

Yaron Matras and Anton Tenser

Complementizers in Romani

1 Introduction

1.1 The language

Romani is the only Indo-Aryan language that has been spoken exclusively in Europe (and by European emigrants to the New World) since the Middle Ages. Historical reconstruction and analysis of loanword layers, coupled with the social-ethnographic profile of the speaker population and comparisons with similar Indian diaspora populations, suggest that the language may have been brought into Anatolia under Byzantine rule sometime around the eleventh century by a caste-like population specialising in itinerant services such as metalwork and entertainment. Today, Romani constitutes one of the largest minority languages of Europe, with upwards of 3.5 million speakers residing mainly in southeastern Europe (Romania, Bulgaria, Serbia, Macedonia, Greece, Albania) as well as in central Europe (Slovakia, Hungary). There are also sizeable Romani-speaking communities in the United States and in South America as well as in urban centres of western Europe. In the western fringe regions of Europe – Scandinavia, Britain, and the Iberian Peninsula – Romani populations abandoned their language toward the beginning of the nineteenth century or even earlier, but they maintain an in-group vocabulary of Romani origin, which typically comprises around 400–600 lexical roots. All Romani speakers are bilingual and often multi-lingual from a very young age. The language has traditionally been limited to oral usage, primarily within the extended family and with neighbouring clans. Codeswitching and language mixing are common and dialects of Romani have absorbed considerable structural and lexical influences from their respective contact languages.

The lexicon of most Romani dialects contains a shared inventory of only around 1,000–1,200 lexical roots. Of those, only around 800 are pre-European, consisting of a core vocabulary of around 600 Indo-Aryan roots along with layers of early loans from Iranian languages and from Armenian (as well as other Caucasian languages). All Romani dialects also contain a significant layer of Greek loans, as well as a much smaller number of shared lexical items of Balkan Slavic and Balkan Romance origin. The remainder of the vocabulary is typically derived from subsequent contact languages, and thus differs among the individual dialects of the language.

Romani morphology is partly fusional and partly agglutinative. The fusional element tends to be older and is best represented by the person conjugation on the verb, especially in the present tense, and by the retention of declension classes in the first layer of nominal inflection. The agglutinative element is generally younger and is best represented by second-layer case-markers, which are enclitic and not sensitive to declension class. Some tense and modality markers as well as verb derivational markers indicating transitivity and intransitivity are also agglutinative. On the whole the language can be said to be drifting toward stronger reliance on analytical structures. This is in part a contact phenomenon. The Romani dialects of western Europe have largely abandoned flexional derivations in the verb and rely instead on analytical constructions with auxiliaries. In southeastern Europe (Albania, Macedonia, and Greece), Romani dialects have developed analytical perfect tenses, while in central and eastern Europe medio-passive constructions tend to rely on the active verb with a reflexive pronoun. Inflectional case markers also show strong competition from prepositions, though the extent varies among the dialects. An outstanding characteristic feature of Romani morphology is the division between pre-European vocabulary and European loanwords in both nominal and verbal morphology. In nouns, the difference is expressed primarily in the choice of declension class, which determines the expression of the nominative and partly also the oblique form. In verbs, European loans require a loan-verb adaptation marker, which functions essentially as a derivational marker that mediates between the verb stem and its tense-aspect and person inflection.

Romani generally displays the properties of a SVO language with a flexible, pragmatically determined word order. Relative clauses follow their head nouns and contain resumptive pronouns wherever there is reference to the head in a non-subject role. Pro-dropping occurs especially in the chaining of subjects, and in this respect Romani is a pro-drop language to a limited extent. Clauses are generally finite and subordinations are introduced by conjunctions, which are typically derived from interrogatives or local relations expressions, or borrowed.

Geographical dispersion and the influence of diverse contact languages have led to the emergence of considerable difference among the dialects of Romani. A dense cluster of isoglosses, referred to as the Great Divide (Matras 2005), separates Romani dialects on both sides of what was, during the relevant period, the border zone (and war zone) between the Austrian and Ottoman empires. Once speakers adjust to a monolingual mode of discourse, it is generally possible for Romani speakers east of the Great Divide (from Greece to the Baltics) to understand one another, while the (much smaller) population of speakers to the west of the Great Divide speak dialects that are more fragmented.

Traditionally an oral language of the household, Romani is increasingly being used in electronic communication among community activists, who practice a pluralistic approach to codification and the selection of dialect features. Some programmes to promote the language in education and media are underway, though there is no standard written form of the language, nor a prestige variety that would lend itself for the creation of such a standard.

1.2 Methodological remarks

Functionalist typology is based on the assumption that structural configurations that are language-specific serve to encode communicative needs that are universal and which are grounded in cognitive processes of conceptualization. The methodological challenge that arises in this perspective is how to capture structural categories in a way that would do justice both to the way they are arranged in the individual language under discussion and in others that bear similarities to it (and in that respect may be said to share a ‘type’), and to the universal conceptual meaning that they represent. Complements are generally understood to be propositions that are conceptually linked to a predicate, and which, structurally, are expressed at the sentential level as arguments of a verb (cf. Givón 1990: 515). The functionalist approach to complementation seeks to explore how the conceptual relationship between the verb and its complement proposition is linked to the structural relationship between them. Cristofaro (2003: 95–154) for instance, for a cross-linguistic sample of languages, maps a series of structural features (such as the presence/absence of TAM distinctions, case-marking, and person-agreement) to the semantic characteristics of predicates (such as modals, phasals, manipulative, desideratives, perception, utterance, and more). The cross-linguistic comparison yields implicational hierarchies of the likelihood of certain structural devices to correlate with certain semantic relations. Boye (2012) relates this kind of continuum to the notion of knowledge or ‘epistemology’: certain meanings require justification through explicit knowledge, or epistemic justification. Evidential and modal procedures provide various degrees of support for the transmission of knowledge. The intensity or strength of such support relates to the (universal) need to provide support. The structural system that provides the cues for conceptual support is the epistemic morphosyntactic system of a language. ‘Epistemic’ in Boye’s (2012) terminology is thus the use of explicit (morphosyntactic) devices to provide support for those conceptual relations or elements of knowledge that require such support (epistemic or modal relations).

In the approach that we take below we follow a similar principle: we accept that structural devices convey instructions on how to process knowledge and

links between information chunks, including the conceptual links between verbs and their complements. We accept that such links are not of equal value, and that some require more (epistemic) support than others. We further accept that there is reason to anticipate a correlation between the distribution of structural devices that mark complementation, and the continuum of types of semantic relations, and in particular between the employment of explicit epistemic devices and the need to provide strong support. Since, however, we are dealing with dialects of a single language and thus with a limited set of morphosyntactic devices involved in complementation, we allow ourselves to focus on the principal parameters of the semantic-conceptual continuum that are of direct relevance to the category distinctions found in the relevant data. We therefore adopt Givón's (1990) notion of a continuum that represents the degree of 'semantic integration' between the two propositions in a complex clause (in our case: the main predicate and its complement). Semantic integration refers to the degree of independence of the two propositions, in terms of the cognitive processing of their content. In the centre of the semantic integration continuum are two main parameters. The first is the likelihood that the information conveyed by the complement is based on secure or real-world knowledge. Predicates of utterance and perception, for instance, are on the 'factual' side of the continuum: their complements are more likely to represent factual reality that can be confirmed independently of the event to which they are linked as complements. For this reason, they represent a lower degree of semantic integration, or a lower degree of dependency on the main predicate. Complements of phasal or desiderative – or non-factual – verbs, on the other hand, owe their factual realization to that of their main predicates (i.e. the likelihood of the phasal or desiderative action to yield the intended results). They are thus highly dependent on their main predicates, and so they represent tight semantic integration between main verb and complement. The second parameter is the degree of agentive control. Like factuality, agent control is a factor in the likelihood that the proposition conveyed by the complement will be realized. With desiderative verbs, for instance, agent identity across the two propositions is a strong guarantee for the likelihood of the action that is conveyed by the complement to be realized since it is dependent on the same conditions as the main predicate; semantic integration is therefore tight. By contrast, agent differentiation (manipulation) requires a weaker degree of control, representing a stronger potential for event independence and therefore weaker semantic integration. The aspect of control is not limited to complements. Predicates that are conceptually linked in a final or purpose construction offer a mixture of the two parameters in which independent predicates are linked through a non-factual relationship. For that reason, we include in our discussion also the occasional reference to the structure of purpose clauses in Romani.

1.3 An overview of complementizers in Romani

Complementation in this work is described separately from subordination following the structure of the Romani Morpho-Syntactic (RMS) database (<http://romani.humanities.manchester.ac.uk/rms>) and the typology of Romani complements along the factuality continuum, as proposed by Matras (1999: 18–20, 2002: 179–185, 2004). The two contrasting poles in this continuum are the non-factual predication (as in ‘I want to go home’) and factual predication (as in ‘I know that he went home’), which together reflect the variation in the independent truth-value of the complement clause. The former (non-factual) accompanies modality verbs and is usually characterized in Romani by the complement particle *te*. The latter is best characterized as complementation with cognition-utterance verbs (cf. Givón 1990: 517). It accompanies factual complement phrases and is commonly expressed in Romani by the complementizer *kaj*. In the middle of this continuum lie the manipulation and purpose clauses, which are more ambiguous with respect to their independent truth-value, and for which the clause-linking strategies are subject to greater variation across dialects.

While the complementizer *te* is fairly stable across most Romani dialects, the functional slot that is represented in some dialects by inherited *kaj* is often filled by a borrowing from a contact language. The semantic functionality of the factual complementizer remains intact, however, even when the complementizer is replaced by a borrowing, and it is therefore customary in the context of Romani linguistics to refer to the *KAJ*-type complementizer (see Matras 2002, 2004). The distinction between *te* and *kaj* type complementation in Romani can be illustrated by the following examples (all data are taken from the Romani Morpho-Syntactic (RMS) database, based on original fieldwork carried out by the Manchester Romani Project: <http://romani.humanities.manchester.ac.uk/rms>. Sample numbers refer to RMS sample codes, detailed descriptions of which are fully accessible online, accompanied by audio files and transcriptions.

- (1) *me kam-om [jekvar amerika te dž-au]*
 I want-1SG once America COMP go-1SG
 ‘I want to visit America someday.’
 (Čuxny, Estonia, EST-005)
- (2) *šundž-om [kaj inne romovja tiš ak beš-en]*
 heard-1SG COMP other Roma also here live-3PL
 ‘I heard that other Roma live here as well.’
 (Bergitka, Poland, PL-007)

In the intermediate position of the independent truth-value continuum lie manipulation and purpose clauses. Their independent truth-value is more ambivalent,

and is dependent on the factor of agent control. Since control is a gradient, it is not surprising that we see greater variation in the forms of complementizers both across and within individual dialects. With the highest degree of agent control we often find the simple complementizer *te*; cf. example (3).

- (3) *ov čjind-a nev-e furjavipe [te šaj dž-al ando foro]*
 he bought-3SG new-PL clothes **COMP** can go-3SG to town
 ‘He bought new clothes so that he could go into town.’
 (Gurbet, Croatia, HR-001)

With lower degrees of agent control we often find duplex or complex forms, where *te* appears alongside a ‘reinforcer’ or ‘reinforcers’ (Matras 2002: 181). One of the common reinforcers is in fact the factual complementizer *kaj*. Examples can be found in purpose and manipulation clauses:

- (4) *phutregj-om e fereastră [kaj te aštik alear-ap tut]*
 opened-1SG ART window **COMP COMP** can hear-1SG you.OBL
 ‘I opened the window so that I could hear you.’
 (Keremidarea, Romania, RO-025)

- (5) *mje manglj-em la [kaj oj the žja-l tar]*
 I asked-1SG her.OBL **COMP** she **COMP** go-3SG away
 ‘I asked her to go away.’
 (Kishinjovcy, Ukraine, UKR-007)

The present discussion will focus on the types of structures mentioned above in the context of the integration continuum, and will include modal, factual, and manipulation complements as well as occasional reference to purpose constructions. Indirect questions will not be discussed here, but for the sake of completeness it should be mentioned that in Romani these are realized through the use of the regular stock of interrogatives, without the use of complementizers.

- (6) *phen man-ge [so kerd-al adjejs]*
 tell me-DAT **what** did-2SG today
 ‘Tell me what you have done today!’
 (Czech Vlax, Czech, CZ-001)

- (7) *jou puht-as [kōn jēl-o fōros]*
 he asked-3SG **who** went-PTCP.M.SG town
 ‘He asked who went to town.’
 (East Finnish Romani, FIN-002)

There are no examples in the RMS database of the minimal pairs of the type “I know that...” vs. “I know whether...”. Embeddings with the potential conditional

‘whether’ usually align themselves with modality (non-factual) clauses and are introduced by the non-factual complementizer *te*:

- (8) *buč-om o profesar-es [te v-el-a pe bijav]*
 asked-1SG ART teacher-OBL **COMP** come-3SG-FUT to wedding
 ‘I asked the teacher whether he will come to the wedding.’
 (Sinti, Romania, RO-022)

The interpretation of the example in (8) as ‘I asked the teacher *to* come to the wedding’ is precluded by the use of the non-subjunctive form of the verb *v-* ‘come’.

The position of an interrogative element in the complement is often occupied by loan elements from the contact languages. For example, dialects in contact with Russian often use the Russian interrogative-conditional particle *li*, those in contact with Polish often use the Polish *czy*, and many dialects of the Balkans use the South Slavic *dali*:

- (9) *me phučlj-om sykljaribnas-tar [av-ela li vov po bjav]*
 I asked-1SG teacher-ABL come-3SG **COND** he to wedding
 ‘I asked the teacher whether he will come to the wedding.’
 (Servi, Ukraine, UKR-004)
- (10) *phučl-om pes te ranjen-dar [či av-ela pe vera]*
 asked-1SG REFL ART teachers-ABL **COND** come-3SG to wedding
 ‘I asked the teacher whether he will come to the wedding.’
 (Bergitka, Poland, PL-007)
- (11) *phučh-um e nastavniko [dali kam av-ol ko bjav]*
 asked-1SG ART teacher **COND** FUT come-3SG to wedding
 ‘I asked the teacher whether he will come to the wedding.’
 (Arli, Serbia, YU-011)

An example of a minimal pair, where *te* and *kaj* might be used with a knowledge verb would be the non-factual *me džinav te rakirav* (‘I know how to speak’) vs. the factual *me džinav kaj jov rakirel* (‘I know that he speaks’).

Considering the dialect diversity of Romani and its geographical dispersion, we will mostly discuss tendencies rather than hard-set rules. The general trends will be exemplified using data samples from various dialects, in order to provide an appreciation of the diversity of Romani. The exceptions to the tendencies will be noted and exemplified. When a certain phenomenon is stated as a tendency and exemplified using data samples from only a few dialects, the reader should assume that the same phenomenon exists in most Romani dialects and that more extensive cross-dialectal exemplification is avoided due to the considerations of space.

1.4 Other aspects of complementation

1.4.1 Non-agreement

Romani is a predominantly finite language, in the sense that subordinate verbs agree with their subjects (Matras, 1999: 17). The overwhelming tendency across dialects is for the complement clauses to be finite, which is also a prominent feature of the Balkan languages. In the modal complement, the verb is marked for person and number:

- (12) *me kam-av* [*te sov-av*]
 I want-1SG **COMP** sleep-1SG
 'I want that I sleep.'
 (General)

There are some exceptions, however. The neutralization of person agreement in complements has been referred to as 'the new infinitive' (Boretzky 1996). This, essentially 'de-balkanization' process (Matras 2002: 161) has affected the Sinti (German) and Finnish Romani dialects, the dialects of Poland, Bohemia and Slovakia, as well as the Romungro dialect of Hungary. The most frequently used 'infinitive' forms in Romani are the historical present tense personal markers: 2SG in *-(e)s* and 3SG in *-(e)l*:

- (13) *jou pyryjlä* [*ceer-es putti*]
 he begins **do-INF** work
 'He begins to work.'
 (East Finnish Romani, FIN-008)

- (14) *kamj-om* [*te dža-l khere*]
 wanted-1SG **COMP go-INF** home
 'I wanted to go home.'
 (Polish Xaladytka, Poland, PL-014)

- (15) *ljubin-av* [*rano te vypi-et jikh šjtekla kali*]
 like-1SG morning **COMP drink-INF** INDF cup coffee
 'I like to have a cup of coffee in the morning.'
 (Polska Roma, Poland, PL-018)

- (16) *na kam-av* [*te dža-l andro foros*]
 NEG want-1SG **COMP go-INF** to town
 'I do not want to go to town.'
 (East Slovak, Slovakia, SK-002)

Eastern Romungro dialect of Slovakia uses a historical 2/3PL *-(e)n* form as the 'new infinitive':

- (17) *na kam-es [te dža-n]*
 NEG want-2SG COMP **go-*INF***
 ‘You do not want to go.’
 (Eastern Romungro, Slovakia, SK-027)

A remarkable development is found in the samples of the Ukrainian Servi and Xandžari speakers, where the infinitive form does not match any personal verb forms. Instead it looks like, on the surface, a present subjunctive form of the verb without the final consonant:

- (18) *vov ačh-el [te tire-Ø buti]*
 he begin-3SG COMP **do** work
 ‘He begins to work.’
 (Servi, Ukraine, UKR-004)
- (19) *tu-te vydž-ala [te dža-Ø man-sa]?*
 you-LOC can-3SG COMP **go** me-INS
 ‘Can you come with me?’
 (Kubanski Servi, Ukraine, UKR-008)

The most likely development path for this verb form is through the 2SG verb marker *-(e)h*, e.g. *tir-eh* → *tir-eØ*. Synchronically, it is clear that grammaticalization has been completed, since the 2SG marker in the Servi dialect is fully pronounced, and even velarized: *tir-ex* ‘you do’. Thus, in this case one can speak of the emergence of a true infinitive form.

1.4.2 Tense, aspect, modality

Modal (non-factual) complement verbs in Romani almost always appear in the subjunctive:

- (20) *me but mangl-em [te ža-v ande Indija]*
 I very wanted-1SG COMP **go-1SG.SBJV** to India
 ‘I wanted very much to go to India.’
 (Sofia Erli, Bulgaria, BG-024)

There are some exceptions, however. The Romacilikanes dialect of Greece shows agreement between the modal verb and the complement verb in tense, aspect and modality – here using the remote past which has indicative imperfect/habitual reading in the given context:

- (21) *panda kam-am-as [te dža-v-as ti Indija]*
 always want-1SG-REM COMP **go-1SG-REM** to India
 ‘I have always wanted to go to India.’
 (Romacilikanes, Greece, GR-002)

In other Romani dialects, the above phrase shows the complement verb form in the subjunctive:

- (22) *me uvek mang-av-as [te dža-v ki India]*
 I always want-1SG-REM COMP **go-1SG.SBJV** to India
 'I have always wanted to go to India.'
 (Kovacki, Macedonia, MK-004)

With inherited modal verbs of ability, which are impersonal and do not carry tense markers, it is the verb of the complement clause that carries tense marking:

- (23) *ame našy [∅ arakhlj-am rupun-i angrusti]*
 we cannot ∅ **found-1PL** silver-F ring
 'We couldn't find the silver ring'
 (Crimean, Ukraine, UKR-001)
- (24) *me naštik [te putard-em o vudar]*
 I cannot COMP **opened-1SG** ART door
 'I couldn't open the door.'
 (Gurbet, Serbia, YU-002)

A rare strategy to mark the tense of the action, found in the Ukrainian Gimpeny dialect sample, is to use the 3rd person past tense copula to form the past tense of the impersonal modal of ability:

- (25) *me našty sis [te rakh-as e jangrusti]*
 we cannot **were.3SG/PL** COMP find-1PL.SBJV ART ring
 'We couldn't find the ring.'
 (Gimpeny, Ukraine, UKR-020)

2 A descriptive account of Romani complementizers

2.1 Semantic aspects

Romani follows a general areal trend found in the languages of Balkans, where complements, including same-subject modal complements, are generally finite, and a semantic opposition between non-factual and factual complementizers is indicated through the identity of the complementizer. In Romani, the inherited and thus most widespread forms of the respective complementizers are *te* for the modal/non-factual, and *kaj* for the factual. Since the semantic-functional distinction tends to be maintained in the dialects even where the forms have undergone

structural change through substitution or borrowing, we can speak of complementizers of the type *TE* and *KAJ*, with individual formal representations in the individual dialects.

The inherited Romani modal complementizer *te* is realized in some dialects as *ti* (Sinti Romani) or *ta* (Lombard Italian Romani). The form is extremely stable across the dialects, and is rarely replaced by a borrowing. The one exception found in the samples is from the Xoraxani dialect, which borrows Bulgarian *da*:

- (26) *bejkim* [da d-av ka-le romes-te duj-trin džigares]
 can.1SG COMP give-1SG this-OBL Gypsy.man-LOC two-three cigarettes
 ‘I can give this Gypsy man some cigarettes.’
 (Xoraxani, Bulgaria, BG-015)

The complementizer *te* is used in Romani consistently with the usual stock of modals: volition (‘want’), positive and negative ability (‘can’, ‘cannot’), inception and termination (‘begin’, ‘start’, ‘end’, ‘finish’), obligation (‘have to’, ‘need’, ‘must’) and attempt (‘try’), as well as other predicates, such as ‘expect’, ‘like’, ‘fear’, ‘know how’, ‘dare’ and ‘have the strength’ among others:

- (27) *na kam-eha [te dž-a-s ko čjüteti]*
 NEG want-2SG COMP go-2SG.SBJV to town
 ‘You do not want to go to town.’
 (Mečkari, Albania, AL-001)

- (28) *na mogindž-am [dava nikaj te rakh-as]*
 NEG could-1PL this nowhere COMP find-1PL.SBJV
 ‘We couldn’t find it anywhere.’
 (Polish Xaladytka, Poland, PL-014)

- (29) *severim [te pj-av ek čaša kaffa sabajlen]*
 like.1SG COMP drink-1SG INDF cup coffee morning
 ‘I like to have a cup of coffee in the morning.’
 (Kalajdži, Bulgaria, BG-009)

- (30) *na has la zor [te dž-a-l pale khere]*
 NEG was.3SG her.OBL strength COMP go-3SG back home
 ‘She did not have the strength to walk back home.’
 (East Slovak, Slovakia, SK-011)

The emphasis of *te*-complements is on non-factuality, which is captured in the unconfirmed truth-value of the embedded proposition. Thus, even a verb that may, on the face of things, convey perception, but whose object cannot be confirmed in its factuality, will trigger the use of non-factual/modal *te* as the complement initiator:

- (31) *džaker-ava m-e čhav-es [te av-el kate sahati]*
 wait-1SG my-OBL son-OBL **COMP** come-3SG here soon
 'I expect my son to come here soon.'
 (Romacilikanes, Greece, GR-002)
- (32) *kam-au [ti v-el mr čao kati an haki minuta]*
 want-1SG **COMP** come-3SG my son here at any minute
 'I expect my son to come here any minute.'
 (Sinti, Romania, RO-022)
- (33) *me adžukar-av mungr-e čhav-e [te ares-el kate sa jekh momento]*
 I wait-1SG my-OBL son-OBL **COMP** return-3SG here any one
 moment
 'I expect my son to come here any minute.'
 (Gurbet, Serbia, YU-002)

Co-acting with the factuality constraint is the constraint on agent control, which constitutes the second relevant semantic dimension on the continuum that represents Romani complementation. Semantic control is taken for granted in plain modality, of the kind illustrated above in examples (27) through (30). By contrast, we find variation in the structure of complementation when different degrees of semantic control appear. This is relevant to manipulation clauses, where the degree of control is a reflection of the power relations between the manipulator (the agent of the modal verb) and the manipulee (the agent of the embedded complement verb); see (34) through (42) and (5). It is equally relevant to purpose clauses, where control is reflected in the degree to which the intentional action (encoded in the final clause) is achievable through the action attributed to the agent in the main clause, in the case of same-subject purpose constructions as in (43) and (45), or indeed, again, in the power relations between the manipulator and manipulee, in the case of different-subject purpose constructions, as in (44).

While there is a continuum along the axis of agent control, there is also the independent factor of manipulative intent of the agent. If the manipulative intent of the agent is the focus of the proposition, complementation is not necessarily bound by the confines of the control continuum, and there is a tendency to express it through *te* across the dialects "irrespective of whether or not the target action is actually realized by the manipulee" (Matras 2002: 182):

- (34) *mang-ava [akava ti naš-el]*
 want-1SG he **COMP** leave-3SG
 'I want him to go away.'
 (Sofades, Greece, GR-004)

- (35) *phend-em la-ke [te čin-el purum aj šax]*
 told-1SG her-DAT **COMP** buy-3SG onion and cabbage
 ‘I told her to buy onions and cabbage.’
 (Mexican Vlax, MX-001)
- (36) *ou rodizj-a man [te d-au les love]*
 he asked-3SG me.OBL **COMP** give-1SG him money
 ‘He asked me to give him money.’
 (Ursari, Romania, RO-004)
- Many dialects also show a strong tendency to use *te* with all manipulation clauses, even the ones with the strongest degree of control. This is true for samples from Albania, Serbia and Kosovo, most samples from Bulgaria, Croatia and Macedonia, samples from Hungary (Lovari and Gurvari), about half of the samples from Romania, as well as Crimean Romani samples from Russia and Ukraine, Romacilikanes dialect of Greece, and some Servi samples from Ukraine:
- (37) *o dad kerd-a le [te bičhav-el o lil]*
 ART father made-3SG him.OBL **COMP** send-3SG ART letter
 ‘His father made him send the letter.’
 (Mečkari, Albania, AL-001)
- (38) *me rod-em la-tar [te dža-l-tar so maj sigo]*
 I demanded-1SG her-ABL **COMP** go-3SG-away what very quickly
 ‘I demanded for her to leave immediately.’
 (Gurbet, Serbia, YU-002)
- (39) *oj nateringj-a le [te dža-l pes]*
 she made-3SG him.OBL **COMP** go-3SG REFL
 ‘She made him leave.’
 (Arli, Macedonia, MK-002)
- (40) *bar-e manuša thejard-e [te thar-as amar-e khera]*
 big-PL men made-3PL **COMP** burn-1PL our-PL houses
 ‘The government made us burn our houses.’
 (Crimean, Ukraine, UKR-001)
- (41) *oj kerd-as os [te naš-el]*
 she made-3SG him.CL **COMP** leave-3SG
 ‘She made him leave.’
 (Romacilikanes, Greece, GR-002)
- (42) *voj muk-el e farfurja [te pher-el]*
 she let-3SG ART plate **COMP** fall-3SG
 ‘She lets the plate fall.’
 (Laeši Kurteja, Moldova, MD-007)

The control and integration continuum with purpose clauses is manifested in Romani primarily in terms of how (un)contentiously the action of the main clause will lead to the outcome of the complement clause. The more certain the outcome, the more integrated the event is, and thus the more likely we are to see the simplex complementizer *te*. The less certain the outcome, the less control the agent is perceived to have over it, the more likely we are to see the more complex, multi-element complementizers (see below). The weakest degree of control can be seen with potential purpose propositions, as in ‘in order to A, one has to B’. Out of the available samples, the continuum is best illustrated with data from Goli Cigani dialect of Bulgaria, where we see, in order of decreasing integration, *te* vs. *za te* vs. *za da te*:

- (43) *avij-om khere [te dikh-av tut]*
 came-1SG home **COMP** see-1SG you.OBL
 ‘I came home to see you.’
 (Goli Cigani, Bulgaria, BG-011)

- (44) *oj veče dij-as t-e phales-te o paras [za te*
 she already gave-3SG your-OBL brother-LOC ART money **COMP COMP**
dža-l maškare]
 go-3SG to.the.middle
 ‘She already gave your brother the money to go to town.’
 (Goli Cigani, Bulgaria, BG-011)

- (45) *oj tho-ela o xurd-es te skemba [te beš-el] [za da*
 she put-3SG ART child-OBL on chair **COMP** sit-3SG **COMP COMP**
te parvar-el les]
COMP feed-3SG him.OBL
 ‘She sits the child down on the chair to feed him.’
 (Goli Cigani, Bulgaria, BG-011)

While Goli Cigani shows the full spectrum of the continuum, many dialects show a uniform use of simplex *te* with all purpose clauses. Many of these are the same samples that use exclusively simplex *te* with manipulation clauses, as discussed above. We find this non-differentiation along the control continuum with all samples from Croatia, Macedonia, Serbia and Kosovo, both Hungarian Lovari and Gurvari, many dialects of Romania and Moldova (plus Vlach dialects elsewhere: Russian Lovari, Czech Vlach and Ukrainian Shanxajcy), some samples from Bulgaria (Xoraxani), as well as Crimean Romani dialect. All of these dialects show simplex *te* even with the weakest degree of control:

- (46) *oj pučlj-a ma so te čjer-el [te av la*
 she asked-3SG me.OBL what COMP do-3SG **COMP** come her.OBL
majbut love]
 more money
 ‘She asked me what to do to earn some more money.’
 (Gurbet, Croatia, HR-001)
- (47) *d-em e gajž-es love [te kin-el kaveja]*
 gave-1SG ART man-OBL money **COMP** buy-3SG coffee
 ‘I gave the non-Gypsy man some money so that he could buy coffee.’
 (Lovari, Hungary, HU-004)
- (48) *[te džja-n ke djukjana] tumen-ge kam-el [te džja-n karjin ke*
COMP go-3PL to shop you.PL-DAT want-3SG COMP go-3PL toward to
khangerji]
 church
 ‘To go to the shop, you have to walk towards the church.’
 (Crimean, Russia, RUS-011)

The *KAJ*-type complementizer in Romani (as mentioned above, we use the arche-type *KAJ* to capture the fact that this function is often filled by borrowed forms) occurs with various cognition-utterance verbs, to use the term from Givón (1990: 517), such as ‘see’, ‘hear’, ‘say’, ‘know’, ‘understand’, ‘think’, ‘feel’, ‘seem’ and ‘be certain’, among others. There are no co-reference restrictions between arguments of the main and the complement clause. The argument of the complement clause can refer either to the same subject as the main clause, or to a different subject:

- (49) *voj phend-a [kaj či pindžard-a khanikaj]*
 she said-3SG **COMP** NEG knew-3SG nobody
 ‘She said that she did not know anyone.’
 (Gurbet, Croatia, HR-001)
- (50) *me šundž-om [kaj jov adaj džid-o ot červca]*
 I heard-1SG **COMP** he here live-PTCP.M.SG from June
 ‘I heard he has lived here since June.’
 (Polish Xaladytka, Poland, PL-014)
- (51) *me džan-au [kaj sas tu but bučji gadalaj duj brš]*
 I know-1SG **COMP** was you much work these two years
 ‘I know that you had a lot of work during the past two years.’
 (Manuša Čurjaja, Croatia, HR-003)
- (52) *me dedum-av [kaj kažno manuš gindisar-el kajci pal pe]*
 I reckon-1SG **COMP** every man think-3SG only of REFL
 ‘It seems to me everybody thinks only of themselves’
 (Xandžari, Ukraine, UKR-010)

While these examples show how the prototypical function of *KAJ* is related to factuality (i.e. potentially confirmed truth-value of the embedded proposition), *KAJ* may also combine with *te* to form a duplex complementizer. Typically, this strategy of using *KAJ* as a ‘reinforcer’ for the non-factual complementizer *te* is found in cases of weak agent control, often in cases of different-subject manipulation and purpose clauses, or more generally, with weaker semantic integration of the two clauses, the main and embedded clause:

- (53) *zaphand-av e blaka [kaj amen te na šun-el]*
 shut-1SG ART window **COMP** US.OBL **COMP** NEG hear-3SG
 ‘I will shut the window so that (he) can’t hear us.’
 (East Slovak, SK-002)
- (54) *jej čuv-el bajatos pale skamint [kaj te d-el te*
 she put-3SG child.OBL on table **COMP** **COMP** give-3SG COMP
xa-l les]
 eat-3SG him.OBL
 ‘She seats the kid on the chair to feed it.’
 (Plasčuny, Ukraine, UKR-019)
- (55) *avilj-om khere [kaj te dikh-av tut]*
 came-1SG home **COMP** **COMP** see-1SG you.OBL
 ‘I came home to see you.’
 (Romungro, Romania, RO-059)
- (56) *ov kind-as neve cavala [ka te dža-l ti poli]*
 he bought-3SG new.PL clothes **COMP** **COMP** go-3SG to town
 ‘He bought new clothes so that he could go into town.’
 (Romacilikanes, Greece, GR-002)
- (57) *vov phendj-a [kaj me te bikin-au e tradini]*
 he told-3SG **COMP** I **COMP** sell-1SG ART car
 ‘He told me to sell the car.’
 (Laeși Kurteja, Moldova, MD-007)
- (58) *kam-av [kaj te dža-l het]*
 want-1SG **COMP** **COMP** go-3SG away
 ‘I want him to go away.’
 (East Slovak, SK-002)

The overall tendency across Romani dialects is to use simplex *te* with tighter integrated events, and to use a duplex or complex linking elements for those events that are less tightly integrated. In cases where a duplex complementizer is used, one of the elements is usually *te*, while the other, reinforcing element is usually the *KAJ*-type complementizer used by the particular dialect. With more complex linking, the additional reinforcing elements often come from borrowed particles

(conjunctions and prepositions) that are used in similar constructions in the contact language(s). Matras (2002: 182) suggests that this demonstrates two-level iconicity: tighter semantic integration correlates to a) simpler form of the subordinator (in the case of *te* complements), and b) to inherited forms (both *kaj* and *te*), while less tight integration is associated with more complex subordination strategies and is much more susceptible to influence from the outside language. Thus, Lombard Sinti of Italy uses duplex *par te*, where the first element is the borrowed Italian purpose particle. Romungro dialect of Hungary uses duplex *hodj te*:

- (59) *pirad-om i vali [par ta sun-a-to]*
 opened-1SG ART window **COMP COMP** hear-1SG-you
 'I opened the window so that I can hear you.'
 (Lombard Sinti, Italy, IT-011)

- (60) *khēre āj-om [hodj te dikh-av tut]*
 home came-1SG **COMP COMP** see-1SG you.OBL
 'I came home to see you.'
 (Romungro, Hungary, HU-009)

Many dialect samples use the opposition of *te* vs. borrowed *KAJ*-type reinforcer + *te* in both manipulation and purpose clauses. This is true of the dialects with strong Russian influence, which use *so(b)* as the reinforcer. Sofades dialect of Greece uses Greek loan *ja* as a reinforcer in both types of clauses. Some East Slovak samples use Hungarian-loan *hoj*, while many Finnish Romani samples often utilize either Swedish *at* or Finnish *et(tä)*. In cases where the strategies of manipulation clause and purpose clause linking differ, the purpose clause strategies seem to be more complex and specialized, and often involve more specific borrowing from the other languages, while manipulation clause linking strategies are more likely to match in form the *KAJ*-type particles found with factual complementation. Examples of more specialized purpose clause complements that are borrowed include Italian *par* in *par ta*, Slavic *bi* in *ta bi ta*, South Slavic *za* and *da* in *za da te*, while examples of more complex Romani-internal developments for purpose clause linking are Finnish Romani *toolesko khaal*, and *kaš (de) te* found in several Vlach-type samples (see Section 2.5 below on combinability complementizer elements).

Table 1 summarises the effect of semantic integration on the choice of complementizer in Romani. Tighter semantic integration (higher agent control and greater likelihood of the success of the main predication to entail the realisation of the proposition contained in the embedded predication) is likely to correlate with the use of a simplex complementizer, which is almost invariably the inherited form *te*. By contrast, loose semantic integration (weaker control of the agent of the main predication over the agent of the embedded predication, or weaker

likelihood of realisation of the proposition encoded by the embedded predication) is more likely to correlate with the use of a complex form.

Table 1: Manipulation and purpose clauses – tendencies

INTEGRATION:	tighter		→		looser
COMPLEXITY:	simplex	→	duplex	→	complex
SOURCE:	inherited		→		borrowed
FORMS:	<i>te</i>	→	+ <i>KAJ</i> -type	→	+ other

2.2 Distribution

Romani complementizers appear with modal (non-factual) and factual predicates, as well as in purpose and manipulation clauses. The inherited modal complementizer *te* is stable and found in most of the dialects with all modals, including volition ('want'), positive and negative ability ('can', 'cannot'), inception and termination ('begin', 'start', 'end', 'finish'), obligation ('have to', 'need', 'must') and attempt ('try'), as well as other predicates, such as 'expect', 'like', 'fear', 'know how', 'dare' and 'have the strength'. It is also used regularly across dialects in manipulative and purpose clauses. Factual complementation is accomplished through the use of the inherited complementizer *kaj*, or through borrowed complementizers. The *KAJ*-type complementizer appears with various factual cognition-utterance verbs, such as 'see', 'hear', 'say', 'know', 'understand', 'think', 'feel', 'seem' and 'be certain', among others. In addition, it often appears as a reinforcer with manipulation clauses, including desiderative predicates ('want', 'wish'), manipulating predicates ('ask', 'demand', 'tell'), as well as with purpose clauses. In addition various borrowed and calqued markers are used with factual, and especially manipulation and purpose complements. Table 2 gives examples of the distribution of complementizer types across different main clause predicates from four distinct Romani dialects.

In the table [above](#), Mečkari represents the simplest system, where only two forms *-te* and *kaj-* are in binary opposition: *kaj* is found with factual complementation, whereas *te* is used with modal, manipulation and purpose clause predicates. This binary opposition is found in most samples from Croatia, Macedonia, Serbia and Kosovo, many of the samples from Bulgaria and Romania, as well as Hungarian Lovari and Gurvari. This is also true for Crimean Romani, and Šanxajcy and Xandžari samples from Ukraine.

Table 2: Complementizers with various predicates

predicate	Mečkari, Albania (AL-001)	Bergitka, Poland (PL-007)	Kalderaš, Romania (RO-065)	Gimpeny, Ukraine (UKR-020)
epistemic				
see	n/a	<i>kaj</i>	<i>kaj/kə</i>	<i>kaj</i>
hear	<i>ka(j)</i>	<i>kaj</i>	<i>kaj</i>	<i>kaj</i>
say	<i>ka(j)</i>	<i>kaj</i>	<i>kaj/kə</i>	<i>kaj</i>
know	<i>ka(j)</i>	<i>kaj</i>	<i>kaj</i>	<i>kaj</i>
understand	<i>ka(j)</i>	<i>kaj</i>	<i>kaj</i>	<i>kaj</i>
think	<i>ka(j)</i>	<i>kaj</i>	<i>kaj/kə</i>	<i>kaj</i>
feel	n/a	<i>kaj</i>	<i>kaj/kə</i>	<i>kaj</i>
modality				
want	<i>te</i>	<i>te</i>	<i>te</i>	<i>te</i>
can	<i>te</i>	∅	<i>te</i>	<i>te</i>
cannot	<i>te</i>	∅	<i>te</i>	<i>te</i>
must	<i>te</i>	<i>te</i>	<i>te</i>	<i>te</i>
begin	<i>te</i>	<i>te</i>	<i>te</i>	<i>te</i>
try	<i>te</i>	n/a	n/a	<i>te</i>
like	<i>te</i>	<i>te</i>	<i>te</i>	<i>te</i>
fear	<i>te</i>	<i>te</i>	<i>te</i>	<i>te</i>
know how to	<i>te</i>	<i>te</i>	n/a	<i>te</i>
dare	<i>te</i>	<i>te</i>	<i>te</i>	<i>te</i>
manipulation				
want	<i>te</i>	<i>kaj te</i>	<i>ka(j) te</i>	<i>kaj te</i>
demand	<i>te</i>	<i>kaj te</i>	<i>te</i>	<i>kaj te</i>
ask	<i>te</i>	<i>kaj te</i>	<i>te</i>	<i>kaj te</i>
tell	<i>te</i>	<i>kaj te</i>	<i>te</i>	<i>te</i>
convince	<i>te</i>	<i>kaj te</i>	<i>te</i>	<i>te</i>
let	<i>te</i>	<i>kaj te</i>	<i>te</i>	<i>te</i>
independent clause				
purpose clause	<i>te</i>	<i>kaj te</i>	<i>kaš te, te</i>	<i>te, kaj te</i>

Many dialect samples show a three-way distinction, with the usual modal *te* and factual *kaj* (or *KAJ*-type), and a duplex complementizer *kaj te* used consistently with both manipulation and purpose clauses. This three way distinction with inherited *kaj*, exemplified with Bergitka in the table above, is found in many dialect samples from Slovakia, some samples from Romania (Romungro, Spoitori, Ursari), all of the samples from Poland (Polska Roma, Bergitka, Polish Xaladytka), and several samples from Ukraine (Plasčuny, Kišinjovcy). The three way distinction where a borrowed or calqued element is used instead of *kaj* is

found in many samples that are in strong contact with Russian, including dialects from Russia, Lithuania, Latvia, and the Servy dialects of Ukraine.

Romanian Kalderaš in the table above is one of the few dialects of Romani where manipulation and purpose clauses show different complementation strategies. In all such cases, the strategies found in purpose clauses are more complex and specialized than those found in manipulation clauses. Other such dialects include Lombard Sinti, which uses *par ta* in purpose clauses; Italian Molise, which uses the Slavic irrealis particle (*ta bi ta*); some dialects from Bulgaria, which use complementizer elements borrowed from South Slavic (*ta te, za da te*); and some of the samples from Finland, which use a language-internal derivation *toolesko*, a demonstrative pronoun ‘there’ with a genitive case ending.

Finally, the Gimpeny sample in Table 2 illustrates that some of the dialects do not have a clear-cut split in the purpose and manipulation clauses; rather, the simpler form *te* is used when there is a tighter integration between the predicate and the complement clause, and more complex strategies are used when there is looser integration.

2.3 Complementizers as a system

The Romani system of complementizers can be discussed at three levels: functional-semantic, structural, and diachronic (etymological). At the functional-semantic level, the principal poles are factuality and non-factuality, with predicates and predicate combinations arranged, potentially, on a continuum in between the two, and continuum points being sensitive to the likelihood of realization of the embedded predicate, which entails the degree of semantic control of the main agent over the action depicted by the embedded predicate, either directly (over that action itself) or indirectly (through control of a secondary agent or manipulee). At the structural level, the continuum shows a range from the absence of complementizers (see the discussion in Section 2.4 below), through the use of simplex complementizers, until the use of duplex and complex complementizers. By and large, the structural continuum maps directly onto the functional-semantic continuum (see Table 1). As a language that is permanently in contact, and with dispersed dialects showing a history of contact with different languages, Romani offers a unique laboratory for inquiries into contact-induced structural change. Etymology thus figures prominently in the area of complementizers, which are often prone to borrowing, though under certain constraints, which often equally map onto the semantic-functional continuum (see also discussion in Elšík & Matras 2009). Table 3 summarizes all the complementizers commonly found across the dialects, including the types of predicates that they

are associated with, their etymology, and the additional functions of the complementizer forms, where appropriate (for a discussion of additional semantic functions see Section 2.6 below).

Table 3: Complementizer forms, functions and etymology

complementizer (including simplex and duplex)	type of predicate (semantics)	etymology (inherited/borrowed)	additional functions
<i>te</i>	modals, purpose, manipulation	inherited	conditional, optative/imperative
∅	inherited modals of (in) ability		
<i>kaj</i>	factual	inherited	interrogative ‘where’, relativizer
<i>ke, oti, hod/hoj, ani</i>	factual	borrowed	
<i>so, sy(r)</i>	factual	calqued	interrogative
<i>kaj te</i>	manipulation, purpose	inherited	
<i>et(tă), ta</i>	factual, manipulation, purpose		
<i>ke te, hod/hoj te, ja ti</i>	manipulation, purpose	borrowed + inherited	
<i>sob(y), sob(y) te</i>	manipulation, purpose	semi-calqued	conditional <i>by</i>
<i>za te, za da te, par te, ta te, kaš te</i>	purpose	borrowed + inherited	
<i>toolesko (khaal)</i>	purpose	language-internal	genitive case

2.4 Complementizer omission

The general tendency across Romani is to use the overt complementizer *te* for modal complementation. The form is very stable, both in terms of being highly immune to replacement through borrowing, and in terms of covering the usual stock of modals. There are several contexts, however, where this complementizer is “omitted” (we prefer to speak simply of the absence of a complementizer, rather than its “omission”). Absence of a complementizer (of the *KAJ*-type) with factual complements is rare and can be attributed to a paratactic-like structure, where the main predicate is usually separated from the embedded clause by a pause, or

a re-formulation or sequential (rather than embedded) structure can at least be inferred:

- (61) *dikh-ava* [Ø *savre* *düşünüiler* *sāde* *pumen-ge*]
 see-1SG Ø everybody thinks only REFL.PL-DAT
 ‘Everybody seems to be thinking only of themselves.’
 (Xoraxani, Bulgaria, BG-023)

- (62) *šund-an* [Ø *vi* *aver* *rom* *beš-en* *kote*]
 heard-2SG Ø also other Roma live-3PL here
 ‘You’ve heard that other Roma live here as well.’
 (Lovari, Hungary, HU-004)

Elsewhere, zero-complementizer is restricted to modality clauses:

- (63) *me kam-am* [Ø *nodža-l* *pes-ke* *jou*]
 I want-1SG Ø leave-3SG REFL-DAT he
 ‘I want him to go away.’
 (Čuxny, Latvia, LV-005)

Here too, however, it is often subject to variation; cf. (64) and (65).

- (64) *ame našy* [Ø *arakhlj-am* *rupun-i* *angrusti*]
 we cannot Ø found-1PL silver-F ring
 ‘We couldn’t find the silver ring.’
 (Crimean, Ukraine, UKR-001)

- (65) *ov ni može-t* [*te dža-l*]
 he NEG can-3SG.BOR COMP go-3SG
 ‘He can’t leave.’
 (Crimean, Ukraine, UKR-001)

The generalisation of zero-complementizer as a rule is bound to certain modality predicates, which are arranged on a hierarchical continuum (cf. Elšík & Matras 2009). The Ø complementizer is commonly found with the inherited modals of positive and negative ability (‘can’ and ‘cannot’), both of which are impersonal (not inflected for person and number). Out of the two, the modal of negative ability (*našti(k)* / *naši*) is diachronically much more stable across the dialects, while the modal of positive ability is much more prone to borrowing from the contact languages. The inherited form of this latter modal, *ašti(k)*/*sašti*/*hašti*/*vašti*/*šaj*, is replaced in various Romani dialects with Slavonic *mog-/mož-*, Greek *bor-*, Italian *pot-*. As a rule, the inherited forms of the modals of ability are more likely to take a zero-complementizer, cf. examples (66)–(69), while the borrowed forms take the usual modal complementizer *te*, cf. examples (70)–(71).

- (66) *naši* [Ø *sikjav-a*] *soskətu trjabva* [*tə pomeizə-a m-ə da-ke*]
cannot Ø study-1SG because needed COMP help-1SG my-OBL mother-DAT
 'I cannot study because I have to help my mother.'
 (Muzikantska Roma, Bulgaria, BG-010)
- (67) *nasti* [Ø *putr-av-as i porta*]
cannot Ø open-1SG-REM ART door
 'I couldn't open the door.'
 (Sofades, Greece, GR-004)
- (68) *me ašti* [Ø *d-av les xari cigares*]
 I **can** Ø give-1SG him.OBL some cigarettes
 'I can give him some cigarettes.'
 (Sofia Erli, Bulgaria, BG-024)
- (69) *ov či darajlo kaj ša* [Ø *per-o*]
 he NEG scared.M COMP **can** Ø fall-PTCP.M.SG
 'He wasn't afraid that he might fall.'
 (Gurbet, Hungary, HR-001)
- (70) *tu možyn-es* [*te zasuv-es txaves-a da ljoxi*?]
 you **can-2SG** COMP sew-2SG thread-INS these holes
 'Can you mend these holes with thread?'
 (Lithuanian Romani, LT-007)
- (71) *borin-eha* [*ti pus-es les*]
can-2SG COMP believe-2SG.SBJV him.OBL
 'You can believe him.'
 (Sofades, Greece, GR-004)

There are exceptions to this rule, however. Many dialects of the Southern Balkans, while keeping the inherited forms of the modals of ability show a tendency to use the overt complementizer *te*:

- (72) *ame našti* [*te irin-as amen*]
 we **cannot** COMP return-1PL REFL.1PL
 'We cannot go back.'
 (Sofia Erli, Bulgaria, BG-024)
- (73) *mislisar-av kaj šaj* [*theara te iriv kir-e pare*]
 think-1SG COMP **can** tomorrow COMP return your-PL money
 'I think that I will be able to pay you back tomorrow.'
 (Gurbet, Macedonia, MK-001)
- (74) *šaj* [*te pomožin-av tumen-gje*]
can COMP help-1SG you.PL-DAT
 'I can help you to.'
 (Arli, Macedonia, MK-002)

- (75) *našti* [te irin amen]
cannot COMP return REFL.1PL
 ‘We cannot go back.’
 (Arli, Macedonia, MK-002)

Exceptions of the opposite kind – borrowed modals of ability with zero-complementizer – are also found, usually with borrowed modals that are impersonal:

- (76) *može* [Ø inanasən-əs les-ke]
possible Ø believe-2SG him-DAT
 ‘(You) can believe him.’
 (Xoraxani, Bulgaria, BG-015)

- (77) *nam pot* [Ø studin-a pe]
 NEG **can Ø** study-1SG REFL
 ‘I cannot study.’
 (Molise, Italy, IT-010)

- (78) *ni miga* [Ø v-as pali] finke joj na sasto-la
 NEG **can Ø** come-1PL back until she NEG become.healthy-3SG
 ‘We cannot go back until she gets well.’
 (Lombard, Italy, IT-011)

When the borrowed modal has personal forms, however, the tendency is to use an overt complementizer *te*:

- (79) *bori-s* [te ker-es ola gures šeles-a]?
can-2SG COMP do-2SG these holes thread-INS
 ‘Can you mend these holes with thread?’
 (Romacilikanes, Greece, GR-002)

- (80) *jesli tumə kam-ən mə možyn-ou* [tumən-ge te pomožyn-ou]
 if you.PL want-2PL I **can-1SG** you.PL-DAT **COMP** help-1SG
 ‘If you (pl) want I can help you.’
 (Lithuanian Romani, LT-005)

There are also cases when the inherited modals of ability become personal in individual dialects, and inflect for person and number. In these cases, the overt complementizer *te* is always used:

- (81) *naštisar-as* [te bold-as]
cannot-1PL COMP return.1PL
 ‘We cannot go back.’
 (Mexican Vlax, Mexico, MX-001)

- (82) *dašti-s* [te av-es man-ca]?
can-2SG COMP come-2SG me-INS
 ‘Can you come with me?’
 (Mexican Vlax, Mexico, MX-001)
- (83) *nasčind-em* [te pətr-av o udar]
could.not-1SG COMP open-1SG ART door
 ‘I couldn’t open the door.’
 (Piculesti, Romania, RO-013)
- (84) *či šajnd-em* [te arak-au les niči-sar]
 NEG **could-1SG COMP** find-1SG him.OBL NEG-how
 ‘I couldn’t find it anywhere.’
 (Kurturare, Romania, RO-015)
- (85) *dašti-l* [te d-el bryšynd i ando Julio]
can-3SG COMP give-3SG rain is in July
 ‘It is possible that it will rain in July.’
 (Rakarengo, Romania, RO-002)

The same phenomenon exists in various dialects of Moldova, as well as in those dialects of Ukraine that have arrived there relatively recently, in the last 150 years. This seems to be an areal phenomenon, affecting different dialects with current and recent contact with the Romanian language.

Sometimes the inherited modals of ability are replaced through Romani-internal material, as in the case of the Čuxny (Estonian Romani) verb *dole-* (< Russian aktionsart *do-* + Romani verb *le-* ‘take’). Another example is the Kubanski Servi *vydža-* (< Russian aktionsart *vy-* + Romani verb *dža-* ‘go’), which is a calque from the Russian *vy-xodit’* ‘go out, come out’. In these cases the modal is inflected for person, and, as expected, it takes an overt complementizer:

- (86) *dol-esa* tu [man-ca te j-ēs]?
can-2SG you me-INS **COMP** come-2SG
 ‘Can you come with me?’
 (Čuxny, Estonia, EST-005)
- (87) *tu-te vydža-la* [te dža man-sa]?
 you-LOC **can-3SG COMP** go me-INS
 ‘Can you come with me?’
 (Kubanski Servi, Ukraine, UKR-008)

While it is quite common to find zero-complementizers with the impersonal modals of positive and negative ability, all other modals in Romani are usually personal and necessarily take the overt complementizer *te*. Exceptions are

Finnish Romani and Lotfitka Romani of Latvia; both these dialects show optional zero realization of modal complementizers:

- (88) *tu mustul-as [Ø aav-en man-go]*
 you must-REM Ø come-2PL me.DAT
 'You should visit me.'
 (East Finnish Romani, FIN-008)

- (89) *mir-i tykn-i čajori straxadžo-la [Ø dža-l pirdal phurt]*
 my-F little-F daughter fear-3SG Ø go-3SG across bridge
 'My little daughter is scared to go across a bridge.'
 (Lotfitka, Latvia, LV-005)

- (90) *tu na kam-es [Ø dža-s po foros]*
 you NEG want-2SG Ø go-2SG to town
 'You do not want to go to town.'
 (Lotfitka, Latvia, LV-006)

As with the modals of ability, zero realization of the complementizer is more likely to appear in impersonal modal constructions. East Slovak Romani has the zero modal complementizer with the borrowed modal *rado* 'like', which is impersonal.

- (91) *rado [Ø pij-av tosara jekh kuči kava]*
 like Ø drink-1SG morning INDF cup coffee
 'I like to have a cup of coffee in the morning.'
 (East Slovak, Slovakia, SK-011)

Likewise, Prekmurski dialect of Slovenia has the zero modal complementizer with the borrowed modal *mereš* 'like':

- (92) *mereš [Ø pij-av džek kūči kofē račas-kro]*
 like Ø drink-1SG INDF cup coffee morning-GEN
 'I like to have a cup of coffee in the morning.'
 (Prekmurski, Slovenia, SLO-001)

Borrowed verbs in Romani normally require the use of loan verb adaptation markers. A number of dialects, however, show wholesale borrowing of verbs along with their original (L2) inflectional morphology. Russka Roma, in contact with Russian, and Xoraxani Romani of Bulgaria, in contact with Turkish, are two such dialects. In both of these dialects we find zero modal complementizers with non-adapted complement verbs:

- (93) *m-i phej bašladi [Ø bārinmā] kana thard-e amar-e khera*
 my-F sister began.BOR Ø **scream.BOR** when burned-3PL out-PL houses
 'My sister began to scream when they burned down our house.'
 (Xoraxani, Bulgaria, BG-023)

- (94) *o la nināzda kuveči [∅ dōnsin khere]*
 ART her.OBL was.not.BOR strength ∅ walk.BOR home
 ‘She did not have the strength to walk back home.’
 (Xoraxani, Bulgaria, BG-023)

- (95) *me kam-am [∅ ujexatj po kurko, kaj gožo i kuč]*
 I want-1SG ∅ leave.BOR for week where nice and pretty
 ‘I want to go somewhere nice and peaceful for a week.’
 (Russka Roma, Russia, RUS-003)

The overall tendencies that condition the absence vs. presence of a complementizer in modal complements across Romani dialects are summarized in Table 4.

Table 4: Zero- vs. overt modal complementizer: tendencies

∅	<i>te</i>
with modals of positive and negative ability	vs. with other modals
with impersonal modals	vs. with personal modals
with inherited modals of positive and negative ability	vs. with borrowed modals of positive and negative ability
with unadapted complement verbs	vs. with adapted complement verbs
Finnish and Lotfitka Romani	vs. other dialects

2.5 Combinability issues

Two issues can be flagged in relation to combinability: word order constraints, and the formal combination of complementizer forms. The principal word order constraint applies to the non-factual or modal complementizer *te*, which always appears immediately before the finite verb, thus functioning as an additional, analytical marker of the subjunctive mood:

- (96) *me kam-om [jekvar Amerika te dža-u]*
 I want-1SG once America COMP go-1SG.SBJV
 ‘I want to visit America someday.’
 (Čuxny, Estonia, EST-005)

Above we already addressed the distribution of duplex and complex complementizers. These are basically an expansion of the inventory of non-factual complementizer forms, used to capture weak semantic integration and low agent control. Prototypically, they involve a combination of the two main complementizer forms, the factual *KAJ* and the non-factual *te*:

- (97) *kam-av* [**kaj** **te** *oddža-l*]
 want-1SG **COMP COMP** leave-3SG.SBJV
 ‘I want him to go away.’
 (Bergitka, Poland, PL-007)

Recall that *KAJ* represents both the inherited form *kaj* itself, and internal grammaticalization and borrowings that take on the function of the factual complementizer in some of the dialects. Many dialects currently in contact with Russian show the form *sob*, derived from inherited *so* ‘what’ and the Russian conditional marker *b(y)*, a semi-calque from Russian *čtob(y)*, which is comprised of *čto* ‘what’ and the conditional/irrealis particle *-b*:

- (98) *me phend-om* [**sob** *voj* *jekhatyr* **te** *udž-al*]
 I told-1SG **COMP** she at.once **COMP** leave.3SG
 ‘I told her to leave at once.’
 (Kubanski Servy, Ukraine, UKR-008)

- (99) *la-te na sys zor* [**sob** **te** *dža-l pale khere*]
 she-LOC NEG was.3SG strength **COMP COMP** go-3SG back home
 ‘She did not have the strength to go back home.’
 (Lithuanian Romani, LT-005)

Borrowed prepositions with the final meaning ‘for’ or ‘in order to’ may also combine with non-factual *te* in cases of predication pairs that are on the weak end of the semantic integration continuum. In the following cases, the combinations *ja ti* in a Greek Romani dialect and *za te* or *za da te* in Bulgarian Romani dialects replicate the respective constructions *ja na* and *za da* in Greek and Bulgarian:

- (100) *i dar ker-ela* [**ja** **ti** *borin-el ti rov-el*]
 ART fear did-3SG **COMP COMP** can-3SG COMP cry-3SG
 ‘The fear made her cry.’
 (Sofades, Greece, GR-004)

- (101) *voj pučlj-as ma so te ker-el* [**za** **da** **te**]
 she asked-3SG me.OBL what COMP do-3SG **COMP COMP COMP**
ker-el pobut love
 make-3SG more money
 ‘She asked me what to do to earn some more money.’
 (Kalajdži, Bulgaria, BG-007)

Note that the word order restriction mentioned at the beginning of this section, namely that of non-factual *te* immediately preceding the verb, determines the order of complementizers in a duplex form: the factual *KAJ*-type complementizer (including *kaj* and the various loans and calques) always precedes the non-factual *te*.

2.6 Non-complementizing functions of complementizer forms

Both ‘prototype’ complementizers, factual *kaj* and non-factual *te*, are aligned with other clause-combining and modality functions, respectively. Factual *kaj* serves in interrogative clauses as an interrogative pronoun ‘where’ and as such it can introduce embeddings that describe location:

- (102) *me ne-bi puč-av tut te džan-av kaj odova*
 I NEG-COND ask-1SG you.OBL COMP know-1SG where this
 ‘I wouldn’t ask you if I knew where it is.’
 (Kosovan Romani, Serbia, YU-018)

It is also the most common relative pronoun in Romani dialects:

- (103) *dikhj-om o kher kaj bori-es-as andar les-te*
 saw-1SG ART house which talk-2SG-REM about him-LOC
 ‘I saw the house that you were talking about.’
 (Ursari, Romania, RO-004)

Non-factual *te* has a general modality function that can mark the optative/imperative, especially in interrogative clauses:

- (104) *So te ker-av?*
 what COMP do-1SG
 ‘What shall I do?’
 (general)

In clause combinations, it can introduce dependent aspectual constructions, and is also the most common inherited conditional conjunction in Romani:

- (105) *palo panč minutora počnisard-a te čero-l lafi*
 after five minutes started-3SG COMP make-3SG words
 ‘After five minutes he started to talk.’
 (Gurbet, Macedonia, MK-001)
- (106) *te av-en man lōve, tu-ke d-os le*
 COMP come-3PL me.OBL money you-DAT give-1SG.REM them.OBL
 ‘If I had some money I would give it to you.’
 (Gurvari, Hungary, HU-007)
- (107) *ta v-es dikk-a-t*
 COMP come-2SG see-1SG-you
 ‘If you come, I shall see you.’
 (Molise Romani, Italy, IT-010)

Finally, in combination with the preposition *bi* ‘without’, non-factual *te* also introduces adverbial clauses that express negative circumstance:

- (108) *sar moži dž-as dži ando foros bi te molisar-av mǎndr-ǎ*
 how can go-1PL until in town without COMP ask-1SG my-OBL
phral-es e kolake?
 brother-OBL ART car
 ‘How can we get to town without asking for my brother’s car?’
 (Kalajdži, Bulgaria, BG-007)

2.7 Diachrony

The etymology of *kaj* is fairly straightforward: it is the inherited Romani interrogative ‘where’. Its use as a relativizer resembles the grammaticalization path of similar forms in other contact languages of the Balkans, most notably Greek, as does its extension to factual complements. The etymology of *te* is less obvious. It is not a cognate of Domari *ta* ‘in order to’, which is borrowed from Arabic. There are however other languages in the Near East that employ *ta* in purpose clauses, e.g. Kurdish and Neo-Aramaic, where it appears to originate in the Iranian preposition *ta* ‘until’. A deictic etymology for *te* has been considered by various authors, linking it with the Old Indo-Aryan pronoun *ta-* (Pobožniak 1964: 58), the Hindi correlative *to* (Pott 1845: 281), or Old Indo-Aryan *iti* ‘so’ (Sampson 1968 [1926]: 363). The correlative function is an attractive etymology as it can be related to the semantic dependency that characterises Romani *te* (see Matras 1994: 231–233, Matras 2002: 180).

The factual complementizer *kaj* is prone to borrowing. Matras (2002: 179–80) lists three main geographic zones where *kaj* has been replaced. The first zone on the list comprises Vlax dialects of Romania, Moldova and elsewhere, where *kaj* has been replaced by the Romanian complementizer *ke*. This change also affects the Balkan type Ursari dialect of Romania. The second zone comprises Arli and Southern Vlax varieties of Greece, where *kaj* has been replaced by the Greek *oti*. The change also affects the Dendropotamos and Agia Varvara Vlax varieties, which are spoken by more recent immigrants to Greece; this illustrates the high susceptibility to contact-induced change of the domain of factual complementation. The third zone comprises the Central dialects of Slovakia and Hungary, where *kaj* is in the process of being replaced by the Hungarian loan *hod/hodž/hod’/hot/hoj*. Consider the following examples from these three zones.

- (109) *voj phendj-as [ke či prinžan-el khanikas]*
 she said-3SG COMP NEG know-3SG nobody
 ‘She said that she did not know anyone.’
 (Kaldaraš, Romania, RO-008)

- (110) *on patjan-ile [ke som ko birtos]*
 they thought-3PL **COMP** am at bar
 ‘They thought that I was in the pub.’
 (Ursari, Romania, RO-004)
- (111) *asund-om [oti bes-ena javer da roma ate]*
 heard-1SG **COMP** live-3PL other also Roma here
 ‘I heard that other Roma live here as well.’
 (Sofades, Greece, GR-004)
- (112) *iërzindj-am [hodj valesosko bajo hi la]*
 felt-1PL **COMP** something bad is her.OBL
 ‘We felt that something was wrong with her.’
 (Gurvari, Hungary, HU-007)
- (113) *džan-el pes pal les-te [hoj hin-o but barval-o]*
 know-3SG REFL about him-LOC **COMP** is-M very rich-M
 ‘It is known that he is very rich’
 (East Slovak, SK-011)

Some additional dialects also have a borrowed factual complementizer. The Lombard Sinti dialect of Italy uses Italian *ke*, many Bulgarian dialects have borrowed Turkish *ani*, and the varieties spoken in Finland use either Finnish *et(tä)* or Swedish *at*:

- (114) *sperar-ava [ke sigo finar-ela ta d-el]*
 hope-1SG **COMP** soon stop-3SG COMP give-3SG
 ‘I hope it stops raining soon.’
 (Lombard Sinti, Italy, IT-011)
- (115) *ašund-em [ani si tumen šukar buki]*
 heard-1SG **COMP** is you.PL.OBL nice job
 ‘I heard that you have a good job.’
 (Kalburdzhu, Bulgaria, BG-008)
- (116) *tenkav-aa [at me vojpuv-aa presav-es tela louve tu-kke]*
 think-1SG **COMP** I can-1SG pay-2SG back money you-DAT
 ‘I think that I will be able to pay you back.’
 (East Finnish Romani, FIN-008)

A slightly more complex replacement of the inherited factual complementizer is found in many dialects that are in contact with Russian and Ukrainian. These include many, but not all, of the varieties of the Northeastern Romani group, namely Russka Roma, Lithuanian Roma and Estonian Roma, as well some of the Servi and Vluxurja type dialects of Ukraine. In these dialects *kaj* has been

completely replaced by *so* ‘what’, following the Russian model, where the factual complementizer *čto* has formal syncretism with the interrogative *čto* ‘what’:

- (117) *me dužakir-av [so mir-o čavo jav-ela s minuty na minute]*
 I wait-1SG **COMP** my-M son come-3SG from minute to minute
 ‘I expect my son to come here any minute.’
 (Russka Roma, Russia, RUS-003)

- (118) *jov sys dasav-o sasto [so vasten-sa vraskir-l-as sastruno]*
 he was so-M strong-M **COMP** hands-INS bend-3SG-REM iron
 ‘He was so strong that he could bend the iron with his hands.’
 (Servi, Ukraine, UKR-003)

The Northeastern dialects of Poland (Polska Roma, Bergitka, Polish Xaladytka) do not have *so*, but rather use the inherited *kaj*, which is not surprising, because Polish, unlike Russian, does not have syncretism of the interrogative ‘what’ (Polish *co*) and the factual complementizer (Polish *że*). This contact-related difference between dialects under Russian vs. Polish influence has been discussed in Tenser (2008: 206). It should be noted, that the Polish Xaladytka dialect is spoken by a group of relatively recent migrants from the Russian-speaking territories, and has most likely used the Russian model *so* until about two generations ago. This quick change from *so* to *kaj* in Polska Xaladytka was due to inter-dialectal contact and once again illustrates the high volatility of non-modal complementation.

Lotfitka Romani (Latvia) also shows an innovation here; it has replaced *kaj* with *sy* ‘how’. This seems to be a calque from Latvian, stemming from the non-differentiation of Latvian interrogative *kā* ‘how’ and factual complementizer *ka*:

- (119) *me žakir-a [sy m-o čavo vraši te ja-l paše]*
 I wait-1SG **COMP** my-M son soon COMP come-3SG back
 ‘I expect my son to come back any minute.’
 (Lotfitka, Latvia, LV-005)

In several samples two different forms of the factual complementizer coexist. In one sample from Bulgaria (Xoraxani) there is an alternation between *či* (of Bulgarian origin) and *ani* (of Turkish origin), and in various samples from Romania inherited *kaj* coexists with the borrowed *ke*; one of the Ukrainian Servi samples has both *kaj* and *so*. Mostly in such cases there does not seem to be a clear-cut grammatical or semantic distribution of the two forms. Thus in the same sample we find:

- (120) *džan-ava [ani bu iki senede sja tut but xazmeči]*
 know-1SG **COMP** last two years was you.OBL much work
 ‘I know that you had a lot of work during the past two years.’
 (Xoraxani, Bulgaria, BG-015)

- (121) *džan-ava [čī odva vakerd-a tumen-ge kal-es]*
 know-1SG **COMP** he said-3SG you.PL-DAT this-OBL
 'I know that he said this to you (pl).'
 (Xoraxani, Bulgaria, BG-015)

In the Servi sample where *kaj* and *so* co-exist, there seems to be a tendency to use *so* with complements that have a stronger independent truth value, and *kaj* with those that have a weaker independent truth value:

- (122) *man-ge na d-el-pe [kaj varokon adava skjerd-a]*
 me-DAT NEG give-3SG-REFL **COMP** someone this did-3SG
 'It does not seem to me that anyone did it.'
 (Servi, Ukraine, UKR-003)

- (123) *me phendj-om les-ke [so banza čekir-el-pe de deš štunde]*
 I told-1SG him-DAT **COMP** store open-3SG-REFL at ten hours
 'I said to him that the shop opens at 10.'
 (Servi, Ukraine, UKR-003)

3 Summary

Romani relies on combining finite clauses in complex predications, and complementizers play the key role in identifying the nature of the semantic links between main and embedded (complement) clauses. The choice of complementizer reflects the degree of semantic integration between the two clauses. The principal distinction is one between factual complementation, and non-factual or modal complementation. In following this basic typology of complement clauses, represented primarily by the choice of complementizer and correlating features such as tense-mood selection and word order, Romani aligns itself with the linguistic area of the Balkans, where most of the languages show a similar distinction. Historically, this goes back to the formation of Early Romani as a language of Early New Indo-Aryan heritage that appears to have undergone significant typological shift in contact with the languages of the Balkans, especially Greek, after the settlement of Romani-speaking populations in the Byzantine Empire in the period around the tenth century CE (cf. Matras 1994, 2002 and Elšik & Matras 2006).

The factuality distinction manifests itself in Romani prototypically through the choice between the complementizers *kaj* (factual) and *te* (non-factual, modal). Factual complements accompany verbs of perception and utterance, and show independent selection of tense-mood and word order patterns. Non-factual, modal complements accompany verbs of intent, command, and manipulation, with purpose clauses aligning themselves with the same type. They normally

appear in the subjunctive (though some dialects show other patterns), and the complementizer *te* that introduces them normally appears in the position immediately preceding the finite verb of the complement clause. Clause integration in complementation is arranged on a continuum of semantic integration, which is relevant primarily for non-factual complements. These can range from immediate, direct or tightly integrated modal complements, to those that are less tightly integrated. The criteria for semantic integration include agent control over the target action (the action depicted by the embedded, complement predicate), control over the manipulee (in different-subject constructions), the degree of event independence (especially in purpose clauses) and the likelihood that the target predication can be accomplished. Dialects that show sensitivity to this semantic continuum of non-factual complements in the form of a variety of complementizers tend to show a correlation between the structural complexity of the complementizer itself and the tightness of semantic integration; that is, tight integration is more likely to be represented by a simplex complementizer, while weaker semantic integration is flagged by a duplex or even triplex complementizer, in which the default non-factual complementizer *te* is extended by a ‘reinforcer’ form (often the factual complementizer, or a preposition, or both). In terms of diachrony, the inherited pattern emerging from Early Romani and which is continued in some form or other in most dialects shows a grammaticalization path that derives complementizers ultimately from deictic forms. More specifically, the non-factual *te* has its roots in all likelihood in an ancient correlative. The factual complementizer *kaj* derives from an interrogative turned relativizer. Contact developments in the various dialects of Romani attest to a high susceptibility of this factual complementizer to borrowing, and in many dialects it is directly replaced by borrowed complementizers, while the inherited non-factual complementizer generally remains stable.

Country abbreviations in RMS sample codes

AL	Albania	LV	Latvia
BG	Bulgaria	MD	Moldova
CZ	Czech Republic	MK	Macedonia
EST	Estonia	MX	Mexico
FIN	Finland	PL	Poland
GR	Greece	RO	Romania
HR	Croatia	RUS	Russia
HU	Hungary	SK	Slovakia
IT	Italy	UKR	Ukraine
LT	Lithuania	YU	Yugoslavia

References

- Boretzky, Norbert. 1996. The 'new infinitive' in Romani. *Journal of the Gypsy Lore Society*, Fifth Series, 6. 1–51.
- Boye, Kasper. 2012. *Epistemic meaning. A crosslinguistic and functional-cognitive study*. Berlin: De Gruyter Mouton.
- Cristofaro, Sonia. 2003. *Subordination*. New York: Oxford University Press.
- Elšík, Viktor & Yaron Matras. 2006. *Markedness and language change. The Romani sample*. Berlin: Mouton de Gruyter.
- Elšík, Viktor & Yaron Matras. 2009. Modality in Romani. In Björn Hansen & Ferdinand de Haan (eds.), *Modals in the languages of Europe*, 267–324. Berlin: Mouton de Gruyter.
- Givón, Talmy. 1990. *Syntax. A functional-typological introduction*, vol. II. Amsterdam: John Benjamins.
- Matras, Yaron. 1994. *Untersuchungen zu Grammatik und Diskurs des Romanes. Dialekt der Kelderaša/Lovara*. Wiesbaden: Harrassowitz.
- Matras, Yaron. 1999. The speech of the Polska Roma: some highlighted features and their implications for Romani dialectology. *Journal of the Gypsy Lore Society*, Fifth Series, 9(1). 1–28.
- Matras, Yaron. 2002. *Romani: A linguistic introduction*. Cambridge: Cambridge University Press.
- Matras, Yaron. 2004. Typology, dialectology and the structure of complementation in Romani. In Bernd Kortmann (ed.), *Dialectology meets typology*, 277–304. Berlin: Mouton de Gruyter.
- Matras, Yaron. 2005. The classification of Romani dialects: A geographic-historical perspective. In Barbara Schrammel, Gerd Ambrosch & Dieter W. Halwachs (eds.), *General and applied Romani linguistics*, 7–26. Munich: Lincom Europa.
- Pobożniak, Tadeusz. 1964. *Grammar of the Lovari dialect*. Kraków: Państwowe wydawnictwo naukowe.
- Pott, August. 1844–1845. *Die Zigeuner in Europa und Asien. Ethnographisch-linguistische Untersuchung vornehmlich ihrer Herkunft und Sprache*. Halle: Heynemann.
- Sampson, John. 1968 [1926]. *The dialect of the Gypsies of Wales, being the older form of British Romani preserved in the speech of the clan of Abram Wood*. Oxford: Clarendon Press.
- Tenser, Anton. 2008. *Northeastern group of Romani dialects*. Manchester: University of Manchester dissertation.

