction. In such cases, the child becomes aligned with the speaker. (Schematically, the direction of the tease may be represented as **Adult and Child > Other**.) (*Kòn i māto?* 'Who is drunk?'; *Kòn i basoso?* 'Who is a fucker?')

– Involving a co-teased child: teasing the child and, simultaneously, someone else too, who was expected to overhear (most often the peer of the opposite sex). In this case, the young child became aligned with the hearer, and the direction of the tease may be represented schematically as **Adult > Child and Other**. (For

example, *Gede bārì t avla, me či dav le e Norbeske* 'When she will grow up, I won't give her to Norbi [the "boyfriend's" nickname]'.)

– The third type may be classified as a complex one where the child is both the receiver and (re)producer of the tease. (That is, she becomes aligned with both the speaker's and the hearer's role. This may schematically be represented as **Adult > Child > Other**.) In this type of teasing, the adult generally gives the child lines to repeat (typically those of some version of the "good wife's routines"). These instances were mostly introduced by the imperative verb form *Phen* "Say!" (see for example lines 3–7, 18 and 23 in Example 2). According to Fig. 3, it was this latter type of indirect teasing strategy that the mother of 2;1 year old Bici preferred above all: this category alone made up 41.0% of the total number of teasing utterances in Record 2 (see also Fig. 3 on p. 301 for the relative proportion of these teasing categories).

Basic speech acts in Record 2

Basic speech acts in adult teasing (both direct and indirect) at this stage were the following:

– Directives (direct or indirect) addressed to the child. They generally aimed at eliciting gender appropriate adult-like speech, action, or emotional display on the part of the child (see for example lines 13–15 in Example 2). Direct modelling of teasing utterances by giving the child lines to repeat, was considered a special subgroup within this category. Directives, including direct modelling, made up 30.4% of teasing utterances in Record 2. (See also Fig. 4 on p. 302, where the proportion of "giving lines" and other directives are represented separately.)

– Test questions[6] focusing on the details of "future life" (see for example lines 3, 5, 12 in Example 3), or involving the child in teasing a third person (see above, p. 305). Test questions of this type seemed to be highly routinized and repetitive. They represented about one quarter of the teasing utterances in Record 2 (see Fig. 4).

– Statements used mostly for describing the child's future life, acknowledging her answers within the teasing routines, or qualifying either the girl or her "boyfriend") amounted to 20.4% of the teasing utterances in Record 2 (see Fig. 4).

– Teasing naming/qualifying (*Dile!* 'Fool!') or the corresponding negatively qualifying statements (*Tiszta diñ la!* 'She is completely crazy!') occurred less frequently than in Record 1: it made up only 1.6% of teasing utterances in Record 2 (see Fig. 4).

---

[6] Test questions are that type of questions to which the speaker knows the answer and expects the addressee to give this particular answer (or verbal response).

EXAMPLE 3: TEST QUESTION IN TEASING

Maron, mother, 37 years old
Bici, daughter, "girlfriend" 2;1 years old
Dombi, playmate, "boyfriend" 2;1 years old
Dragica, sister, 12 years old

1	Mother:	Či kameh leh aba?
		Don't you love him [i.e. Dombi, the "boyfriend"] any more?

2	Child:	Na.
		I don't.

3	Mother:	Kàh kameh, muri śej?
		Whom do you love, my daughter?

4	Child:	E mamàs.
		Mum.

5	Mother:	H anke?
		And anybody else?

6	Child:	E Dombès.
		Dombi.

7	Mother:	De akkor čàk kameh e Dombes!
		Well, then you do love Dombi!

8		Lèh leh tuke?
		Are you going to marry him?

9		Sodè śavoṙa avna tu kathar o Dombi?
		How many children are you going to have from Dombi?

10		Śej!
		Daughter!

11		Bici!

12		Sodè śavora avna tu?
		How many children are you going to have?

		(addressing Dombi)
13		Ker aba vì tu vãreso!
		Say something already, you too!

Table 1 summarizes the discourse strategies used in Records 1 and 2, in a comparative way.

Table 1
Main discourse strategies in teasing

RECORD 1: TEASING AN INFANT	RECORD 2 : TEASING A TODDLER
DIRECT TEASING	
naming/addressing	naming/addressing
negatively qualifying statements/questions	negatively qualifying statements/questions
disclaimers	—
directives (direct or indirect) (aimed at eliciting nonverbal response)	directives (direct or indirect) (aimed at eliciting emotional display)
—	test questions (aimed at eliciting expected utterance)
	giving the child lines to repeat
INDIRECT (OR INDIRECT + DIRECT) TEASING AND ITS SCHEMATIC REPRESENTATION	
involving a co-teaser to tease the child (Adult and Co-teaser > Child)	involving the child as co-teaser in teasing another person (Adult and Child > Other)
—	involving a co-teased (teasing the child and another person simultaneously) (Adult > Child and Other)
—	or teasing the child by getting her tease a mate or another person (Adult > Child > Other)
improvising mock dialogue with shifts between the roles of speaker and addressee (Adult and "Child" > Child)	
(Basic speech acts realizing items of indirect teasing are potentially those listed in Direct teasing)	

The high frequency of teasing (also a preferred way of speaking among adults) in the speech directed to children as well as the topics and discourse strategies used in them seem to fit the general patterns of CDS (child directed speech) in traditional Gypsy communities described earlier in Réger–Gleason (1991). Their paper demonstrated that, in traditional Gypsy communities, children from the moment of birth are considered as full members of their community and potential communicative partners and that, almost from the moment of birth on, they are directly taught and trained for future gender-specific roles. Although adults and older children do adjust their speech to their linguistically immature partner by using, for example, baby talk lexicon and phonology (see Réger–Gleason 1991),[7] ways and genres of speaking are often explicitly modelled, demonstrated, and possibly tested from very early on. Some of the specific features of teasing babies and young children—among others, the direct modelling of teasing sequences or, on the level of the content, the involvement with future gender specific roles—thus proved to fit the general patterns of adult–child interaction in traditional Gypsy communities outlined in this earlier paper.

On the other hand, the great emphasis on teaching culture-specific ways of speaking to babies and young children may be related to the fact that knowledge of these skills is considered basic for Gypsy culture and identity (see Réger–Gleason 1991).

Some of the differential features found in early and later teasing, more specifically, the decrease in the occurrence of direct sexual teasing with age may perhaps be related to beliefs concerning the child's changing status in traditional Gypsy communities. Anthropological research done in the same community has demonstrated that rules of behaviour connected with beliefs of pollution are still very widespread in it (Stewart 1987). These rules prescribe, for example, the strict separation of the upper and lower parts of the body, and different customs related to childbirth and women's fertility. Babies are, however, considered to be exempt from pollution, at least once the perinatal period is passed. It is perhaps for this reason that there seem to be no rules restricting the metaphorical expressions of love and tenderness when speaking to them.

How long is the state free of pollution supposed to last? Sutherland (1975) puts its limit at the age of puberty,[8] but our data seem to indicate that, between infancy and childhood, there may be a secondary age limit as well. This point needs further investigation.

---

[7] See also the use of the form *co* [tso] instead of *so* [so] in line 4 and the consonant assimilation in the word *mistoj* > *mittoj* in line 39 in Example 1. These items illustrate typical phonological modifications in conformity with the rules of Baby Talk phonology (see Ferguson 1977; Réger – Gleason 1991).

[8] In the community studied by Sutherland, babies become *wuzho* or free from pollution at six weeks, and preserve this status until puberty (Sutherland 1975, 262).

3. Glimpsing at the earliest teasing skills of a Gypsy child

Beyond types of adult teasing, Examples 2 and 3 show that at the age of 2;1, the young child has already proved to be an eager and relatively skilled partner in performing her part within the teasing sequences. Thus, for example, she has already been able to provide the expected replies in the teasing routines ("filling in" the appropriate lines, see lines 2, 4 and 6 in Example 3). Moreover, she has already been able to creatively connect portions of two different routines (that of the sexual and that of the "good wife" routines, see line 11 in Example 2). She was also able to initiate teasing by herself (see Example 4, line 15) by using playful, sing-song intonation, and to use teasing for other, more sophisticated pragmatic ends as well (for example to "defuse" impending threat, see Example 5, line 16).

EXAMPLE 4

1	Child:	Anyu, butìl mange!
		Mum, undo it for me!
2	Mother:	Bontìl.
		I am undoing.
3	Child:	Anj-
		Mom-
4	Mother:	De v è Dombeh ek cerra, more!
		Give a piece to Dombi, girl!
5	.	Phàg l e Dombehke!
		Break it off for Dombi!
6		De!
		Now!
7	Child:	Na maj-!
		There is no-!
8	Mother:	Hat aś ta, bontij les!
		Wait, I am undoing it!
9		Či mukheh te bontil leh?
		Won't you let me undo it?
10	Child:	Na.
		I won't.

11   Mother: (with pretended surprise)
             Na!
             You won't!

12   Mĭr?
             Why?

13   Bico!
             Bici!

14   Child: ar] #

             ...

             (with sing-song intonation)
15   Me či dav!
             I won't give him any!

16   Mother: De vì leh ek cino, mĭr̄i śej!
             Yet give him a small (piece) my daughter, too!

17   Maj vì lehki dej kerel, dèl tu!
             His mother is going to make, she too, and she is going to give you!

EXAMPLE 5

1   Mother: Pekèl o khàm?
             Is the sun shining?

2   Child: Paj.
             Water.

3   Mother: tu [?] paj ande fejastra?
             Is there [?] water on the window-sill?

4   Child: Aha!
             Yes.

5   Mother: (with pretended anger)
             Kòn śordah les?
             Who poured the water there?

6   Kòn śordęh koth o paj?
             Who poured the water there?

7   Child: Ha?
               Hmm?

8   Mother: (with pretended anger)
               Kòn śordęh koth o paj?
               Who poured the water there?

9   Child: Me!
               Me.

10   Mother: Tu!
               You!

11   Ha!
               Oh!

12   Ha tu sằr buśes?
               And what is your name?

13   Tu sằr buśes?
               What is your name?

14   Child: Karing?
               Where?

15   Mother: Karing.
               Where?

16   Child: (with sing-song intonation)
               Keràv tēlē!
               I wipe it up!

Modification of intonation patterns (the use of sing-song intonation) seemed to be the primary means of the young child to frame her utterance as teasing (see also Eisenberg 1986; Miller 1986 for similar findings). A more thorough description of the development of early teasing in young children will require the study of the developmental patterns of intonation. The examples cited however, seem to demonstrate that as a consequence of their linguistic socialization, children in traditional Gypsy communities are able to recognize and use from very early on some of the specific "contextualization cues" necessary for the identification of the underlying intention of teasing beneath the surface form.

4. Conclusion

Preliminary analysis of child–adult interaction in a traditional Gypsy community has highlighted the pervasive importance of teasing in Gypsy children's early linguistic socialization. Topics and discourse strategies used in early teasing revealed some of the beliefs and expectations surrounding babies and children as well as parents' endeavour to teach them as early as possible the ways of speaking considered as useful and appropriate in their community. Among the topics of teasing, metaphorical use of sexual teasing addressed to infants and teasing related to the baby's or child's future life and expected gender-specific behaviour seemed to be particularly important. Direct sexual teasing, however, seemed to decrease with age.

On the other hand, some of the discourse strategies used in teasing reflected the marked tendency of mothers and other adults to provide the child with explicit models of proper ways of speaking. These included: inviting a co-teaser, shifting roles or using explicit disclaimers in teasing babies, and in later teasing, giving the child lines to repeat or getting her to tease another person. (At the same time, teasing has become increasingly sophisticated over time, with more frequent indirect teasing occurring as the child grows older.) Use of directives (direct or indirect) in later teasing served, among others, to elicit gender-specific behaviour and emotional display from the child. All these findings are in accordance with the results of the earlier study on the general characteristics of adult–child interaction in traditional Gypsy communities.

This quality of input may affect the processes of socialization itself, as well as those involved in language learning. First, evidently, it must have a strong impact on socialization to sex roles as well as on the socialization of affect—the process of learning how to feel and how to express feelings and emotions according to the patterns of the given culture. In a more subtle way, these characteristics of the input language may affect the processes involved in language acquisition as well. For example, in this type of linguistic socialization, intonation may play a particularly important role in "bootstrapping" grammatical acquisition (perhaps even beyond the great importance attributed to it in a recent work, see Morgan–Demuth 1996). As to language use, children learn from very early on that in understanding and interpreting messages, they have to rely strongly on contextual cues—a factor that may lead to a highly context-dependent way of speaking. On the other side, the practice of subtle variation in ways of involving the child in teasing, in changing speaker(s) to hearer(s) alignment may lead to the development of skilful conversationalists who will then be able to skilfully manage and negotiate different conversational purposes in a variety of speech situations.

Gypsy children's increasing skills in teasing may thus be as much a source of advantage as disadvantage in later life. On the one hand, this type of context-dependent linguistic socialization is probably at the core of their difficulties with the use of decontextualized language in the classroom. On the other hand, proficient teasing and what it involves: extraordinary sensitivity to contextual cues, and the skilful manipulation of the frames of communication, may represent a basic survival skill in the hostile and prejudiced social environment in which most of them will have to live their later life.

References

Eisenberg, A.R. 1986. Teasing: verbal play in two Mexicano homes. In: Schieffelin, B.B. – Ochs, E. (eds): Language socialization across cultures, 182–98. Cambridge University Press, Cambridge.

Ely, R. – Gleason, J.B. 1995. Socialization across contexts. In: Fletcher P. – MacWhinney, B. (eds): Handbook of child language, 251–70. Blackwell, Oxford.

Ferguson, C.A. 1977. Baby talk as a simplified register. In: Snow, C.E. – Ferguson, C.A. (eds): Talking to children, 209–35. Cambridge University Press, Cambridge.

Fishman, J.A. 1967. Bilingualism with and without diglossia: Diglossia with and without bilingualism. In: The Journal of Social Issues 2: 29–38.

Fletcher, P. – Garman, M. 1988. LARSPing by numbers. In: British Journal of Disorders of Communication 23: 309–21.

Goffman, E. 1981. Forms of talk (Chapter 3. Footing), 124–57. University of Pennsylvania Press, Philadelphia.

Gumperz, J.J. 1982. Discourse strategies. Cambridge University Press, Cambridge.

Havas, G. – Kemény, I. 1995. A magyarországi romákról [About the Roms in Hungary]. In: Szociológiai Szemle, 3: 3–20.

Heath, S.B. 1983. Ways with words: language, life and work in communities and classrooms. Cambridge University Press, Cambridge.

Hymes, D. 1972. Foundation of sociolinguistics. University of Pennsylvania Press, Philadelphia.

Kaprow, M.L. 1989. Resisting respectability: Gypsies in Saragosa. Urban Anthropology 11 (3–4): 399–431.

MacWhinney, B. 1994. The CHILDES project. Computational tools for analyzing talk. Second edition. Erlbaum, Hillsdale NJ.

Miller, P. 1986. Teasing as language socialization and verbal play in a white working-class community. In: Schieffelin, B.B. – Ochs, E. (eds): Language socialization across cultures, 199–212. Cambridge University Press, Cambridge.

Morgan, J.L. – Demuth, K. (eds) 1996. Signal to syntax. Bootstrapping from speech to grammar in early acquisition. Erlbaum, Mahwah NJ.

Ninio, A. – Snow, C.E. 1996. Pragmatic development. Westview Press, Boulder CO.

Réger, Z. 1986. Report of the research project Disadvantaged Social Background and Language Development. Unpublished manuscript.

Réger, Z. – Gleason, J.B. 1991. Romani child directed speech and children's language among Gypsies in Hungary. In: Language in Society 20: 601–17.

Schieffelin, B.B. 1986. Teasing and shaming in Kaluli children's interactions. In: Schieffelin, B.B. – Ochs, E. (eds): Language socialization across cultures, 165–81. Cambridge University Press, Cambridge.

Silverstein, Michael. 1976. Shifters, linguistic categories and cultural description. In: Basso, K. – Selby, H.A. (eds): Meaning in anthropology, 11–55. University of New Mexico Press, Albuquerque.

Stewart, M.S. 1987. Brothers in song. Thesis submitted for the degree of Ph.D. at The London School of Economic and Political Science, Faculty of Economics.

Sutherland, A. 1975. Gypsies: The hidden Americans. The Free Press, New York.

Vekerdi, J. 1977. A magyarországi cigány nyelvjárások [Gypsy dialects in Hungary]. Unpublished academic doctoral dissertation. Hungarian Academy of Sciences, Budapest.

Address of the author: Zita Réger
Linguistics Institute
Hungarian Academy of Sciences
Színház u. 5–9.
H–1250 Budapest P.O.Box 19
Hungary
e-mail: reger@nytud.hu

THE ROMANI LANGUAGE IN THE REPUBLIC OF MACEDONIA: STATUS, USAGE, AND SOCIOLINGUISTIC PERSPECTIVES*

VICTOR A. FRIEDMAN

Abstract

The development of Romani language use in the Republic of Macedonia is discussed here primarily in terms of its relation to education and other public and official contexts. The paper gives a detailed account of the statistical and legal position of the Roms in Macedonia and former Yugoslavia, followed by a linguistic account of the standardization of Romani for official use in Macedonia with comments on dialects, political action, publishing, the educational establishment, and the situation in other Balkan countries. Romani has gradually risen in status from total legal absence to legal equality with all other minority languages, even if the *de facto* realization has not yet met the *de jure* possibilities. The standardization of Romani and the fixing of the Arli dialectal base with other elements, together with a consistent orthography, has made significant progress.

As an ethnic group in Macedonia, the Romani people have occupied a marginalized place similar to that in which they find themselves elsewhere in Europe, but with an important difference. While they have been subjected to discrimination (see Silverman 1995a; 1995b), they have not been the target of the kind of racist violence that has occurred and still occurs elsewhere in Europe (cf. Bárány 1994; 1995; Kyuchukov 1995). In fact, in the complex ethnic mosaic of Macedonia, the Roms[1] have maintained their separateness while at the same time functioning as an integral and accepted part of everyday Macedonian life. Since the constitutional revi-

[1] See note at *Rom* in Table 2.

sions of 1974, when the Roms received legal recognition as an ethnic collective, there have been various attempts to advance education in Romani—educational linguistic rights being a cornerstone of identity politics. With the independence of the Republic of Macedonia, the Roms were recognized in the 1991 constitution as a *narodnost* 'nationality' on the same level as Albanians, Turks, Vlahs, and others (mainly Serbs). Although Roms seek to participate in formal educational institutions using their own language, however, the Romani language itself is still in the process of the kind of standardization associated, among other things, with the institutionalized structures of formal education. In this paper, I shall discuss the development of Romani language use in the Republic of Macedonia primarily in terms of its relation to education and other public and official contexts, utilizing both published sources and my own experiences during more than twenty years of field research in Macedonia. In order to frame these issues, I shall first turn to the statistical and legal position of the Roms in Macedonia and former Yugoslavia.

According to the 1994 census Roms constitute approximately 2.3% of the population of the Republic of Macedonia.[2] In former Yugoslavia, Roms constituted between 0.6% and 0.7% of the total population, approximately two thirds in the Republic of Serbia and about one quarter in the Republic of Macedonia. As can be seen from Tables 1 and 2, the figures on people declaring Romani nationality have varied considerably over the years. This is not due to changes in birth or death rates, since the Roms have a natality rate almost twice the Yugoslav average (Stanković 1992, 173). While mechanical growth (migration) could account for some fluctuations, there were no migrations massive enough to account for these differences. Rather the magnitude of changes in the figures reflect differences in social pressure concerning the declaration of Romani as opposed to some other nationality. The tremendous drop in 1961 in the number of self-declared Roms in Serbia but not in Macedonia has been labeled a "statistical enigma" (Stanković 1992, 160), but must clearly reflect a difference between either the censusing or social position (or both) of Roms in Serbia and in Macedonia.

[2] Romani ethnopoliticians have on occasion claimed as many as five times the official figures (Čangova 1991), and political leaders of all the other major ethnic groups inside Macedonia as well as politicians in neighboring states with histories of territorial claims against Macedonia have at one time or another claimed numbers as much as 60 times in excess of the 1994 census figures. The point is not one of statistical accuracy but rather claims to political power and hegemony (see Friedman 1996). Nonetheless, given the social stigma attached to Romani identity and the marginalized position of Roms in society, it is well known in demographic literature that Roms often choose to declare a different nationality (Stanković 1992, 161). Another example of conflicting tensions was the appeal to Roms during the 1991 census to register as Roms and not as Albanians or Turks or "something else" out of fear or due to pressure (Nova Makedonija 28.III.91, 4). To the extent that the Macedonian government is seeking to reduce Albanian influence, this has led to increased recognition for both Roms and Macedonian Muslims.

Table 1

Figures concerning Romani nationality in Macedonia and Former Yugoslavia
in the five uncontested post-war censuses. Source: Stanković (1992)

	1981	1971	1961	1953	1948
Macedonia	43125=25.7%	24505=31.2%	20606=65%	20462=24.2%	19500=26.8%
Serbia	110959=66%	49894=63.6%	9826=31%	58800=69.5%	52181=71.8%

The 1991 Yugoslav census was boycotted by part of the population, mainly ethnic Albanians.

Table 2

The ethnic structure of the Republic of Macedonia since World War Two

Census total by year, number, and percentage (rounded upward where necessary)

Declared nationality (narodnost)	1948	%	1953	%	1961	%	1971	%	1981	%	1991	%	1994*	%
Macedonians	789548	68.5	860699	66.0	1000854	71.2	1142375	69.3	1279323	67.0	1328187	65.3	1295964	66.5
Albanians	197389	17.1	162524	12.4	183108	13.0	279871	17.0	377208	19.8	441987	21.7	441104	22.9
Turks	95940	8.3	203938	15.6	131481	9.4	108552	6.6	86591	4.5	77080	3.8	78019	4.0
Roms**	19500	1.7	20462	1.6	20606	1.5	24505	1.5	43125	2.3	52103	2.6	43707	2.3
Vlahs	9511	0.8	8668	0.6	8046	0.6	7190	0.6	6384	0.3	7764	0.4	8601	0.4
Serbs	29721	2.6	35112	2.7	42728	3	46465	2.8	44468	2.3	42775	2.1	40228	2.0
Muslims	1560	0.1	1591	0.1	3002	0.2	1248	0.1	39513	2.1	31356	1.5	15418	0.8
Bulgarians	889	0.1	920	0.1	3087	0.2	3334	0.2	1980	0.1	1370	0.0	1682	0.1
Greeks	–	–	848	0.1	836	0.1	536	0.0	707	0.1	474	0.0	368	0.0
Egyptians	–	–	–	–	–	–	–	–	–	–	3307	0.2	3080	0.2
Bosniacs	–	–	–	–	–	–	–	–	–	–	–	–	6829	0.4
Yugoslavs	–	–	–	–	1260	0.1	3652	0.2	14225	0.7	15703	0.8	595***	0.0
Others****	8928	0.8	9752	0.8	10995	0.7	29580	1.7	15612	0.8	31858	1.6	9797	0.4
Total	1152986	100	304514	100	1406003	100	1647308	100	1909136	100	2033964	100	1945932	100

Sources: Antonovska et al. (1991; 1994a; 1994b; 1996), Latifić et al. (1970), Pekevski et al. (1973), Savezni zavod za statistiku (1954; 1981).

* According to Dr. Svetlana Antonovska (p.c. 95/05/25), Director of the Republic Bureau of Statistics, the lower figures for some nationalities in 1994 vs. 1991 is due to the fact that citizens living abroad for more than one year were included in the 1991 census, whereas in the 1994 census—in accordance with international norms—only those citizens living abroad for one year or less were counted.

** The predominantly Romani-speaking ethnic group popularly known as *Gypsies* in English and *Cigani* in Macedonian (similar ethnonyms are used in most of the languages of Central and Eastern Europe) is now referred to by the native ethnonym *Rom* (singular) in scholarly literature as well as official documents in many countries. (The term was official in the 1971 Macedonian census.) Although in languages other than English this word has been unhesitatingly adapted to the grammar of the language

in which it is used (e.g. Bulgarian, Macedonian, Romanian, Serbo-Croatian pl. *Romi,* Albanian pl. *Romë*), considerable inconsistency has arisen in English usage. Thus as the plural of *Rom* some scholars and other serious writers use the Romani form *Roma,* others adapt the word to English morphology and write *Roms,* others use a pluralized adjective *Romanies,* and some use the Kalderash collective *Rom* for both singular and plural. I have argued elsewhere (Friedman – Hancock 1995) that just as in English the plural of *Turk* is *Turks* and not *Turkler,* so the plural of *Rom* should be *Roms* and not *Roma.* I would argue that the form *Roma* exoticizes and marginalizes rather than emphasizing the fact that the group in question is an ethnic group just as are *Turks, Magyars* (not *Magyarok*), *Bulgars* (not *Bulgari*), etc.

*** This figure includes those who declared "Yugoslav" as well as nationalities not counted separately in the census, mostly from Africa, East Asia, and the Middle East (MIC 95/01/05).

**** Yugoslav and Macedonian censuses distinguished up to 34 nationality categories as well as several other types including those who declared a regional indentity and those who did not declare a nationality. For the sake of conciseness, I have grouped all the smaller categories, none of which are relevant for this paper, under the designation Other. This designation includes the following specified groups: Austrian, English, Belgian, German, Danish, Jewish, Italian, Norwegian, Polish, Romanian, Russian, Rusyn, Slovak, Slovenian, Ukrainian, Hungarian, French, Dutch, Croatian, Montenegrin, Czech, Swiss and Swedish. This category also includes those who protested the use of nationality as a classification by making facetious declarations, among which the most popular were Lightbulb and Refrigerator (Robert Hayden, University of Pittsburgh, personal communication).

Between the 1971 and 1981 censuses, there was a change in the legal status of Romani that both reflected and encouraged a rise in consciousness of Romani identity—viz. the federal and republic constitutional reforms of 1974 in which Romani (along with Vlah) received the official status of *etnička grupa* 'ethnic group', a step below *narodnost* 'nationality' (the term which came to replace 'national minority' [Macedonian *nacionalno malcinstvo* Serbo-Croatian *nacionalna manjina*] during the 1960's and became official in the 1974 constitutions).[3] This rise in national con-

---

[3] The 1974 constitution recognized three types of ethnically defined collectives: *narod* 'nation', *narodnost* 'nationality', and *etnička grupa* 'ethnic group'. The difference between a *narod* and a *narodnost* was that a *narod* was considered a constitutive nation of Yugoslavia and of its constituent republics (Slovene, Serb, Croat, Macedonian, Montenegrin, and Muslim) whereas a *narodnost* was *de facto* a minority that was a constituent of a national-state other than Yugoslavia, e.g. Turks. An ethnic group was a minority with no nation state beyond the borders of Yugoslavia, i.e. the Vlahs and the Roms. An exception to this principle were the Ruthenians (Rusyni), who live primarily in Vojvodina and who did not have an external nation-state but were nonetheless given the status of *narodnost*. A major complaint of the Albanians during this period was that while they constituted a numerically larger group than Macedonians or Montenegrins, they were considered a *narodnost* while the latter each constituted a *narod*. Each category implied a different level of linguistic and other collective rights mitigated by factors of size and distribution: the language of a *narod* (Slovenian, Macedonian, Serbo-Croatian) was official at the federal level. However, federal laws and regulations were also to be published in Albanian and Hungarian, making them semi-federal. The language of a *narodnost* was official at the republic or provincial level (e.g. Turkish in Macedonia, Hungarian in Vojvodina), the communal (municipality) level (e.g. Italian in Slovenia, Bulgarian in Serbia), or not at all (e.g., German, Polish, and Russian) (see Bugarski 1992; Škiljan 1992). The languages of ethnic groups did not receive guaranteed official support, but their constitutional recognition positioned them to seek such support. Although the Roms had the status of *narodnost* in the Republic constitution of Bosnia-Hercegovina, this had no practical effect (Škiljan 1992, 40).

sciousness was parallel with a rise in linguistic consciousness. It was during this period that the first serious attempts in the direction of Romani-language education were made in Macedonia.

In discussing the relationship of the Romani language to Romani nationality, it is important to keep in mind that there is not an absolute one-to-one correspondence between the two. The figures in Table 3 show the correlation between declared nationality and declared mother tongue for the first and last uncontested Yugoslav censuses conducted in Macedonia and the 1994 Macedonian census.

Table 3
Difference between declared nationality and declared mother tongue for the six main languages
of the Republic of Macedonia: 1953, 1981, 1994

1953
Declared Mother Tongue

Declared nationality	Macedonian	Albanian	Turkish	Serbo-Croat	Romani	Vlah
Macedonians	853971	1986	281	934	277	2565
Albanians	2152	153502	6569	181	70	1
Turks	32392	27087	143615	534	70	10
Roms	1040	860	2066	25	16456	1
Vlahs	137	4	2	14	0	8130
Serbs	3945	0	8	31070	41	9
Muslims	*	*	*	*	*	*
Yugoslavs	2152	25	50	563	2	4
Others	322	341	569	5258	173	31
Total	896651	183805	153160	38579	17089	10751

1981
Declared Mother Tongue

Declared nationality	Macedonian	Albanian	Turkish	Serbo-Croat	Romani	Vlah
Macedonians	1276878	190	160	547	316	*
Albanians	1218	374181	3	440	1697	*
Turks	16608	8592	60768	366	94	*
Roms	4160	1697	808	24	36399	*
Vlahs	1111	1	0	3	2	5257
Serbs	8521	10	3	35867	14	*
Muslims	15075	4968	2038	16325	308	30
Yugoslavs	7645	1943	274	2746	530	*
Others	13282	4247	2853	17031	1280	*
Total	1334498	391829	64907	63349	37780	5931

*Not specified

Table 3 (continued)

1994
Declared Mother Tongue

Declared nationality	Macedonian	Albanian	Turkish	Serbo-Croat[†]	Romani	Vlah	Bulgarian
Macedonians	1289868	124	124	1938	94	259	125
Albanians	2063	426418	210	135	•	•	0
Turks	10885	906	62726	86 + •	11	0	•
Roms	5974	1212	1311	14 + •	34955	•	•
Vlahs	1800	•	0	12	0	6747	0
Serbs	11693	0	8	27843	•	•	32
Muslims	5552	605	180	1181	20	•	•
Bosniacs[††]	36	21	•	312 + •	0	0	•
Bulgarians	358	0	•	80	0	•	1216
Egyptians	961	1856	42	10	•	0	0
Others	3793	221	64	3484	40	30	75
Total	1332983	431363	64665	35095	35120	7036	1448

Sources: Savezni zavod za statistiku (1953a), Savezni zavod za statistiku (1988), Antonovska *et al.* (1996)

† This figure represents *Serbian* and *Croatian* which were listed as separate languages in the 1994 census.
†† 6426 Bosniacs were listed in the column 'Other' for mother tongue. Presumably the overwhelming majority declared Bosnian. 7795 Muslims were also in the 'Other' column and presumably also listed Bosnian.
• Under 10

As can be seen from the table, there is a fairly high correlation between declared Romani nationality and declared Romani mother tongue. In fact, the correlation of over 96% is well above the Yugoslav average of 79.1% of those with Romani nationality declaring Romani mother tongue (Petrović 1992, 120). This can be taken as an indicator of the strength of the correlation between declared Romani language and nationality in Macedonia. What these figures do not—and cannot—reveal, however, is the fact that many Roms declare another nationality (and/or mother tongue) due to the social stigma attached to Romani.[4] Since the majority of Roms in Macedonia are Muslim, and moreover urban, Turkish represents a significant prestige language while

[4] Albanian and Turkish are probably more frequent than Macedonian due to the fact that most Roms in Macedonia are Muslim and the former two languages are more closely identified with Islam (cf. note 2). According to Faik Abdi, cited in Čangova (1991), most Roms who declare another nationality and/or mother tongue in western Macedonia declare themselves as Albanians while in eastern Macedonia they declare themselves as Turks. This would reflect the relative numerical strength of these two groups in these two regions.

Albanian represents numerical strength as the language of the largest predominantly Muslim minority. We can also note here that the drop in Roms declaring Turkish mother tongue between 1953 and 1981 correlates with the migration of Turks to Turkey (largely for economic reasons) in the late 1950s (cf. Katona 1969; Jašar-Nasteva 1992). Many other Muslims also declared themselves as Turks on the basis of religion in order to emigrate to a non-communist country. The tremendous discrepancy between declared Turkish and Albanian nationalities in 1948 and 1953 was politically motivated. The 1948 census was conducted before the Tito–Stalin break, when relations with communist Albania were good and relations with non-communist Turkey were bad. By 1953, Yugoslavia had been expelled from Cominform and was not on good terms with Albania whereas by contrast relations with Turkey had thawed considerably (cf. Tanasković 1992). Although not readily ascertainable from census figures, these changes in relations also affected Roms, albeit not those declaring Romani mother tongue. At present, the issue of education for non-Romani speaking Gypsies also involves ethnic politics.[5] There is pressure on Muslim Gypsies to go to Albanian or Turkish rather than Macedonian schools, the better to justify expanding minority language education (cf. e.g. Flaka e Vëllazërimit 86.01.06, 10 on the situation in Kumanovo, also Birlik 84.10.01, 14). Current concern in Macedonia with Romani education is not merely connected to Article 48 of the Republic's constitution, which guarantees minority language rights, but can also be seen as aimed at reducing challenges from Albanian and Turkish.

There are also the *Gupci,* or *Egipkani,* endogamous, non-Romani speaking groups of Romani descent who do not identify as Roms and who in Macedonia speak Albanian (e.g., in Ohrid and Struga) or Macedonian (e.g., in Bitola) as their first language (Friedman 1985b; Ljubisavljević 1990; Risteski 1991; Duijzings 1992; 1997; Hadži-Ristik 1994; Zemon 1996).[6] Although they sought to be recognized as a separate category in the 1981 census, they were placed in the category "unknown" (Nova Makedonija 82.03.06, 9), whereas in the 1991 and 1994 censuses they were recognized as a distinct group (but see Abduramanoski 1994). Both Romani and Albanian ethnopoliticians claim them for their own, but the Gupci identify with neither.

While Romani-speaking groups are often associated in the popular mind with nomadism, and indeed many groups were and some still are peripatetic—in France, for example, more than half the Gypsy and Traveler population is nomadic or semi-seden-

---

[5] I am using the term *Gypsy* here as a cover term for all those European groups descended from Indic speakers who arrived in Europe during the Middle Ages.

[6] Like the English term *Gypsy,* these ethnonyms derive from ethnonymic terms meaning 'Egyptian'.

tary (Chalumeau–Gualdaroni 1995)—Romani people have been settled in the Balkans
in general and Macedonia in particular for centuries (see Friedman–Dankoff 1991).
The social situation of many if not most Roms in Macedonia is thus quite different
from that of many groups living elsewhere. The education issues facing many
European countries, and the United States as well, involve significant differences
between Romani and non-Romani culture—issues such as how to deliver societal ser-
vices to nomadic groups, adapting the educational curriculum to the needs of children
from a very different culture, etc. The Romani-speaking people of Macedonia, howev-
er, are part of a region where multilingualism and multiculturalism are a centuries-old
tradition. The very existence of the Balkan linguistic league is testimony to this. The
Balkan linguistic league (or *Sprachbund*) consists of the Balkan Slavic languages
(Bulgarian, Macedonian, and the Torlak dialects of Serbian), the Balkan Romance lan-
guages (Romanian, Megleno-Romanian, and Aromanian [Vlah]), Greek and Albanian,
all of which share a variety of significant structural similarities as the result of centuries
of language contact. As the Slovene linguist Jernej Kopitar wrote in 1829, these lan-
guages gave the impression of having a single grammar (*Sprachform*) with different
lexicons (*Sprachmaterie*). Moreover, although they are not usually included in studies
of the Balkan linguistic league, the Balkan Romani dialects do in fact share a number
of significant grammatical features with the other Balkan languages (see Kostov 1973;
Friedman 1985a; Matras 1994). A brief illustrative example is given in Table 4.

Table 4
Optative-Subjunctive particle replaces infinitive and other structures in the Balkan Languages

Romani	te džas	mangav	te	hramonav
Albanian	të shkojmë	dua	të	shkruaj
Greek	ná páme	thelō	ná	gráfō
Bulgarian	da trŭgnem	iskam	da	piša
Macedonian	da odime	sakam	da	pišuvam
Torlak Serbian	da idemo	oču	da	pišem
Romanian	să mergem	vreau	să	scriu
Vlah (Kruševo)	s- neádzimŭ	voi	si	scriu
gloss	'let us/if we go'	'I want to write'		

However, while Romani-speakers constitute an integral albeit distinct and
sometimes marginalized segment of Macedonian society, and while issues in
Romani education parallel issues in other minority language education in
Macedonia, there is a significant difference: while the Albanian, Turkish, and
Serbian minorities in Macedonia have codified languages to serve as the bases of
education, Romani, like Vlah (Aromanian), lacks such a standard. In the case of

Vlah, there is a codified literary language, viz. Romanian, which is sufficiently close that it can provide a model if not a substitute. In fact, there is a movement that would replace Vlah with Romanian, but this is resisted by the majority of Vlahs in Macedonia (cf. Jašar-Nasteva 1997) in the same way that, e.g., the Macedonians resisted the imposition of Bulgarian, the Ukrainians resisted Russian, the Slovaks resisted Czech, or the Norwegians resisted Danish (cf. Haugen 1968). Romani, however, is faced with a different problem which makes it unique in Macedonia and unusual in Europe: not only does there not exist at this time an established Romani linguistic norm, but Romani's closest relatives—the languages of western India such as Hindi and Punjabi—are too distant from Romani to have even the slightest chance of substituting for it, although their relationship to Romani in terms of vocabulary enrichment is a separate issue, one which we shall discuss below (cf. also Friedman 1989). Thus Romani education cannot take place effectively without settling the Romani "questione della lingua" (čhibakoro phučipe).

As was mentioned earlier, the 1970s saw both a change in Romani legal status and attempts at advancing Romani-language education. In general, however, these attempts met with a variety of difficulties. In 1971, Šaip Jusuf, a Rom who had earned a B.A. in physical education from the University of Belgrade, began work on a Romani grammar with Krume Kepeski, a professor at the Skopje Pedagogical Academy (Nova Makedonija 80.02.15, 10; cf. also Koneski 1950; Lunt 1952, v). By 1973 Jusuf and Kepeski had completed the manuscript of their grammar (Prof. Kepeski was kind enough to show me the manuscript while I was in Macedonia), and they were seeking publication. Due to various complicating factors, however, the grammar did not appear until 1980. The appearance of Jusuf–Kepeski (1980) in a tirage of 3,000 copies signaled a new phase in the development of the standardization of Romani in Macedonia. The book is written in both Romani and Macedonian on facing pages and was the most ambitious attempt of its kind at the time. The express purpose of the book was the creation of a Literary Romani for use by Roms in Macedonia, Kosovia,[7] and adjacent parts of Serbia, with a view to the creation of Romani-language schools in these areas and to the use of this literary standard as a basis for the creation of a Romani literary language for use by Roms in general (Jusuf–Kepeski 1980, 4–5).[8] The language of the grammar is

[7] Because both the [original] Slavic Kosovo and the Albanian [and Turkish] Kosova currently have political implications whose complexities I wish to eschew, I have chosen to use the productive English suffix when writing about this place in English.

[8] The question of whether Romani is to be considered a single language with numerous dialects or as a group of closely related languages is not of immediate concern to this article. The most commonly held opinion among linguists is that Romani is to be treated as a single language (cf. Kochanowski 1963, 184–92; Hancock 1975, 26; Ventcel' – Čerenkov 1976, 283; Cortiade 1984) despite varying degrees of mutual intelligibility. Questions of the definition of language and dialect

based on the Arli dialect of Skopje, although Jusuf makes frequent use of his native Džambaz dialect—especially when citing Romani forms in the Macedonian text—and occasionally Gurbet and Burgudži forms are also mentioned.[9] I have published a detailed analysis of this grammar elsewhere (Friedman 1985c). For the purposes of this paper it will suffice to point out some of the most salient types of problems raised by Jusuf–Kepeski (1980) with respect to the standardization of Romani and its use in education:

1. Orthographic conventions were not standardized as illustrated by the following examples: syllable final jot is indicated by both *i* and *j* as in the spellings *muj* and *mui* 'mouth', the automatic fronting of velars before front vowels is inconsistently indicated, e.g. *kerdo* and *kjerdo* 'done', the opposition between a uvular fricative /x/ and a glottal glide /h/—phonemic in some Romani dialects but not in others—is not made consistently, e.g. *xor* 'depth' but *hordaripe* 'deepening', *xramonel* 'write' but *hramondikano* 'written', etc.

2. Competing dialectal forms are not selected but rather mixed, as seen in the following examples. The basic form of the instrumental singular marker is {-sa} but the /s/ is lost intervocalically in Arli. On the Romani side of one of the nominal paradigms, the instrumental singular of the word for 'wind' is given as *bavlal-aa, -asa* while on the Cyrillic side it is given as бавлалаја (*bavlalaja*). In fact, *bavlal* is the Arli dialectal form, the Džambaz and etymologically older form being *balval*. Similarly, the second singular present tense morpheme, which also has the basic shape {-sa} and has both the Arli-specific loss of /s/ and a morphological variant without the final /a/ in all the dialects, is used in various places in all its possible

---

involve more factors than mere intelligibility, as can be seen from such classic examples as the mutually unintelligible *dialects* of Chinese or the mutually intelligible *languages* of Scandinavia, but the discussion of the relationship among language, dialect, ethnic identity, and national identity are beyond the scope of this article. For our purposes, we shall adopt the commonly held view of linguists just alluded to (cf. also Haugen 1966; Lunt 1984).

[9] We are accepting here as a useful heuristic device the distinction between the so-called Vlax and Non-Vlax dialects of Romani. Although the Romani dialectal situation in the Republic of Macedonia is quite complex, the majority of speakers use dialects of a Non-Vlax type that are described by the self-ascriptive cover term Arli (< Turkish *yerli* 'local'). Next in importance for Macedonia is Džambaz (< Turkish *cambaz* 'acrobat, horse-dealer', known elsewhere as Gurbet, related to Kalderaš, Lovari, Čurari, Mačvano, etc.), which is a Vlax type dialect that has undergone Non-Vlax influence. Also of significance for Macedonia is Burgudži or Bugurdži (< Turkish *burgucu* 'gimlet-maker', also known as Rabadži [< Turkish *arabacı* 'drayman'] or Kovačja [< Slavic *Kovač* 'blacksmith'], a name which is also used for other groups including the non-Romani speaking Gupci of southwestern Macedonia), which is also a Non-Vlax dialect. The details of Romani dialectal differentiation are beyond the scope of this paper. For a summary of current theories see Hancock (1995). Vencel' – Čerenkov (1976) and Boretzky – Igla (1994, 361–415) give useful comparative sketches of Romani dialects. Cf. also Table 7 for some illustrative examples.

realizations: *keresa/kereja/kerea* ↔ *keres/kere* 'you do'. Similarly, for the nominative plural definite article both Arli/Burgudži *o* and Džambaz/Gurbet *e* are used, e.g. *o Roma* and *e Roma* 'the Roms', feminine nouns in consonants are used with both jotated and non-jotated oblique stems, e.g. *chiba-* and *chibja-* 'tongue, language', etc.

3. Neologisms are coined from Hindi, sometimes with disregard for the Romani phonological system, rather than based on native material or borrowed from languages familiar to the speakers e.g. *bhaga* 'consciousness'.

4. The grammar was written on a level for use in a high school or pedagogical academy, but at the time there were no textbooks at the elementary school level. The grammar could thus at most have been used to prepare teachers, but the cadre of educated and motivated individuals and particularly the organizational structure was lacking.

Although the cultural organization *Phralipe* 'Brotherhood' was formed in 1948, it was not until the 1970s and 1980s that sporadic attempts were made at advancing Romani education and related linguistic rights such as use in the mass media. Thus there were radio programs broadcast out of cities and towns such as Belgrade and Niš in Serbia and Tetovo in Macedonia (cf. Puxon 1979, 89), a monthly entitled *Krlo e romengo* 'Voice of the Roms' was published for nine months in 1973 in Belgrade (Dalbello 1989). Books in Romani, most of them poetry, were published in all of the capitals of the ex-Yugoslav republics and autonomous regions as well as in smaller towns such as Leskovac and Preševo. Classes in Romani were begun in Gilan (Gnjilane) and Ferizaj (Ferizovik, Uroševac) in Kosovia (Birlik 9.X.94, 14) and informal classes outside the regular school structure were also organized in the predominantly Romani Skopje suburb Šuto Orizari (Šutka), which was where a large number of Roms from Skopje and later elsewhere settled after the disastrous Skopje earthquake of 1963.[10] In 1977, Šaip Jusuf translated a book about Tito into Romani with significant press coverage (Nova Makedonija 77.09.28-30, 9; Jusuf 1978). It was the first non-periodical publication in Macedonia (and Yugoslavia) by a Rom for Roms. Shortly thereafter, a number of anthologies of Romani poetry and stories were published, mostly in

---

[10] According to the 1994 census (Antonovska *et al.* 1996), 48% of Macedonia's Romani population lives in the five Skopje municipalities, more than half of them in the municipality of Čair, where Šutka is located. The next largest concentrations are Prilep (8.2%), Kumanovo (7.1%), Tetovo (5.6%), Gostivar (4.9%), Bitola (3.9%), Štip (3.3%), Debar (2.5%), and Vinica (2%). In terms of proportions, Roms constitute 3.9% of the population in the five municipalities of Skopje, but 14.7% in Čair. Other relatively sizable proportions are Vinica (4.6%), Debar (4.3%), Prilep (3.8%), Berovo (3.3%), Štip (2.9%), Kičevo (2.6%), Delčevo (2.5%), Kumanovo (2.4%), Kočani (2.3%), and Kriva Palanka (2.2%).

bilingual editions (e.g. Djurić 1979; Dimić 1979; Krasnići 1981; Salijesor 1984; Balić *et al.* 1984; Dimić 1986; see also Dalbello 1989) as well as a Romani–Serbo-Croatian–English dictionary (Uhlik 1983). More than ten years after Jusuf (1978), however, Trajko Petrovski's (1989) translation of the pre-World War Two Macedonian poet Kočo Racin's collection *Beli Mugri* ('White mists') into Romani was still an unusual event.

Throughout this period, pedagogical materials were virtually nonexistent. The classes mentioned above were conducted without formal textbooks. Jusuf– Kepeski (1980), while it brought attention to the Romani language, did not function in an institutional context. Although translations, original belle lettres, folklore collections, and scholarly studies[11] appeared with increasing frequency, they did not change the educational situation. Cortiade (1984), published in an expanded version in Titograd (Podgorica) in 1986, was an attempt to formulate a transdialectal orthography that would serve as the basis of both literary communication and a literary language for use in schools. Although this orthography has been gaining increasing acceptance in Western Europe (see also Cortiade *et al.* 1991; Cortiade 1994), and was even the basis of a primer published in Sarajevo (Kurtiàde 1990) in a tirage of 2,000 with a teachers manual in a tirage of 1,000, this orthography has not had a significant impact on publications in Macedonia. Problems with this orthography will be discussed below (see also Friedman 1995).

In 1990, Yugoslavia along with the rest of Eastern Europe entered an era of political pluralism. The tragic results of the nationalist hijacking of that political process, a course of action that had already begun during the previous decade, are too well known to require further comment here.[12] The Romani contribution to the ethnopolitical movement, however, is not perceived as a challenge to the legitimacy of the Macedonian state and has been careful to insist on its loyalty to the Republic of Macedonia.[13] The Party for the Complete Emancipation of the Roms (Macedonian: *Partija za Celosna Emancipacija na Romite,* Romani: *Partija Saste Emancipacijake e Romengiri;* PCER or PSER) was founded on 12 August 1990

---

[11] An especially prolific scholar has been Rade Uhlik, who published a Serbo-Croatian–Romani dictionary as early as 1947 and had published Romani poetry even prior to that (Uhlik 1937; see also Dalbello 1989).

[12] Of the plethora of books that have sprung from this tragedy, Woodward (1995) is the first major scholarly work in English (see Hayden 1995) and gives a particularly clear analysis of the role of international involvement. Silber – Allen (1995) represents the best journalistic account in English in the opinion of many educated observers (Vesna Pusić, University of Zagreb, personal communication).

[13] This is in contrast to, e.g., some Albanian ethnopoliticians in Macedonia, who on occasion dispute the legitimacy of Macedonian statehood and even national identity (e.g. Xhaferi 1995).

with Faik Abdi (Faik Abdieskoro), a wealthy businessman from Shutka, as its leader.[14] The first item in the party program concerned education. Among the planks in the educational platform were the following (Predlog 1990):

1. Opening of pre-school/day-care centers in which children could be exposed to both Macedonian and Romani simultaneously
2. Elementary education in Macedonian with 2–3 hours a week Romani instruction
3. Founding of a Department of Romology at Skopje University, one which would contribute to the overall improvement of the condition of the Romani nationality
4. Post-graduate studies and exchanges with India
5. Romani children learn the history of their people, especially about Hitler's holocaust against the Roms during World War Two, which is to be made part of the facultative language classes 3–4 hours a week.

The third section of the party program was concerned with the Romani language, and called not only for minority language rights equal with other minority languages but also expressed concern for the normativization of Romani and close cooperation with institutions in India.

On 8 September 1991 a referendum was held concerning independence for the Republic of Macedonia. Material encouraging people to vote was printed in all six major languages of Macedonia including Romani.[15] It was perhaps the first time Romani appeared on the front page of the principal Macedonian daily newspaper, Nova Makedonija (91.09.08, 1).[16]

In Fall 1991, Faik Abdi sent a letter in the name of PCER to the rector of the University of Skopje demanding the opening of a Romani Studies Department but the establishment of such a Department remains a desideratum that has yet to be achieved. In early 1992, a group of Romani intellectuals formed a second Romani political party, the Democratic Progressive Party of the Roms in Macedonia, headed

---

[14] Like other ethnopolitical parties in Macedonia, the leadership of PCER denies that the party is "mononational" and points to the existence of party members belonging to other nationalities to demonstrate this (Čangova 1991). As with other ethnopolitical parties, however, the fact remains that the party's political concerns are focused on a specific ethnic group. In 1991 the party changed its name to Party for the Complete Emancipation of the Roms of Macedonia (Romani: *Partija Saste Emancipacijake e Romengiri tari Makedonija*, Macedonian: *Partija za Celosna Emancipacija na Romite na Makedonija*).

[15] The other five are Macedonian, Albanian, Turkish, Aromanian, and Serbian.

[16] The phrase was *Referendumeske '91. Va: sijam suvereno thaj korkorifundirimi Makedonijake* '[pertaining to] Referendum '91. Yes: I am for a sovereign and independent Macedonia' (Arli dialect). The advertisement was hexaglossic, but there were also monoglot Romani-language posters encouraging Roms to vote in the referendum.

by Bekir Arif. A major issue for the new party was increasing the pace of education-
al reforms (Nova Makedonija 21.10.1992, 4). There was also disagreement between
the two parties over questions of language standardization, dialectal compromise and
the place of Romani in educational institutions. Among the debated issues were
whether Romani should be a language of instruction (Macedonian: *nastaven jazik*) or
a language of study (Macedonian: *nastaven predmet*) and whether or not the standard
was to be based entirely on the Arli dialect, spoken by the majority or Roms in
Macedonia, or whether elements of other dialects should be included. Table 5 gives a
selection of diagnostic words illustrating some of the most salient phonological, mor-
phological, and lexical differences among the main dialects spoken in Macedonia.
Among the features illustrated are the following: palatal mutation of dentals before
stressed /i/ ('work'), treatment of inherited intervocalic retroflex *ṇd ('bread'),
palatalization and loss of /n/ intervocalically before stressed /i/ ('water'), formation of
the aorist ('I gave'), the shibboleth 'thus', intervocalic /s/ in grammatical suffixes
('with God'), shape and gender in the 3pl pronoun ('they'), and shape of the definite
article (nominative plural = masculine nominative singular vs nominative plural =
oblique singular and plural and distinct feminine nominative singular vs feminine
nominative singular = nominative and oblique plural).

Table 5
Examples of Romani dialectal differences

Burgudži	buci	maro	pani	diyom	kidjal	devlesa	on, ol	o-i-o-c
Arli	buti	maro	pani	dindjum	agjar	devlca	on, ola	o-i-o-c
Džambaz	buki	manro	pai	diyem	gčja	devlesa	von	o-i-c-c
Gurbet	buči	marno	pai	diyem	gaja	devlcha	von	o-c-c-c
gloss	*work*	*bread*	*water*	*I gave*	*thus*	*with God*	*they*	*the:*
								Ms-Fs-
								Np-Obl

It was shortly after this split in the Romani political scene occurred that, on Novem-
ber 20–21, 1992 the Ministry of Education of the Republic of Macedonia and the
Philological Faculty of the University of Skopje sponsored a conference for the
purpose of reaching an agreement concerning the introduction of Romani as a
course of study in Macedonian schools.[17] The conference was attended by a num-
ber of Macedonian Roms active in Romani intellectual life, including Šaip Jusuf,

---

[17] Concerned Macedonian intellectuals were already attempting to respond to the need for
Romani-language education by the summer of 1991, when Dimitar Mirčev, who was then a professor
of Sociology at the University of Skopje, discussed with me the possibility of holding a normativiza-
tion conference in Skopje.

Trajko Petrovski, Guneš Mustafa, Šaip Isen, Ramo Rušidovski, Tahir Nuhi, Iliaz Zendel, and others. Also present were Donald Kenrick and myself as well as members of the Philological Faculty of Skopje University and the Macedonian Academy of Sciences, most notably Olivera Jašar-Nasteva and Liljana Minova-Ġurkova as well as Živko Cvetkovski, head of the Macedonian Department. Representatives of the political factions were also present at the opening session, but the chair of the meeting deftly prevented the meeting from becoming a series of political speeches and expeditiously turned the conference into a language standardization working group.

It quickly emerged that the Roms present at this conference were not in favor of the establishment of Romani as a language of instruction in a parallel education system but rather the teaching of Romani as a subject in elementary schools and pedagogical academies with a view to preparing a cadre of teachers and ultimately a lectureship and Department of Romani at the University of Skopje. As mentioned above, one of the explicit goals of Romani politics in Macedonia is the establishment of such a Department, but a qualified cadre of faculty has yet to be trained.

The document that resulted from these deliberations, which was reproduced in full and analyzed in detail in Friedman (1995), was agreed upon by representatives of the various political currents as well as by the intellectuals that produced it. The document addresses a number of issues in Romani language standardization, e.g. the Arli dialect is specified as the base, with elements from other dialects being incorporated into it, and basic orthographic, morphophonological and morphological rules are specified in a series of twelve points. The document should be viewed in the context of Jusuf–Kepeski (1980), Kenrick (1981), and Cortiade *et al.* (1991). As indicated above, both Jusuf and Kenrick were present at the conference. Moreover, both Jusuf and Kenrick participated in the deliberations of the Language Commission at the Fourth World Romani Congress, at which Cortiade *et al.* (1991) was discussed and signed. Jusuf was a signatory to that document, but Kenrick was not. Mention should also be made here of Hancock (1975; 1993), which, while important for the history of Romani standardization, did not have a direct bearing on the 1992 conference. The former had been superseded by subsequent publications and events while the latter had not yet appeared.

Orthography has always been an issue for the standardization of Romani. Because efforts at Romani education have taken place in the context of the languages of other countries, as many orthographies have been used for Romani as there are standard languages with which is has been in contact. Although Romani in Cyrillic-using countries such as Russia and Bulgaria has been written in Cyrillic, a consensus has emerged to use a Latin based orthography as the most universally accessible (cf. Kyuchukov *et al.* 1995)—considerations which also influenced the

choice of alphabet for Albanian (see Skendi 1967, 366–90). In the case of Macedonia, which in the context of former Yugoslavia had an established bi-alphabetical tradition, Romani has always been written using a Latin orthography similar to that of Kenrick (1981), although Jusuf and Kepeski (1980) also use a Macedonian-based Cyrillic orthography for Romani in their Macedonian parallel text. At the 1992 Skopje conference, Macedonian Roms preferred to continue developing an orthography like that of the Second World Romani Congress (Kenrick 1981) rather than the Fourth.

Table 6 illustrates some of the salient differences between the Fourth World Romani Congress orthography and that of the 1992 Macedonian Conference.

Table 6
Comparison of current Romani orthographies

	Cortiade *et al.* (1991)	1992 Macedonian Conference	dialectal pronunciations	
Rom (loc. sg.)	Romesθe	Romeste	[romeste]	[romesće]
Rom (loc. pl.)	Romenθe	Romende	[romende]	[romende]
Rom (abl. sg.)	Romesθar	Romestar	[romestar]	
Rom (abl. pl.)	Romenθar	Romendar	[romendar]	
Rom (dat. sg.)	Romesqe	Romeske	[romesće]	[romeske]
Rom (dat. pl.)	Romenqe	Romenge	[romende]	[romenge]
done (pl. pt.)	kerde	kerde	[ćerde]	[kerde]
Rom (instr. sg.)	Romeça	Romesa	[Romea]	[Romesa]
you do (sg.)	keresa	keresa	[ćerea]	[keresa]

The treatment of underlying or historical dental and/or velar stops is an area of both considerable and salient dialectal variation and morphophonemic alternation in Romani. These phonemes can be pronounced as palatals and/or with affricated or fricativized articulation in various dialects of Macedonia and elsewhere (see Ventcel'–Čerenkov 1976 and Boretzky–Igla 1994 for details). Moreover, as seen in Table 6 (and mentioned earlier), some Romani dialects eliminate intervocalic /s/ in certain grammatical morphemes. Cortiade *et al.* (1991) articulates the principle of using underlying forms in most environments, but has special graphic symbols for the above mentioned morphophonemic alternations in their function as case markers (which Cortiade *et al.* 1991 treats as postpositions, but see Friedman 1991), viz. θ, q, and ç for dentals, velars, and /s/, respectively. Thus in the orthography of

Cortiade *et al.* (1991) the same morphophonemic alternations have different spellings, while the same graphic symbols have different pronunciations whose dialectal variations are represented by different letters in roots and verbal affixes on the one hand and in case affixes on the other, as illustrated in Table 6.

On 17 November 1993 the first issue of a Romani monthly newspaper, *Romani Sumnal/Romski Svet* 'Romani World', was published in Skopje under the editorial leadership of Oskar Mamut, who is also employed in the Romani-language division of Radio-Television Skopje.[18] The newspaper is bilingual, with all material in both Romani and Macedonian. The issue of the codification of a Romani standard language is explicitly addressed on the first page of the first number, where the editorial board states that one of the tasks they have set themselves is contributing to the development and use of literary Romani. As such, the paper can be taken as a measure of the progress and ongoing concerns of the standardization of Romani in the Republic of Macedonia. The role of the mass media is potentially of great importance in language standardization.

In its basic principles, *Romano Sumnal* represents a development in the direction described by the decisions reached at the 1992 Skopje conference and indicated in Jusuf–Kepeski (1980), namely an Arli base with elements from other dialects using a Latin orthography of the type in wide use in Eastern Europe, including Jusuf–Kepeski (1980), and recommended at the 1971 standardization conference (cf. also Hancock 1993; 1995). Nonetheless, specifics of the solutions reached by *Romano Sumnal* differ from those seen elsewhere (see Friedman 1997). Taken as a whole, *Romano Sumnal* clearly represents a step forward in the standardization of Romani in the Republic of Macedonia. The editors are aware of standardization issues and are attempting to make concrete contributions towards a consistent and usable norm.

Of particular importance to *Romano Sumnal* was the issue of education. Four articles were dedicated to the topic in the first issue (Bajramovska 1993; Mamut 1993; Darman 1993; Jašarov 1993). Darman (1993) speaks directly to the concerns of Romani parents for creating a home environment conducive to the success of children who are just beginning school. The other three articles are all critical of the fact that at the time they were written, Romani was still not a subject in any school curriculum in Macedonia, that the rate of educational success among Romani children is not showing any signs of increase, and that the few Romani intellectuals either hide their origins or bicker with one another rather than cooperating. While the very existence of a newspaper complaining about these conditions is itself

---

[18] Although the newspaper was intended as a monthly, it has so far appeared only thrice: 17 November 1993, 10 December 1993, and 1 April 1994.

something of a step forward, the fact remains that Romani education, like many other social programs, has not progressed with alacrity. Emilija Simoska (p.c. 8.XII.95), former Minister of Education, observed that education is functioning as a proxy for interethnic relations so not enough attention is being paid to curriculum.

Nonetheless, progress is being made. When in June–July 1994, at the behest of the International Conference on Former Yugoslavia (ICFY) and under the sponsorship of the Council of Europe, an extraordinary census was carried out in the Republic of Macedonia (see Friedman 1996a), Romani was one of the six official languages of census forms and documents, including the training manual for enumerators, in accordance with Article 35 of the census law. The published materials connected with the 1994 census represent the first official use of Romani in the Republic of Macedonia and were thus intimately connected with the standardization of Literary Romani in that country. The language of the census forms displays significant progress in the achievement of standardization and as such represents a significant development of Romani in official usage (Friedman 1996b).

In September 1995 Šaip Jusuf's Romani textbook for elementary schools was finally in press at Prosvetno Delo, the reviews having been completed in July of that year. Although Jusuf originally envisioned a series of textbooks beginning with grade one, the current book is intended for grade 3. The manuscript contained about a hundred texts, in poetry (24) and prose (80). The amount of material is about twice as much as can be covered in a year at two hours a week (i.e. 70 hours), and the question of norm versus dialect is not explicitly addressed. The final version also included a vocabulary and pages of pictures for stories. One problem with the review process was that the two reviewers with pedagogical expertise knew no Romani, while the Romani reviewer had no pedagogical experience. The Romani reviewer criticized the text for excessive Indicism (e.g., using *namaste* 'hail' (Hindi) instead of *sar sijan* 'how are you' or *šukar/lačho dive* 'good day', the Indicism *badali* instead of the colloquial Turkism *buluti* 'cloud'), but sometimes picked on dialectal details (suggesting *kanzavuri* for *kanauri* 'hedgehog', *farba* (from German) for *renki* (from Persian, probably via Turkish) 'color' (cf. Friedman 1989). However, he also caught orthographic inconsistencies, e.g. the need to treat the syllable *o-* in the 3sg. acc. pronoun as part of the stem: *ole* and not *o le* (as if *o* were the definite article). The textbook was officially published and announced to the public on 8 April 1996, but as of this writing (March 1997) it had not yet been released due to financial complications.

Meanwhile, developments in neighboring and other Balkan countries are taking place each independent of the other. In Albania (Kurtiàde 1994) the Fourth World Romani Congress orthography and pedagogical materials sponsored by the European commission are circulating, but it is unclear if any of them are in actual

use. Similarly, in Romania a pedagogical manual for teacher training complete with lessons using the Fourth World Romani Congress orthography (Cortiade *et al.* 1991) has been published (Sarău 1991; cf. also Sarău 1992) and European Commission-sponsored projects are being undertaken (Interface 15[8/94], 5), but so far Romani is only a language of study at the university level (Lemon: Romnet 95.XII.20). In Bulgaria some materials use an adaptation of Bulgarian Cyrillic for Romani (e.g. Malikov 1992), but an English-type Latin orthography using digraphs rather than diacritics is also in use (Kyuchukov et al. 1995), e.g. *sh=š, ch=č, chh/chsh=čh, j=dž, zh=ž, x=ks, h=x* or *h, y=j, ts=c, ph, th, kh, w*=schwa, and studies have been conducted for bilingual literacy (Kyuchukov 1995).[19] Romani is also being taught at the University of Sofia (1995–97) by Birgit Igla, a non-Rom specialist in Romani with extensive fieldwork experience in Balkan Romani. In Greece, various studies have been conducted and conferences held, but none of them are concerned with education of Romani children in Romani (Interface 18[5/95], 18; 13[2/94], 15–20; 8[11/92], 12). Rather, in accordance with Greece's assimilationist language policies towards its minorities (cf. Human Rights Watch 1994), the concern is with teaching Greek. Although former Yugoslavia was home to some of the most progressive Romani activities such as the first Romani summer school, which was held in Belgrade (Interface 16[11/94], 3), the war has resulted in the persecution of Roms living in the FRY and other Former-Serbo-Croatian speaking lands, and many have fled to Western Europe (see Interface 19[8/95], 20–2).

In conclusion we can say that while progress in Romani language education in the Republic of Macedonia has been slow, it has been made. Romani has gradually risen in status from total legal absence to legal equality with all other minority languages, even if the *de facto* realization has not yet met the *de jure* possibilities. The standardization of Romani and the fixing of the Arli dialectal base with other elements, together with a consistent orthography, has made significant progress from Jusuf–Kepeski (1980), to Romano Sumnal (1993) to the census (1994) to Jusuf's third grade textbook (1996). Similarly, the orthography conference of 1992 probably helped make actors aware of the need for consistency. While activity in the Republic of Macedonia has not been coordinated with that going on in neighboring countries or western Europe, it is endeavoring to meet the needs of the people for whom it is intended, and is certainly in advance of, e.g., Greece or FRY. Although politics is clearly playing a role, nonetheless, the essential issues remain pedagogical and normative. The introduction of Romani as a language of study at the elementary level has the potential to exert an enormous influence on the future codification of Romani both within the Republic of Macedonia and beyond its bor-

---

[19] Cf. also Hübschmannová *et al.* (1991), which uses a Czech based orthography.

ders. While Romani education in Macedonia is progressing slowly, it is nonetheless perceptibly progressing, and if other circumstances in the region allow, the future promises to be better than the past.

References

Abduramanoski, I. 1994. Veštačka delba na Romite. In: Nova Makedonija 30.X.94: 7.

Antonovska, S. *et al.* 1991. Broj i struktura na naselenieto vo Republika Makedonija po opštini i nacionalna pripadnost: Sostojba 31.03.1991 godina. Republički zavod za statistika, Skopje.

Antonovska, S. *et al.* 1994a. Statistički godišnik na Republika Makedonija 1993. Republički zavod za statistika, Skopje.

Antonovska, S. *et al.* 1994b. Popis '94: Podatoci za segašnosta i idninata. Prvi rezultati. Soopštenie 1. Zavod za statistika, Skopje.

Antonovska, S. *et al.* 1994c. Popis '94: Podatoci za segašnosta i idninata. Prvi rezultati. Soopštenie 2. Zavod za statistika, Skopje.

Antonovska, S. *et al.* 1996. Popis naselenieto, domak'instvata, stanovite i zemjodelskite stopanstva vo Republika Makedonija, 1994 godina, kn. 1, Naselenie po nacionalna pripadnost, veroispoved, majčin jazik i državjanstvo. Republički zavod za statistika, Skopje.

Bajramovska, V. 1993. Romani siklana—suno ja nane suno!? In: Romano Sumnal 17.XI.93: 4.

Balić, S. *et al.* 1984. Jaga—Vatre. Savet smotra kulturnih dostignicá roma SR Srbije, Leskovac.

Bárány, Z. 1994. Living on the edge: East European Roma in postcommunist politics. In: Slavic Review 53: 321–44.

Bárány, Z. 1995. The Roma in Macedonia: Ethnic politics and the marginal condition in a Balkan state. In: Ethnic and Racial Studies 18: 515–31.

Boretzky, N. – Igla, B. 1994. Wörterbuch Romani–Deutsch–Englisch: Mit einer Grammatik der Dialektvarianten. Harrassowitz, Wiesbaden.

Bugarski, R. 1992. Language in Yugoslavia: Situation, policy, planning. In: Bugarski, R. – Hawkesworth, C. (eds.): Language planning in Yugoslavia, 10—26. Slavica, Columbus.

Chalumeau, N. – Gualdaroni, R. 1995. France: The general and scholastic situation of the Gypsy community. In: Interface 18: 9—17.

Cortiade, M. 1984. Romani fonetika thaj lekhipa. Peskekerdo, Filăn Than (expanded and reprinted 1986 Pobjeda, Titograd).

Cortiade (Kurtiàde), M. 1990. Rromano startoro. Svjetlost, Sarajevo.

Cortiade, M. *et al.* 1991. I alfabèta e standardone Rromane ćhibaqiri, Décizia: I Rromani Alfabèta. In: Informaciaqoro Lil e Rromane Uniaqoro 1–2: 7–8.

Cortiade, M. 1994. Sikavipen sar te siklŏn e ćhavorre e širpustikaça. Centre National de Documentation Pedagogique de Midi-Pyrenées, Toulouse.

Čangova, K. 1991. Na Romite treba da im se pomogne: Razgovor so Faik Abdi, pretsedatel na PCERM. In: Nova Makedonia 15.X.95: 4.

Dalbello, M. 1989. Prilog bibliografiji o Romima (Ciganima) u SFR Jugoslaviji. In: Šipka, M. (ed.): Jezik i kultura Roma, 429–99. Insitut za proučavanje nacionalnih odnosa, Sarajevo.

Darman, L. 1993. E avgo sikljovneskere avgune hape. In: Romano Sumnal 17.XI.93: 6.

Dimić, T. 1979. Kana vavas ando foro—Dolazeći sa vašara. Irig, Novi Sad.

Dimić, T. 1986. Ćidimata e rromane đilendar—Romska narodna poezija. Matica Srpska, Novi Sad.

Djurić, R. 1979. Bi kheresko bi limoresko—Bez doma bez groba. Nolit, Belgrade.

Duijzings, G. 1992. De Egyptenaren in Kosovo en Macedonië. In: Amsterdams Sociologisch Tijdschrift 18(2): 24—38.

Duijzings, G. 1997. The making of the Egyptians in Kosovo and Macedonia. In: Vermeulen, H. – Govers, C. (eds): The politics of ethnic consciousness, St. Martin's, New York.

Friedman, V.A. 1985a. Balkan Romani modality and other Balkan languages. In: Folia Slavica 7: 381–9.

Friedman, V.A. 1985b. The sociolinguistics of Literary Macedonian. In: International Journal of the Sociology of Language 52: 31–57.

Friedman, V.A. 1985c. Problems in the codification of a Standard Romani literary language. In: Grumet, J. (ed.): Papers from the Fourth and Fifth Annual Meetings: Gypsy Lore Society, North American Chapter, 56–75. Gypsy Lore Society, New York.

Friedman, V.A. 1989. Toward defining the position of Turkisms in Romani. In: Šipka, M. (ed.): Jezik i kultura Roma, 251–67. Insitut za proučavanje nacionalnih odnosa, Sarajevo.

Friedman, V.A. 1991. Romani nominal inflection: Cases or postpositions? In: Studia gramatyczne 11 (=Problemy opisu gramatycznego języków słowiańskich): 57–64. Polish Academy of Sciences, Warsaw.

Friedman, V.A. 1995. Romani standardization and status in the Republic of Macedonia. In: Matras, Y. (ed.): Romani in contact: The history, structure, and sociology of a language, 203–17. Benjamins, Amsterdam.

Friedman, V.A. 1996a. Observing the observers: Language, ethnicity, and power in the 1994 Macedonian census and beyond. In: Rubin, B. (ed.): Toward comprehensive peace in Southeastern Europe: Conflict prevention in the South Balkans, 81–105/119–26. Council on Foreign Relations/Twentieth Century Fund, New York.

Friedman, V.A. 1996b. Romani and the census in the Republic of Macedonia. In: Journal of the Gypsy Lore Society (Fifth Series) 6: 89–101.

Friedman, V.A. 1997. Linguistic form and content in the Romani-language press. In: Matras, Y. – Bakker, P. – Kyuchukov, H. (eds.), The Typology and dialectology of Romani, 181–96. Benjamins, Amsterdam.

Friedman, V.A. – Dankoff, R. 1991. The earliest text in Balkan (Rumelian) Romani: A passage from Evliya Çelebi's Seyāḥatnāme. In: Journal of the Gypsy Lore Society (Fifth Series) 1: 1–20.

Hadži-Ristik, M. 1994. Prašanjeto za etnogenezata na Egipḱanite vo Makedonija. In: Nova Makedonija 13–17.IX.1994: 16, 16, 15, 22, 9.

Hancock, I. 1975. Problems in the creation of a standard dialect of Romanés. (= Social Sciences Research Council Working Papers in Sociolinguistics No. 25.) ERIC, Arlington VA.

Hancock, I. 1993. The emergence of a union dialect of North American Vlax Romani, and its implications for an international standard. In: International Journal of the Sociology of Language 99: 91–104.

Hancock, I. 1995. A handbook of Vlax Romani. Slavica, Columbus.

Haugen, E. 1966. Semicommunication: The language gap in Scandinavia. In: Sociological Inquiry 36: 280–97.

Haugen, E. 1968. The Scandinavian languages as cultural artifacts. In: Fishman, J. – Ferguson, C. – Das Gupta, J. (eds): Language problems of developing nations, 267–84. John Wiley & Sons, New York.

Hayden, R. 1995. Review of Woodward 1995. In: Slavic Review 54: 1114–5.

Hübschmannová, M. *et al.* 1991. Romsko-český a česko-romský kapesní slovník. Státní pedagogické nakladatelství, Prague.

Human Rights Watch/Helsinki. 1994. The Macedonians of Greece. Human Rights Watch, New York.

Interface. Information newsletter of the Gypsy Research Centre. Gypsy Research Centre, Université René Descartes, Clichy.

Jašar-Nasteva, O. 1992. Soziolinguistische Aspekte des makedonischen und der anderen Sprachen in der Republik Makedonien. In: Welt der Slaven 37: 188–210.

Jašar-Nasteva, O. 1997. A contribution to the study of the revival of the Aromanians of Macedonia (based on an autographed manuscript by G. Prličev). In: Balkanistica 10: 24–54.

Jašarov, S. 1993. Simfonijakoro orkestri, prima violina agorutne hevja ki bašalni. In: Romano Sumnal 17.XI.93: 7.

Jusuf, Š. 1996. Sikljova romani čhib. Prosvetno Delo, Skopje.

Jusuf, Š. – Kepeski, K. 1980. Romani gramatika—Romska gramatika. Naša Kniga, Skopje.

Jusuf, Š. *et al.* (transl.) 1978. Amen sam e Titoske, O Tito si amaro. (Mi smo Titovi, Tito je naš by Drago Zdunić *et al.*) Univerzum, Ljubljana.

Katona, L. 1969. Le dialecte turc de la Macédoine de l'Ouest. In: Türk dili araştırmaları yıllığı belleten, 57–194.

Kenrick, D. 1981. Romano alfabeto. In: Loli Phabaj 1: 3–4.

Kochanowski, J. 1963. Gypsy studies, Part 1. International Academy of Indian Culture, New Delhi.

Koneski, B. 1950. Za padežite. In: Makedonski jazik 1: 165–72.

Kopitar, J. 1829. Albanische, walachische und bulgarische Sprache. In: Jahrbücher der Literatur 46: 59–106.

Kostov, K. 1973. Zur Bedeutung des Zigeunerischen für die Erforschung grammatischer Interferenzscheinungen. In: Balkansko ezikoznanie 14: 99–113.

Krasnići, A. 1981. Čergarendje Jaga—Čergarske Vatre. Edinstvo, Prishtina.

Kurtiàde 1994 = see Cortiade 1994.

Kyuchukov, H. 1995. Romany children and their preparation for literacy: A case study. Tilburg University Press, Tilburg.

Kyuchukov, H. *et al.* 1995. Romani alfabeta—Romska azbuka. Bulvest-2000, Sofia.

Latifić, I. *et al.* 1970. Vitalni, etnička i migraciona obeležja. Popis stanovništva 1961. Savezni zavod za statistiku, Belgrade.

Ljubisavljević, M. 1990. Nismo ptice selice. In: Svet 9.218: 34—35.

Lunt, H. 1952. A grammar of the Macedonian literary language. Državno knigoizdatelstvo, Skopje.

Lunt, H. 1984. Some sociolinguistic aspects of Macedonian and Bulgarian. In: Stolz, B. – Titunik, I. – Doležel, L. (eds.): Language and literary theory: In honor of Ladislav Matejka (= Papers in Slavic Philology 5), 83–132. University of Michigan, Ann Arbor.

Mamut, O. 1993. Soske o sikle stagnirinena!? In: Romano Sumnal 17.XI.93: 5.

Malikov, J. 1992. Cigansko-bălgarski rečnik. Fondacija Otvoreno obštestvo, Sofia.

Matras, Y. 1994. Structural Balkanisms in Romani. In: Reiter, N. (ed.): Sprachlicher Standard und Substandard in Südosteuropa und Osteuropa, 195–210. Harrassowitz, Wiesbaden.

Pekevski, B. *et al.* 1973. Naselenie po narodnost vo SR Makedonija. Popis na naselenieto i stanovite, 1971 god. Republički zavod za statistika, Skopje.

Petrović, R. 1992. Demografske osobenosti Roma u Jugoslavii. In: Mačura, F. (ed.): Razvitak Roma u Jugoslavii, 115–27. SANU, Beograd.

Petrovski, T. (transl.) 1989. Parne detharina (Beli mugri by Kočo Racin). Naša kniga, Skope.

Predlog. 1990. Programa na Partijata za celosna emancipacija na Romite od SR Makedonija. MS. 9pp.

Puxon, G. 1979. Romanès and language policy in Jugoslavia. In: International Journal of the Sociology of Language 19: 83–90.

Risteski, S. 1991. Narodni prikazni, predanija i običai kaj egipḱanite/eǵupcite vo Makedonija. Nezavisni izdanija, Ohrid.

Salijesor, S. 1984. Bahtale dive atar mire sune. Univerziteti e bućengor, Preševo.

Sarău, G. 1991. Limba romaní. Editura didactică și pedagogică, Bucharest.

Sarău, G 1992. Mic dicționar rom-român. Kriterion, Bucharest.

Savezni zavod za statistiku. 1954. Stanovništvo po narodnosti (Knj. 9): Konačni rezultati popisa stanovništva od 15 marta 1948 godine. Savezni zavod za statistiku, Belgrade.

Savezni zavod za statistiku. 1981. Statistički bilten broj 1295 (1981). Popis stanovništva, domaćinstva i stanova u 1981 godini: Nacionalni sastav stanovništva po opštinama. Savezni zavod za statistiku, Belgrade.

Savezni zavod za statistiku. 1988. Statistički Godišnjak Jugoslavije (Knj. 35). Savezni zavod za statistiku, Belgrade.

Silber, L. – Little, A. 1995. The death of Yugoslavia. Penguin, London.

Silverman, C. 1995a. Persecution and politicization: Roma (Gypsies) of Eastern Europe. In: Cultural Survival Quarterly 19: 43–9.

Silverman, C. 1995b. Roma of Shuto Orizari, Macedonia: Class, politics, and community. In: Kideckel, D. (ed.): East-Central European communities: the struggle for balance in turbulent times, 197–216. Westview Press, Boulder.

Skendi, S. 1967. The Albanian national awakening. Princeton University Press, Princeton.

Stanković, V. 1992. Romi u svetlu podataka jugoslovenske statistike. In: Mačura, F. (ed.), Razvitak Roma u Jugoslavii, 159—79. SANU, Beograd.

Škiljan, D. 1992. Standard languages in Yugoslavia. In: Bugarski, R. – Hawkesworth, C. (eds.): Language planning in Yugoslavia, 27–42. Slavica, Columbus.

Tanasković, D. 1992. The planning of Turkish as a minority language in Yugoslavia. In: Bugarski, R. – Hawkesworth, C. (eds.): Language planning in Yugoslavia, 140–61. Slavica, Columbus.

Uhlik, R. 1937. Romane gjilja—Ciganske pesme. Vučen Štrbac, Prijedor.

Uhlik, R. 1947. Srpskohrvatsko–ciganski rečnik (romane alava). Svjetlost, Sarajevo.

Uhlik, R. 1983. Srpskohrvatsko–Romski–Engleski rječnik (Romengo alavari). Svjetlost, Sarajevo.

Ventcel', T.V. – Čerenkov, L.N. 1976. Dialekty ciganskogo jazyka. In: Konrad, N. I. (ed.): Jazyki Azii i Afriki I, 283L. N. 332. Nauka, Moscow.

Woodward, S.L. 1995. Balkan tragedy. The Brookings Institution, Washington DC.

Xhaferi, A. 1995. Regjistrimi i jashtëzakonshëm i popullsisë '94. In: Tribuna-Sh 2(23/24): 12–3.

Zemon, R. (ed.) 1996. Zbornik na trudovi za etnogenezata na Egipḱanite vo Makedonija. Logos-T, Skopje.

Address of the author: Victor A. Friedman
                          The University of Chicago
                          Dept. of Slavic Languages and Literatures
                          1130 East 59th Street
                          Chicago, Illionis 60637
                          e-mail: vfriedm@midway.uchicago.edu

INFORMATION FOR CONTRIBUTORS

Please provide typewritten double-spaced manuscripts with wide margins (50 characters per line, 25 lines per page). Type on one side of the paper only and number all the pages consecutively. The following information should be provided on the first page: the title of the article (or, in case of reviews, the title of the book under discussion, the name of the publisher, the place and date of publication as well as the number of pages), the author's full name and address and an abbreviated title of the paper (not exceeding 45 characters including spaces). Only original papers will be published and a copy of the Publishing Agreement will be sent to the authors of papers accepted for publication. Manuscripts will be processed only after receiving the signed copy of the agreement. Authors are requested to send two hard copies of their manuscript + a floppy disk under DOS.

The manuscript should be accompanied by an abstract of about 100 words. It should be typed on a separate sheet of paper. Tables, diagrams and illustrations (with the author's name and an Arabic number) should be presented on separate sheets. Captions should be typed on a separate sheet and placed at the end of the manuscript. Footnotes should be typed double-spaced on a separate sheet at the end of the article. All language examples (and only these) are to be italicized (single underlining). Citations in the text should be enclosed in double quotation marks (" " in case of a paper written in English, „ " in German and « » in French).

Reference to a publication should be made by the name of the author, the year of publication and, when necessary, by page numbers in the following ways:

. . . as described by Schmidt (1967) . . .
. . . as referred to by Hoover (1967, 56 78; 1976, 43).
. . . mentioned by several authors (Elgar 1978, 77; Williams 1981, 154 6) . . .

An alphabetically arranged list of references should be presented at the end of the article as follows:

Bárczi, G. 1958a. Magyar hangtörténet [The history of Hungarian sounds]. Akadémiai Kiadó, Budapest.

Bárczi, G. 1958b. A szótövek [Word stems]. Akadémiai Kiadó, Budapest.

Lakoff, G. Peters, P.S. 1969. Phrasal conjunction and symmetric predicates. In: Reibel, D. Schane, S. (eds): Modern studies in English, 113 41. Prentice-Hall, Englewood Cliffs.

Leben, W.R. 1980. A metrical analysis of length. In: Linguistic Inquiry 11: 497 509.

Ross, J.R. 1967. Constraints on variables in syntax. Ph.D. dissertation. MIT, Cambridge MA.

For marking subsections decimal notation should be applied. Do not use more than four digits if possible.

Examples within the text should be marked in italics. Meanings are to be rendered between inverted commas (' '). If glosses are given morpheme by morpheme, the initial letter of the gloss should be placed exactly below that of the example. Grammatical morphemes can be abbreviated in small case letters connected to the stem or the other morphemes by a hyphen. No period should be applied in such abbreviation. For example:

(1) (a)   A   sólymaid   elszálltak
            the falcon-gen-pl-2sg away-flew-3pl
            'Your falcons have flown away.'

Examples can be referred to in the text as (1a), (1a d), etc.

One proof will be sent to the author. Please read it carefully and return it by air mail to the editor within one week.

CONTENTS

ACTA LINGUISTICA HUNGARICA

VOLUME 46   1999

CONTENTS