15. Syntax

1. Code-switching and code-mixing
2. Lexical borrowing
3. Relexification
4. Syntactic borrowing
5. Pidgins and creoles
6. The contribution of language contact studies to linguistic theory
7. Bibliography (selected)

The relation between syntactic theory and the study of language contact is a two-way relation: on the one hand, syntactic theory can help refine the structural analysis of different language contact phenomena. On the other hand, language contact phenomena can help deepen the understanding of syntactic structure and its role in language behaviour. Thus one could depart from the different syntactic notions relevant to the study of language contact, or alternatively from different contact scenarios in specific situations and from specific phenomena of language contact. The last approach has been adopted here.

1. Code-switching and code-mixing

There are two dominant approaches to the problem of syntactic constraints on intra-sentential code-switching (Muysken, in press):

(a) those that depart from alternation, and view the constraints on it in terms of the compatibility or equivalence of the languages involved at the switchpoint, and (b) those that depart from the notion of insertion, and view the constraints in terms of the structural properties of some matrix, base, or host structure. The alternation models (a) involve bidirectional compatibility checking on the horizontal level of constituent concatenation, the insessional models (b) unidirectional compatibility checking within a structure.

Under model (a) code-switching is akin to the switching of codes between turns or utterances. This will be termed code-switching here. Under model (b) the process of code-switching is conceived of as something akin to borrowing: the insertion of an alien lexical or phrasal category into a given structure, termed code-mixing here. The difference between borrowing and code-mixing would simply be the size and type of element inserted, e.g. noun versus noun phrase, respectively. The two ‘models’ in fact correspond to two different phenomena, namely to the alternation between languages and to the insertion into a matrix or base language.
III. Bedingungsfaktoren: Kontaktfaktoren sprachlicher Ebenen

Shana Poplack of the University of Ottawa is the principal exponent of the alternation perspective. In her work on Spanish/English code-switching in the Puertorican community (1980), Poplack discovered that code-switching (i.e. the alternation between languages) was constrained to occur largely at sites of equivalent constituent order.

The kind of data that have been the basis for Poplack's views were from Spanish/English code-switching. They include examples such as the following, in which the elements switched do not form a single constituent:

(1) Bueno, in other words, el flight [que sale de Chicago around three o'clock]. Good, in other words, the flight [that leaves Chicago around three o'clock.]

Here que sale de Chicago 'that leaves Chicago' or even el flight que sale de Chicago (assuming Chicago to be part of the Spanish stretch for the sake of the argument — in fact it may be the trigger for the subsequent switch to English) is a constituent, but not a unique one, since it also includes the English fragment around three o'clock.

An example where the several switched elements do not form a constituent at all (unless this includes the entire utterance) is again from Spanish-English code-switching:

(2) He was sitting down en la cama, mirándonos peleando, y really, I don't remember si él nos separó. He was sitting down [on the bed, watching us fight, and] really, I don't remember [if he took us apart.]

Here the Spanish stretch in the middle consists of three separate constituents. The stretch at the end is a unique constituent, an embedded clause.

It is clear that this type of data cannot be handled very well in a model which takes insertion into a matrix and a dependency relation between matrix and inserted material as its primes. Rather, it has led to the idea that order equivalence across the switch point is what constrains code-switching here. Poplack and her associates have developed a more elaborate typology of code-switching phenomena in later work, but always taking the issue of whether a given code-switching pattern conforms to the equivalence constraint as the starting point (Poplack/Sankoff 1988).

The import for syntactic theory is how the notion of equivalence should be defined: is it a deep 'structural' notion or a surface 'linear' notion? Furthermore, notice that linear equivalence presupposes categorial equivalence (but not vice-versa), since you cannot say that the order between A and B is equivalent in two languages unless A and B are themselves equivalent cross-linguistically. Thus there is a challenge for the theory of universals here: to what extent are e.g. clitic pronouns equivalent to full pronouns in another language?

The type of data analyzed by Poplack and other researchers who have worked on Spanish-English code-switching contrasts rather sharply to the cases of Swahili-English mixing that have been the basis for Myers-Scotton's work, which exemplifies the insertion approach. Carol Myers-Scotton (1993 a, b) tries to develop a comprehensive psycholinguistically embedded linguistic model for intra-sentential code-mixing. The model proposed in her latest book (1993 b, 75–119), the Matrix Language Frame Model, crucially incorporates the idea that there is an asymmetrical relation between the matrix and the embedded language in the switching situation; this contrasts with the symmetry implied in Poplack's Equivalence Constraint, which involves properties of both languages. Furthermore, content and function morphemes behave differently in Myers-Scotton's model: the former can be inserted, when congruent with the matrix language categories, into mixed constituents, the latter cannot.

Consider some of the cases that form the basis of Myers-Scotton's analysis, which are representative for the data reported on in her work:

(3) Na kweli, hata mimi si-ko sure lakini na-suspect i-ta-kuwa week kesho. [Well, even I am not sure, but I suspect it will be next week.]

Here the elements sure, suspect, and week are single elements inserted into a Swahili utterance. The same holds for (4), where housing allowance is not necessarily a single word, but certainly a single constituent:

(4) Ujue watu wengine ni funny sana. Wa-na-clain ati mishahara yao iko low sana. Tena wanasaem eti hawapewi housing allowance. [You know, some people are very funny. They are claiming that their salaries are very low. They also say — eh — that they are not given house allowances.]
From a syntactic perspective, the theoretical interest of Myers-Scotton's work lies in the cross-linguistic exploration of the distinction between function and content morphemes, and of the role of function morphemes in defining grammatical structure.

2. Lexical borrowing

Do grammatical constraints hold for all types of language contact? One possible perspective could be that code-mixing and -switching are subject to constraints, but that borrowing is unconstrained by grammar. The opposite appears to be the case. The traditional observation, with long roots in language contact research, is that different categories are more or less easily borrowable, or at least, actually borrowed. This observation, which had a somewhat shaky empirical base until recently, has received massive support from the work reported in Poplack, Sankoff and Miller (1988). The finding that nouns are the most frequently borrowed element is confirmed for other language pairs as well. For borrowing, constraints can then be formulated in terms of a categorial hierarchy: words of one specific lexical category are borrowed more easily than those of another (cf. Art. 14). Such hierarchies predict that a noun such as French automobile can be borrowed more easily into English than a conjunction such as que, and this prediction holds reasonably well in the extreme cases. The problem, however, with this hierarchy is that there is no explanation given for the order of the lexical categories in the hierarchies. In addition, there are very striking language-specific deviations, as it turns out.

Here aspects of lexical borrowing will be focussed on that are constrained by the syntax or that have syntactic implications. In Van Hout and Muysken (in press) we explore the possibility of a probabilistic approach to borrowing by considering the interaction of a number of factors, syntactic and other.

Thus content words such as adjectives, nouns, verbs may be borrowed more easily than function words (articles, pronouns, conjunctions) since the former have a clear link to cultural content and the latter do not. A second cluster of factors is structural in nature. To what extent do syntagmatic and paradigmatic constraints on lexical items, again both in the donor language and in the recipient language, influence their borrowability, it is clear from a number of cases that words which play a peripheral role in sentence grammar, particularly the grammar of the recipient or matrix language: interjections, some types of adverbs, discourse markers, and even sentence coordination markers, are borrowed relatively easily. Note that this is the same class of elements that participates in emblematic switching, the type of phenomenon half-way between inter-sentential and intra-sentential code-switching. What this suggests is that switching and borrowing may to some extent be subject to the same type of constraints: both are difficult when the coherence of the clause is disturbed. A related way to approach the same question is to see to what extent categories are directly implied in the organization of the sentence: a verb is more crucial to that organization than a noun, and perhaps therefore it may be harder to borrow verbs than nouns. Thus for a noun/verb asymmetry in borrowability the principal explanation could lie in the different role that these categories play in the organization of the sentence. Nouns denote elements referred to, verbs link the elements referred to. In other words, nouns are inert as far as the syntactic make-up of the clause is concerned, while verbs are active in the syntax, and form the nucleus of the clause. More generally, complementizers, auxiliaries, copulas, and verbs play a role in structuring the clause, and prepositions, determiners, and demonstratives help structuring the argument constituents in the clause.

The central role of the verb is also reflected in its assigning different cases, which may be specific to that verb and ideosyncratic, to different elements in the sentence. This also stands in the way of their being taken from one system to another. Prepositions share this property with verbs, which may stand in the way of their being borrowed. On the other hand, prepositions are rarely inflected themselves. Nonetheless, they may not be frequently borrowed. Additional factors hindering their borrowability include the fact that often their meaning is grammaticalized (and hence language specific), that they are sometimes paradigmatically organized (systematically subdividing a semantic field such as space in a language specific way), or that they themselves are part of the subcategorization of a verb or adjective (angry with, afraid of, wait for, await on). This line of thinking would predict that elements such as transitive verbs and prepositions would be harder to
borrow than e.g. nouns. In addition to these factors deriving from syntagmatic coherence: peripherality, structure-building, case-marking, there is also paradigmatic coherence. Paradigmatic coherence is the tightness of organization of a given subcategory: the pronoun system is tightly organized, and it is difficult to imagine English borrowing a new pronoun to create a second person dual in addition to second person singular and plural. For this reason determiners, pronouns, demonstratives, and other paradigmatically organized words may be harder to borrow. Notice also that paradigmatic organization in the donor language may also stand in the way of borrowing since paradigmatically organized elements often have rather abstract, grammaticalized meanings, which are not accessible independently of the subsystem they are part of. Thus the fact that Spanish *este* ‘this’ and *ese* ‘that’ may be hard to borrow into Quechua could be due to two factors: first the fact that their Quechua equivalents *kay* ‘this’ and *chay* ‘that’ also form a tight subsystem (to which it would be difficult to add new members), and second the fact that the meaning of *este* is defined in opposition to that of *ese* (and of *aquel* ‘yonder’), and thus the element is not quite independently transportable into another system.

Finally, there is the factor of *categorial equivalence* to be considered. Weinreich (1953, 61) notes that resistance to borrowing is always a function, not so much of properties of recipient and sources languages by themselves, but of the difference in structures of the recipient and source languages.

3. Relexification

Generally relexification is used to refer to a process of lexical borrowing which (a) involves a large part of the vocabulary, and (b) involves the replacement of native items, rather than the mere addition of vocabulary. The way the term is used here goes one step further: in fact the new relexified forms do not enter as full lexical entries, but as phonological shapes which are grafted onto the original lexical entries. Thus the original entry is not replaced, but merely altered in outer shape. Take the French verb *embrasser* which means ‘embrace’ as well as ‘kiss’ (the original French verb meaning ‘kiss’, * baiser*, now often means ‘sleep with’). One could imagine a French relexified with German vocabulary where *küssen* also has the double meaning and corresponding range of uses of the French original. Since the grammatical properties of the original element are maintained, this type of language contact has the effect of pairing the lexicon of one language with the syntax of another one.

The different components of a lexical entry function so independently of each other that (apparently) a phonological representation can be substituted into an entry without affecting the other sets of features (syntactic, subcategorization, semantic, selectional). A language which emerged through relexification has the same lexical, morphosyntactic and syntactic categories as its source. An example is Media Lengua from Ecuador, Quechua relexified with Spanish; both the Spanish (S) and the Quechua (Q) equivalents of the Media Lengua (ML) phrase are given:

(5) a. ubixa-buk yirba nuwabi-shka (ML) sheep for grass there is not SD ‘There turns out to be no grass for the sheep.’
   (SD = sudden discovery tense)
   b. llama-buk k’iwa illa-shka (Q)
   c. No hay hierba para las ovejas (S)

Relexification necessarily must take place on the basis of meaning correspondences. In Media Lengua, a Spanish stem as close as possible in meaning to the Quechua original is used to replace it. The question now is how closeness in meaning is determined. It is necessary to distinguish here between lexical meaning and grammatical meaning. The former can be determined by reference to some extralinguistic entity, the former only by reference to the language systems themselves. When the Quechua verb *rikus* ‘see’ is reflexified as *bi-* (from Spanish *ver*), this is possible because there is a large shared element of meaning. Reflexification is feasible, with all the difficulties mentioned, for lexical items, but operates in a very incomplete manner for grammatical items. In the latter case, there is drastic restructuring of the system. Function words do not have a meaning outside the linguistic system that they are part of, since their meanings are paradigmatically defined within that linguistic system. So when you relexify a system of paradigmatically organized function words, automatically the semantic organization of the target language

---

The words mentioned in the text are: **embrasser** (French), **küssen** (German), **kay** (Quechua), **chay** (Quechua), **aquel** (Spanish), **ese** (Spanish), **este** (Spanish), **sudden discovery tense** (English), **baiser** (French), **kisse** (German), **kissen** (German).
comes in, and the result is at best a compromise between source and target language systems.

4. Syntactic borrowing

One of the reasons why so little agreement has been reached with respect to the question of what can be borrowed in language contact (in addition to vocabulary) is that the focus has been on the supposed outcome, i.e. the elements borrowed and the directionality of borrowing, and not as much on the processes of borrowing, determined by the type of contact situation (cf. Thomason/Kaufmann 1988, 35-64). As to outcomes of borrowing, there is a distinction, at least in principle, between convergence, the result of bidirectional influence, and grammatical borrowing, generally seen as a unidirectional result. There are a number of ways in which grammatical borrowing could potentially take place. Sometimes lexical borrowing will lead to syntactic borrowing, e.g. when a language with case-suffixes and/or post-positions borrows prepositions from a prestige language. In this case the [P NP] syntactic pattern will be borrowed as well. Sometimes the borrowing is limited to a frozen pattern as in Ecuadorian Quechua sin ...-r pattern borrowed from Spanish: *sin miku-r* 'without eating' modelled on Spanish *sin comer*.

In other cases, the construction may retain a grammatically peripheral status, e.g. the introduction of a Persian complementizer into Turkish. Turkish has undergone extensive lexical influence, from Arabic and Persian successively, and at the Ottoman court a very complex and flexible form of Turkish was spoken, full of Arabic and Persian expressions and phrases. One element introduced was the Persian particle *ki*, somewhat like English 'that', which created the possibility of having Indo-European-like relative clauses such as (6b) in addition to original Turkish patterns such as (6a):

(6a) a. kapıylı kapamam yanı bir çocuk  
    door not-shutting a child

b. bir çocuk *ki* kapıylı kapamaz  
    a child REL door not-shuts  
    'a child who does not shut the door'  
    (Lewis 1972)

In (6a) the relative clause is formed with a participial form of the verb, and in (6b) with the particle *ki* and a fully inflected verb. Furthermore, the original type of relative clause precedes the head noun, and the 'Persian' type follows the head noun.

This would seem to be a clear example of syntactic borrowing, in this case through cultural influence and lexical borrowing: the introduction into Turkish of the element *ki* opened the way, not only for new types of relative clauses, but also of complement clauses, just as with English 'that'. There are two observations to be made, however: first, Lewis (1972) notes that there is an old Turkish interrogative element *kim*, which through its phonetic similarity may have paved the way for the extension in syntactic use of *ki*. Second, there is some doubt that constructions of type (6b) ever really became part of Turkish. Lewis suggests that this construction 'is regarded as alien and is increasingly rare in modern Turkish.' While the latter is not borne out by recent recordings, the use of *ki* may still be a peculiar type of code-mixing, triggering a non-Turkish syntactic pattern in speech production, without this pattern really entering the grammar.

Another type of syntactic borrowing may occur in massive second language learning: semantic distinctions, pragmatically relevant ordering patterns, ways of marking causality and conditionality, etc. may be taken over into the new language. Similar, but perhaps more intensive, are the effects of interference in long-term and massive bilingualism. The latter has been the source for the remarkable phenomena of convergence and drastic syntactic restructuring in the Balkan and Indian sub-continent. Sometimes it seems more appropriate perhaps to speak of one syntax with different lexicons attached to it (the different languages), than of syntactic systems which have converged. Again, more studies are needed of contemporary communities where the apparent convergence is taking place.

An apparently very clear case of this type of syntactic influence involves some dialects of Konkani, an Indo-European language related to Marathi spoken in central India. Some centuries ago a group of Konkani speakers moved into an area where Kannada, a Dravidian language, is spoken, and they were forced by the circumstances to become bilingual: Konkani inside the home, Kannada outside. That their bilingualism was maintained and shows no sign of disappearing is perhaps due to the rigid ethnic, religious, and caste divisions that cut through
Indian society: the Konkani speakers were Brahmans and kept themselves separate socially. Nadkarni (1975) claims, however, that the structure of the Konkani dialects involved was directly affected, becoming very much like the structure of Kannada. The original Konkani relative clause, formed with a relative particle as in (7), was gradually replaced by a Kannada-type relative clause, formed with a question word and a yes/no interrogation element as in (8):

(7) jo mhältāro pepar vāccat āssa, to đākt-aru āssa

REL old-man paper reading is that doctor is

(8) khančo mhältāro pepar vāccat āssa-ki, to đāktaru āssa

which old-man paper reading is-Y/N, that doctor is

'The old man that is reading the newspaper is the doctor.'

This replacement can only be explained through the postulation of Kannada influence; it is not motivated structurally. On the whole the case presented by Nadkarni is very strong and hard to explain otherwise. An alternative explanation could involve two observations: (a) In (8) there is no replacement of one type of relative clause by another one, but rather the loss of the possibility to relativize, and its replacement by a question-like structure, which functions somewhat like a relative clause. (b) Konkani grammar is not undergoing some change, but Konkani grammar is replaced by Kannada grammar, while maintaining Konkani vocabulary. This may be called resyntactization (as opposed to relexification). Centuries of coexistence and massive bilingualism have led to the convergence of the grammars of the Indian languages, but the existing social divisions called for pluriformity. Therefore the languages remained as separate as possible on the lexical level.

Some prestige languages, e.g. Latin in Renaissance Europe or French in eighteenth century Europe, have been involved in syntactic borrowing through calquing and imitation of prestige patterns. Here the effect of the syntactic borrowing is mostly limited to the (extended) lexicon: set phrases and expressions (cf. 'if you please' and s'il vous plaît). Finally, it should be mentioned that under conditions language erosion and attrition these processes will proceed faster and with greater frequency.

5. Pidgins and creoles

Pidgins and creoles are a case of drastic language change due to contact, with far-reaching syntactic consequences. First, contact pidgins, which involve a somewhat symmetrical relationship of often only two ethnolinguistic groups, are different from L2 pidgins, which result from the attempt by different groups to communicate on the basis of an imperfectly mastered dominant language. Creoles result from nativization of L2 pidgins. Syntactic issues in the study of contact pidgins are the nature of the convergent structural adaptations to reach a common medium of communication and the factor of markedness in steering the adaptations towards an unmarked system. Thus Kouwenberg (1992) argues that the SVO-character of Berbice Dutch Creole (spoken in Guyana) resulted from a compromise between surface SVO patterns in spoken colonial Dutch and surface word order patterns in Ijo, the language spoken by most slaves. Remarkably, both contributing languages are underlyingly SOV.

For L2 pidgins the relevant issues are the types of grammatical simplification that characterize stabilized interlanguage systems, and the consequences of the loss of inflectional morphology for the grammatical system as a whole: fixed word order, emergence of grammaticalized auxiliary particles, etc. The idea that contact languages are simple has been taken to mean two things. On one level it has meant that these languages do not have a rich morphology, on another that their overall grammar is less complex than that of other languages. Both interpretations are relevant to grammatical theory. The idea that absence of morphology is related to grammatical simplicity needs to be evaluated in the context of contemporary research into morphology-syntax interactions, and the grammatical status of functional elements. Even more importantly, the idea that contact languages are not grammatically complex in general only makes sense if one has a theory of grammatical complexity to fall back on, and this brings in markedness theory.

In the study of creoles a number of syntactic issues arise. First, it has been noted that
the creoles that emerged in different parts of the world bear striking resemblances to each other. When we say that languages $x$ and $y$ are more alike than $y$ and $z$, we are claiming in fact that in the total (abstract) variation space allowed for by the human language capacity $x$ and $y$ are closer than $y$ and $z$. Consequently, the claim that the creole languages are more alike than other languages implies a clustering in the variation space. If one thinks of the variation space as defined by parameter theory, trying to develop a notion of 'alike' really boils down to developing a theory of parameters, parameters along which similarities and differences between natural languages can be defined. One such parameter is pro-drop, the possibility for the subject to be absent. There is much recent discussion about the status of creole languages in this respect.

A second issue is that of structural completeness and grammatical adequacy. Creoles remain structurally unstable for a long time, and that they keep expanding particularly in the area of distinct functional elements, through a process of grammaticalization. Thus the inventories of complex pronominal and quantificational elements (e.g. reflexives), of determiners, auxiliary particles, and conjunctions keep growing for quite some time. How necessary are these elements for the functioning of natural languages, and how creole growth differs from ordinary language change?

A third issue involves substratum influence, something that remains as hotly disputed as grammatical borrowing, since we do not know much about the way language can be mixed. Mixing implies that elements of one language are put together with elements of another one, and this in turn calls into question the cohesion of the grammatical systems involved. The tighter a particular subsystem (e.g. the vowel system, or the system of referential expressions) is organized, the less amenable it will be to restructuring under borrowing. Tightness of organization in modern grammatical theory is conceptualized within modularity theory: the grammar is organized into a set of internally structured but externally independent modules, the interaction of which leads to the final grammatical output. For this reason, the notion of mixing is important: it raises the issue of which parts of the grammar are tightly organized, and hence the notion of modularity.

6. The contribution of language contact studies to linguistic theory

The contribution that the study of language contact can make to grammatical theory is that it can help to elucidate a number of concepts and explore their role in structuring natural languages. These concepts have been introduced in the preceding sections and include:

- simplicity and markedness
- universality of categories and structures and parametrization
- functional elements and grammaticalization
- mixedness and modularity
- the role of lexical elements in syntactic structure

All these concepts turn out to be very relevant indeed to the central concerns of modern grammatical theory.

7. Bibliography (selected)


16. Lexicon and word formation

1. Introduction

The goal of this paper is to present a critical appraisal of what is presently known about the impact of contact situations on the internal structure of lexical units. In line with this goal the paper includes the following three sections: (1) a state of the art overview of the main types of concerns and achievements in the investigation of the lexicon in contact situations; (2) a specification of two basic notions independently of contact situations, namely, the lexicon and the dynamic processes involved in lexical unit formation; and (3) in conclusion, a brief programmatic statement as to how the impact of contact situations on the internal structure of lexical units needs to be investigated in order to go beyond the achievements of the pioneers in this field.

2. State of the art overview

Three types of concerns can be differentiated in the literature on the lexicon — or the linguistic system as a whole — in contact situations. These are: (1) psychological concerns focussing either on the cognitive processes mediating between the languages in contact, or on questions of lexical storage in memory; (2) concerns for the sociolinguistic factors at play in contact situations; and (3) concerns for the impact of contact situations on the lexicon.

2.1. Haugen and Weinreich are the two pioneers in the investigation of the contact phenomenon par excellence, borrowing. Haugen (1950) accounts for types of borrowing only in terms of the internal structure of lexical units (see 2.4.). Weinreich is the first to offer an interpretation of borrowing in terms of a cognitive mediation process, the notion of interference (1953, 7-70). Mackey (1970) enlarges the cognitive interpretation of borrowing by proposing two cognitive mediation processes, interference and integration. Interference is simply the use of a lexical unit from one language while speaking another. It is an aspect of code switching (Art. 73). Integration is the adoption into a language of lexical units from another language. Wölk (1987/88) attempts to combine Weinreich’s notion of interference with Ervin and Osgood’s (1965) distinction between two cognitive mediation processes, a compounding process and a coordinating process. As interpreted by Wölk (ibid., 3) “in the coordinate mode both linguistic signs, or signifiants, and their corresponding meanings, or signifiés, will be separated, while in the compound mode corresponding signifiants from two different languages will be associated within the same signifiés”. In line with the above, Wölk suggests a distinction between three cognitive mediation processes: unidirectional interference, bidirectional fusion, and complementation. Unidirectional interference is a redefinition of Weinreich’s notion of interference as “unidirectional influence of one language system upon another” (ibid., 9). Complementation is the specification of the coordinating process as “complementary distribution of variant forms across a determined functional range” (ibid., 8). It is the case in which for an individual bilingual, or a specific bilingual community, the particular vocabulary for a specific domain of discussion may only be available in one of the languages” (ibid., 7). Bidirectional fusion is the specification of the