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 This implies that they are older than, and in most environments have been replaced by the current nominative pronouns, which show dialect variants in *ov/oj/on, jov/joj/jon, vov/voj/von*.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-Keldераš (Vlach) dialect from Transylvania (Kovalcsik–Tálos 1991, 114):

(3) The, le dò raklèn   xutjildàha-le   the phanglàh-le ke le grastèhki pourì.
     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 distribution 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 constructions and the anaphoric devices they show.

4.1 Adverbial clauses and embeddings

In adverbial subordinations, the Mark corpus shows almost exclusively free pronouns 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 pronouns 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 koï 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 Universalienforschung 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