In: Buchstaller, Isabelle & Siebenhaar, Beat (eds.) 2017. *Language Variation – European Perspectives VI: Selected papers from the 8th International Conference on Language Variation in Europe (ICLaVE 8), Leipzig 2015.* Amsterdam: Benjamins.

Analytic and synthetic: Typological change in varieties of European languages

MARTIN HASPELMATH & SUSANNE MARIA MICHAELIS

1. The macro-comparative perspective: Language typology and language contact

Since the early 19th century, linguists have sometimes tried to understand language change from a broader perspective, as affecting the entire character of a language. Since A.W. von Schlegel (1818), it has been commonplace to say that Latin was a SYNTHETIC language, while the Romance languages are (more) ANALYTIC, i.e. make more use of auxiliary words and periphrastic constructions of various kinds.

In this paper, we adopt a macro-comparative perspective on language variation in Europe, corresponding to our background in world-wide typology of contact languages (Michaelis et al. 2013) and general world-wide typology (Haspelmath et al. 2005). While variationist studies typically ask for patterns of variation within a single language, we ask whether there is a "big picture" in addition to all the details, in the tradition of A.W. von Schlegel. In particular, the topic of this paper is the replacement of synthetic patterns by analytic patterns that has interested typologists and historical linguists since the 19th century and that has recently been the focus of some prominent research in variationist studies of English and English-lexified creoles (Szmrecsanyi 2009; Kortmann & Szmrecsanyi 2011; Siegel et al. 2014; Szmrecsanyi 2016).

By way of a first simple illustration, consider the examples in Table 1, where the symbol ">>" means that the newer pattern on the right-hand side competes with and tends to replace the older pattern on the left-hand side. Most of the changes that we will talk about are well-known and have been widely discussed: development of a new prepositional genitive (as in German), development of a new auxiliary-based past tense (as in French), development of a new particle-based comparative form (as in Modern Greek); the loss of the old plural form in creoles such as French-based Seychelles Creole, and its replacement by forms such as *bann* (from French *bande* 'group') is less well-known, but also falls into the class of analyticizations. (The replacement of the definite article *la* by the demonstrative *sa* can also be seen as an example of this.)

Table 1: Some illustrative cases of synthetic and analytic patterns

	synthetic (old)		analytic (new)
German	*des Hauses* 'the house's'	>>	*von dem Haus* 'of the house'
French	*Edith chanta* 'Edith sang'	>>	*Edith a chanté* 'Edith has sung'
Modern Greek	*oreó-tero* 'nicer'	>>	*pjo oréo* 'more nice'
French > Seychelles Creole			
	la femme / les femmes 'the woman/the women'	>>	*sa fanm / sa bann fanm* 'the woman/the women'

In this paper, we make three main points. First, we discuss the basic question of how to distinguish analytic and synthetic patterns in the first place, noting that the distinction if understood synchronically rests on the concept of "(auxiliary) word", which is not well-defined except in a trivial orthographic sense (§3.1). But there is no question that a diachronic process of "analyticizing" or "refunctionalizing" is widespread and is involved in a substantial number of salient grammatical innovations (§3.2).

Second, we highlight the strongly analyticizing developments in creole languages, based on examples from the *Atlas of Pidgin and Creole Language Structures* (Michaelis et al. 2013). Compared to other Romance varieties, especially of course the standard varieties, all creoles show drastic loss of inflectional markers, their replacement by new function items, and/or the development of novel function items, mostly from earlier lexical roots (§4).

Third, we propose an explanation of these developments on the basis of the contact history of these languages, invoking general principles of contact-induced grammatical change (§5). The basic idea is that analyticizations are due to the increased need for clarity when a language has many speakers who learn it as adults. We go on to ask whether similar differences can be found within some of the major language families of Europe (e.g. with French being more analytic than Spanish, or Bulgarian more analytic than Russian), or even within the major languages (with some vernacular varieties being more analytic than the standard varieties).

Before getting to these three main points in §§3-5, we briefly discuss the history of the analytic/synthetic terminology.

2. A short history of the analytic/synthetic terminology

The terms *analytic* and *synthetic*, as they are still used today, were coined by von Schlegel (1818). He conceived of them as two subtypes of the class of inflecting languages (comprising all Indo-European languages), which was opposed to the classes of agglutinating languages (such as Turkish) and isolating languages (such as Chinese). In the early 19th century, there was generally a value judgement associated with language classification. More highly inflected languages were regarded as superior to agglutinating and isolating languages. From a modern, strictly linguistic point of view, analytic patterns

are not really different from isolating patterns, but in earlier times there was a general feeling that the situation in Romance languages should not be compared to Chinese (after all, Romance languages still have fairly rich verb inflections), so the terms *analytic* and *synthetic* survived.[1]

Typological considerations had little prestige in the 2nd half of the 19th century, but linguists became increasingly skeptical about the value judgements that were originally associated with them. Otto Jespersen even claimed that the modern analytic languages (such as English) were superior to the cumbersome classical languages (most clearly expressed in his 1894 book *Progress in language*). In the early 20th century, Sapir (1921) tried to apply the typological notions from the 19th century to the North American languages that he had studied, but typology became popular again only with Joseph Greenberg's work since the 1960s. Greenberg became best-known for the word-order correlations that he popularized and discovered, but in 1960 he also published the first paper that approached typological distinctions quantitatively, by computing an analyticity index for different languages.

The idea that whole languages could be classified into categories such as analytic or agglutinating was gradually abandoned in the latter part of the 20th century, but linguists still needed to distinguish between patterns such as *des Hauses* and *von dem Haus*, so they described the latter as ANALYTIC CONSTRUCTIONS. Since the 1990s, interest in the development of such constructions that use grammatical words has increased greatly, but generally under the rubric of GRAMMATICALIZATION (e.g. Lehmann 2016[1982]; Hopper & Traugott 1993). The term *analytic construction* has remained in use, but not prominently, and there has been very little general research on analyticization that uses this term (in fact, our dynamic term *analyticization* is likely to be unfamiliar to most linguists). In the 2000s, it became popular to consider whole languages from the point of view of "complexity" (e.g. McWhorter 2001; Miestamo et al. 2008), which is not unrelated to analyticity.

3. Analytic/synthetic as a synchronic notion

Before moving on, we need to point out that the term pair analytic/synthetic is problematic because it is based on the notion of a "word", which is itself poorly defined. In Matthews's (1997) *Concise dictionary of linguistics*, the terms are defined as in (1a) and (1b). In addition to presupposing a notion of "word", they presuppose "inflection" (1c), a concept that is itself based on "lexical unit" (= "word").

(1) a. *analytic form*
"form in which separate <u>words</u> realize grammatical distinctions that in other languages may be realized by <u>inflections</u>"

b. *synthetic form*
"form in which grammatical distinctions are realized by <u>inflections</u>"

[1] The morphological types *<u>analytic, agglutinating</u>* etc. have also been closely linked to the idea of a general

c. *inflection*
"any form or change of form which distinguishes different grammatical forms of the same lexical unit"

That the notion of "word" cannot be defined consistently across languages (other than orthographically, in languages with spaces between words) has long been known and was recently highlighted again by Haspelmath (2011) and Michaelis (2015). Spelling systems are arbitrary (historically accidental) to a substantial degree, so comparative linguists can hardly rely on them. But while the difference between the written unit "letter" and strictly linguistic units such as 'segment' and 'phoneme' is universally recognized, linguists often still seem to think that a written unit separated by spaces ("word") must reflect an important grammatical unit in (spoken) languages.

In Siegel et al.'s (2014) study of analyticity and syntheticity in creoles, they try to base their counts on the distinction between "free grammatical markers" and "bound grammatical markers" (along the lines of Greenberg 1960), claiming that the writing systems of the languages they are considering "generally represent a free marker as a separate word and a bound marker as a part of another word. Thus, we use this conventional orthography as a basis for our analysis." They are aware that there is a problem, but they do not actually do anything about it:

> "Of course, this is not ideal, as there is not necessarily an unequivocal relationship between spelling conventions and language structure (Haspelmath 2011). However, as a detailed phonological and morphosyntactic analysis of each language's texts would not be feasible, it is the best option." (Siegel et al. 2014: 53)

But the problem is actually deeper than they seem to realize, because there is no rigorous definition of "word" and "affix" that would coincide with linguists' intuition but be based on "a detailed phonological and morphological analysis of languages". Apparently our intuitions are based on the best known spelling systems, and these are not based on anything truly systematic.[2] (Note also that "bound" cannot be equated with "affixal", because clitics are generally regarded as bound elements that are not affixes.)

However, in the next section we will see that the tendency to replace older synthetic patterns by newer analytic patterns is real, despite the definitional difficulty noted here.

4. Synthetic/analytic in diachrony

Even if we cannot say whether a construction is synchronically synthetic, due to our inability to define "word" across languages, we can often determine whether it is diachronically old or freshly created from new material. Thus, the English *-er* comparative (e.g. *likeli-er*) is clearly old, while the *more* comparative (*more likely*) is clearly innovative;

[2] A reviewer asked whether the notion of phonological word might be helpful, but as Schiering et al. (2010) have shown, the diverse criteria for identifying phonological word domains typically do not coincide, so there are as many phonological words as criteria. Another reviewer asked whether the synchronic notion *synthetic* could not be based on some other notion such as uninterruptability. The answer is found in Haspelmath (2011): No other notion gives a good match with the folk concept of the "word"; in particular, most function items (e.g. prepositions and auciliaries) are uninterruptable in the sense that they cannot be separated from the form they occur next to.

and the French simple synthetic future (e.g. *elle chante-r-a* 'she will dance') is old, while the *aller* future (e.g. *elle va chanter* 'she is going to dance') is clearly innovative.

Thus, from a diachronic perspective it is possible to define an analytic pattern as in (2).

(2) *analytic pattern*:
a morphosyntactic pattern that was created from lexical or other concrete material and that is in functional competition with (and tends to replace) an older (synthetic) pattern

This means that if we are dealing with a language whose history is totally unknown, we cannot classify its patterns as synthetic or analytic. Since we know the history of French, we can perhaps say that it is "analytic" in comparison with Latin, but for a language such as Hup in Amazonia (Epps 2008) or Oko in Nigeria (Atoyebi 2010), about whose history we know very little, we cannot say whether they are analytic or synthetic.

We are thus primarily talking about a diachronic process of ANALYTICIZATION or "refunctionalization". The latter term is intended to refer to the new creation of a functional morpheme to express approximately the same grammatical notion or construction that has earlier been expressed by some other construction.[3] This restriction of the term *analytic* to a diachronic context may strike some readers as odd, but there is a general tendency in language typology to view typological generalizations as primarily diachronic (e.g. Bybee 2006; Cristofaro 2012), so even if we are not optimistic about pursuing Greenberg's (1960) research programme of a quantitative synchronic typology in terms of degrees of analyticity, we are convinced that there are strong general tendencies toward analyticization in language change.

Thus, the term *analytic* should be understood as roughly meaning "freshly re-grammaticalized". This definition works, because all patterns that have traditionally been called "analytic" are known to have been created from lexical or other concrete material; there does not seem to be any other way in which such patterns can come about. This definition is somewhat broader than the traditional purely synchronic definition, in that it also includes cases like the English past-tense marker *-ed* as in *play-ed*, which is generally thought to be a much newer pattern than the old pattern represented by ablauting verbs such as *sing/sang, write/wrote* (e.g. Lahiri 2000), and cases like the Sranan definite article *a* (deriving from *da* < *dat* < *that*), which is an analytic form when compared to English *the*, because it is based on a refunctionalization.

On the other hand, our definition of *analytic* is somewhat narrower than the synchronic view in that grammatical morphemes with no earlier counterparts, such as the Germanic and Romance definite articles, cannot be regarded as analytic. Even though the rise of definite articles in Romance and Germanic languages (and more generally in European

[3] Of course, there is no need for a new grammatical pattern to express exactly the same meaning as an older one, and indeed we typically seem to find meaning differences, e.g. between the older French synthetic future and the newer *aller* future (cf. Reinöhl & Himmelmann 2016), or between the German genitive and the possessive construction with *von*, or between the English definite article and the Sranan definite article *a* (cf. §4.1 below). The „identity" of the older and the newer patterns is not more than a relationship of typological matching, in terms of typological comparative concepts (Haspelmath 2010). The notion of „replacement" must be seen in this light, and it cannot be inferred that the newer construction is completely identical in function with the older construction.

languages, cf. Heine & Kuteva 2006: Chapter 3) has often been seen as part of the general tendency toward analyticity, we think that they should be treated differently, because otherwise we would have to include all kinds of other grammatical elements (e.g. discourse particles, focus particles, the new conditional mood) as well.

Changes whereby an earlier more compact pattern is replaced by a newer pattern based on a new periphrasis are quite widespread in European languages. A few types of changes are listed in (3), where the symbol ">>" means that the (approximate) function of a pattern tends to be replaced by a new pattern based on lexical or other concrete material.

(3) a. cases >> prepositions
b. comparative suffixes >> analytical comparatives
c. synthetic past tense >> analytic past tense
d. synthetic future tense >> analytic future tense
e. person-number suffixes on verbs >> personal pronouns
f. infinitive >> person-indexed subjunctive forms (Balkan languages)

In all these cases, the new analytic forms (originally) include transparent components, i.e. they arise from grammaticalization of concrete or lexical items. In a next step, analytic forms may then become compact again, be written together and thus become "synthetic again" (a process which can be called ANASYNTHESIS, cf. Haspelmath 2017), e.g.

(4) anasynthesis in the Romance future:

Latin	*canta-bit* 'will sing'	>>				
Romance		*cantare habet* 'has to sing'	>	*cantare ha*	>	*cantar-á* 'will sing' (Spanish)

(5) anasynthesis in the Romance adverb:

Latin	*fidel-iter* 'faithfully'	>>		
Romance		*fideli mente*	>	*fedelmente* 'faithfully' (Italian)

As mentioned earlier, analytic forms never arise in any other way; in particular, they do not arise from "desyntheticization" of a synthetic form ("antigrammaticalization" does not exist, with very few exceptions; Haspelmath 2004).

Just as new grammatical patterns can arise without any precursors (e.g. definite articles), old synthetic forms can disappear without any replacement, e.g. the old gender inflection in English adjectives, or the dual of older Germanic languages.

While analyticizations are really common in European languages, we will see in the next section that they are found even more widely in creole languages based on European languages.[4]

[4] Note that we do not follow authors such as Bickerton (1981) and Thomason & Kaufman (1988) in regarding creole languages as "new languages" with no historical continuity with their lexifier languages. Our position in this paper is more in line with authors such as Mufwene (2001) and Ansaldo et al.

5. Analyticizations occur very commonly in creoles

In this section, we will give examples of analyticizations in creole languages, from the database of the *Atlas of Pidgin and Creole Language Structures* (*APiCS*, Michaelis et al. 2013). The creoles we mention are all based on Germanic languages (English, Dutch) or Romance languages (Portuguese, Spanish, French). Table 1 lists the creoles from which our examples come, together with their major lexifier and the corresponding *APiCS* contribution.

creole language	lexifier	author in *APiCS*
African American English	English	Green 2013
Batavia Creole	Portuguese	Maurer 2013
Bislama	English	Meyerhoff 2013
Creolese	English	Devonish & Thompson 2013
Diu Indo-Portuguese	Portuguese	Cardoso 2013
Guadeloupean Creole	French	Colot & Ludwig
Guinea-Bissau Kriyol	Portuguese	Intumbo et al. 2013
Guyanais	French	Pfänder 2013
Haitian Creole	French	Fattier 2012
Jamaican	English	Farquharson 2013
Kriol	English	Schultze-Berndt & Angelo 2013
Mauritian Creole	French	Baker & Kriegel 2013
Negerhollands	Dutch	van Sluijs 2013
Palenquero	Spanish	Schwegler 2013
Papiá Kristang	Portuguese	Baxter 2013
Principense	Portuguese	Maurer 2013
Santome	Portuguese	Hagemeijer 2013
Seychelles Creole	French	Michaelis & Rosalie 2013
Sranan	English	Bruyn & van den Berg 2013
Sri Lanka Portuguese	Portuguese	Smith 2013
Tayo	French	Ehrhart & Revis 2013
Ternate Chabacano	Spanish	Sippola 2013
Tok Pisin	English	Siegel & Smith 2013
Vincentian Creole	English	Prescod 2013

The *APiCS* numbers following the subsection headings below are the feature numbers where the relevant information can be found.

5.1. Definite articles (*APiCS* 28, 9)

(6) Sranan *a* (e.g. *a pikin* 'the child') < *da* < English *that* (Bruyn 2009)

(7) Kriol *thet/thad* (e.g. *thad lif* 'the leaf') < English *that:*

(7.i) *Thad lif pat bla mukarra, im gud-wan bla so.*
DEM leaf part POSS river.pandanus 3SG good-ADJ DAT sore
'The leaf of the river pandanus is good for sores.' (Schultze-Berndt & Angelo 2013)

(2007), who stress the similarities between creolization and other kinds of contact-induced historical change.

(8) Haitian Creole =*la* (e.g. *nouvel=la* 'the news) < French *là* 'there'

5.2. Indefinite articles (*APiCS* 29, 10)

(9) Sranan *wan* < English *one* (Bruyn 2009), also in other English-lexified creoles

(10) Guinea-Bissau Kriyol *utru* < Portuguese *outro* 'other'

(10.i) *utru omi musulmanu*
a man Muslim
'a Muslim man' (Intumbo et al. 2013)

5.3. Plural markers (*APiCS* 22, 23)

(11) Seychelles Creole *bann* < French *bande* 'group'

(11.i) *Tou sa bann landrwa mon 'n ale.*
all DEM PL place 1SG PRF go
'It's to all these places that I have been.' (Michaelis & Rosalie 2013)

(12) Tok Pisin *ol* < English *all*

(13) Diu *tud* < Portuguese *tudo* 'all' (also Tayo *tule* < *tous les*)

5.4. Genitive markers (*APiCS* 38, 37)

(14) Vincentian Creole *fo* 'of' < English *for* (also in other English-lexified Caribbean creoles)

(14.i) *di pikni fo di woman*
ART child for ART woman
'the woman's child(ren)' (Prescod 2013)

(15) Tok Pisin *bilong* < *(that) belong (to)* (also Bislama *blong*, Kriol *bla*)

(16) Seychelles Creole *pour*, Tayo *pu* < French *pour* 'for'

5.5. Personal pronouns in subject or possessor function (*APiCS* 62)

(17) Santome obligatory subject person forms (cf. Portuguese optional subject pronouns)

(17.i) ***Bô*** *na sêbê kuma bô so kota mu mon fa?*
2SG NEG know COMP 2SG FOC cut 1SG.OBJ hand NEG
'Don't you know that it was you who cut my hand off?' (Hagemeijer 2013)

(18) Palenquero obligatory subject person forms (cf. Spanish optional subject pronouns)

(18.i) ***Ele*** ***e*** *ta trabahá.*
he he is working (Schwegler 2013)

(19) Diu Indo-Portuguese *d-**el*** 'his', lit. 'of him'
(cf. Portuguese possessive pronoun *seu/sua*)

(20) Guadeloupean Creole *timoun an **mwen*** [child of me] 'my child(ren)'

5.6. Accusative markers (*APiCS* 57)

(21) Batavia Creole *kung,* Papiá Kristang *ku* (< Portuguese *com* 'with'),
Ternate Chabacano *con* (< Spanish *con* 'with')

(22) Sri Lanka Portuguese *-pa* (< Portuguese *para* 'for') (cf. also Afrikaans *vir* < *voor* 'for')

(22.i) *eev* *vosa* *kuɲaadu-pa* *kada* *ɔɔra* *ki-lembraa*
1SG 2SG.GEN brother.in.law-ACC every time HAB-think.of
'I often think of your brother-in-law.' (Smith 2013)

5.7. Dative markers (*APiCS* 60, 61)

(23) Bislama *long* (< English *along*), cf. also Kriol *langa, la*

(24) Mauritian Creole *avek/ek* (< French *avec* 'with')

(24.i) *(av)ek* *ki* *sanla* *to 'n* *don* *larzaṅ* *la?*
with who that.one 2SG.PFV give money DEF
'To whom have you given the money?' (Baker & Kriegel 2013)

(25) Diu Indo-Portuguese *pe* (< Portuguese *para*)

(26) Papiá Kristang *ku*, Batavia Creole *kung*, Chabacano *con/kon* (cf. §2.6)

5.8. Future tense markers (cf. *APiCS* 48)

(27) Negerhollands *lo* < *loo* 'go' < Dutch *lopen* 'run'

(27.i) *Morək* *mi* *lō* *lō.*
tomorrow 1SG FUT go
'Tomorrow I will go.' (van Sluijs 2013)

(28) Seychelles Creole *pou* < French (*être*) *pour*

(29) Tok Pisin *bai* < English *by and by*

5.9. Past tense (or anterior) markers (*APiCS* 45)

(30) Seychelles Creole *ti* < French *était* 'was'

(31) Jamaican *wehn* < English *been* (also in many other English-lexified creoles)

(32) Principense *tava* < Portuguese *estava* 'was'

(33) Batavia Creole *dja* (perfective marker) < Portuguese *já* 'already'

(33.i) *fala kung ile ki eo* ***dja*** *teng aki*
tell OBJ 3SG COMP 1SG PFV be here
'tell him that I have been here' (Maurer 2013a)

5.10. Imperfective aspect markers (*APiCS* 46, 47, 48)

(34) Early Sranan *de* (PROG) < English *there*
(also in other Atlantic English-lexified creoles)

(34.i) *Hangri* ***de*** *killi mi.*
hunger PROG kill 1SG
'I am hungry (lit. Hunger is killing me).' (Bruyn & van den Berg 2013)

(35) Tok Pisin *i stap* (PROG) < English *stop*

(36) Seychelles Creole *pe*, Haitian Creole *ap* (PROG)
< French (*être*) *après* 'near, about (to do)'

(37) Haitian Creole *konn* (HAB) < French *connaître* 'know'

(38) Palenquero *asé* (HAB) < Spanish *hacer* 'do' (cf. Gullah *duhz*)

5.11. Causative construction

(39) Seychelles Creole *fer* < French *faire* 'do'
Mon ***fer*** *Zan manze* vs French *Je* ***fais*** *manger Jean*
'I make Jean eat'

The Seychelles construction in (39) uses the same lexical construction as the older French construction, and the older construction is not normally regarded as "synthetic". But by our definition, the Seychelles Creole pattern qualifies as analytic because it is clearly a new creation, as can be seen from the word order, where the causee (*Zan*) stands between the causative verb *fer* and the caused verb (*manze*). If the Creole construction continued the French construction, this ordering would not be possible.

Thus, we see that creole languages have a substantial additional number of analyticizations. In the next section we propose an explanation for this and link it to language contact in more general terms.

6. Analyticization is generally favoured by language-contact situations

The idea that analyticization is favoured by language contact is not new. In fact, in a sketchy form it is found in the very first work that discussed the analytic/synthetic distinction, August Wilhelm von Schlegel's (1818) work on the Provençal language (and its literature):

> "Mais cette transition au système analytique a lieu bien plus rapidement, et, pour ainsi dire, par secousses, lorsque, par l'effet de la conquête, il existe un conflit entre deux langues, celle des conquérans et celle des anciens habitans du pays. Voilà ce qui a eu lieu dans les provinces de l'empire occidental, conquises par les peuples germaniques, et en Angleterre lors de l'invasion des Normands. De la lutte prolongée de deux langues, dont l'une étoit celle de la grande masse de la population, l'autre celle de la nation prépondérante, et de l'amalgame final des langues et des peuples, sont issus le provençal, l'italien, l'espagnol, le portugais, le françois et l'anglais."[5]

More recently, Carlier et al. (2012) express it in the following way:

> "The more languages spread over large populations and involve frequent language contact between individuals who are related to each other by weak ties, the faster languages may evolve by regularizing mechanisms, utimately also reducing their morphological and grammatical systems." (Carlier, De Mulder & Lamiroy 2012: 292, citing Lupyan & Dale 2010; Trudgill 2011; see also McWhorter 2007)

But what explains increased analyticization in situations of increased contact? We can contrast two possible explanations for the increased tendency to analyticize in situations of language contact, what we call the "Loss-and-Repair Hypothesis" (cf. 40) and the Extra-Transparency Hypothesis (cf. 41). We will argue below that the second hypothesis is the correct explanation. But the Loss-and-Repair Hypothesis does not seem implausible either at first blush. In fact, the idea that languages tend to undergo "decay" and therefore need fresh material to reconstitute its grammar is quite old, going back at least to Schleicher (1860).

(40) Loss-and-Repair Hypothesis (e.g. Siegel 2008: 65-66; Good 2012)
In the transmission bottleneck of pidginization, inflectional and other non-salient grammatical markers are lost, because they cannot be acquired by adult learners. This leaves a void, and when pidgins turn into full-fledged languages again, they need to fill the gaps by new material deriving from content words.

---

[5] "But this transition to the analytic system took place more rapidly, and, so to speak, by jolts, when, due to conquest, a conflict between two languages arises, the language of the conquerors and the language of the earlier inhabitatnts of the country. This took place in the provinces of the Western Roman empire which were conquered by Germanic peoples, and in England after the Norman invasion. The extended struggle between two languages, one of which was the language of the great majority of the people, and other other the language of the ruling group, and the eventual merging of the languages and the peoples gave rise to Provençal, Italian, Spanish, Portuguese, French and English."

This hypothesis is similar to the therapeutic view of grammaticalization, which has been shown to be wrong for grammaticalization in general (cf. Lehmann 1985; Haspelmath 2000). It simply cannot be the case, for a number of reasons, that grammatical forms first reduce and then need to be strengthened again Thus, we favour the following hypothesis:

(41) Extra-Transparency Hypothesis
In social situations with many (or even mostly) adult second-language speakers, people need to make an extra effort to make themselves understood – they need to add extra transparency. This naturally leads to the overuse of content items for grammatical meanings, which may become fixed when more and more speakers adopt the innovative uses.

This is similar to the extravagance-based view of grammaticalization, which offers the best account of unidirectionality (Haspelmath 1999). Grammaticalization in ordinary situations is explained as due to occasional extravagant language use, when no special social circumstances are present. But in high-contact situations, no appeal to extravagance is necessary, and extra clarity can explain the stronger tendency to functionalize content items.

Creolists have recently tended to focus on transparency (Seuren & Wekker 1986; Leufkens 2013), simplification (McWhorter 2001; 2007; Parkvall 2008) or on the uniqueness of creole languages (Bakker et al. 2011), not on particularly fast grammaticalization. But the idea that many of the changes observed in creoles can be seen as accelerated grammaticalization has been expressed earlier (cf. Plag 2002), and despite some problems in distinguishing between true innovative grammaticalization and simple constructional calquing (cf. Bruyn 2009), we think that it is basically correct.

It is clear that simplification by adult learners cannot be invoked in all cases of analyticization, because this also occurs when the older synthetic form is simple to begin with (e.g. *the* >> *that* in §4.1, *faire manger Jean* >> *fer Zan manze* in §4.11, *a* >> *para* (dative marker) in Brazilian Portuguese, *de* >> *pou* (genitive marker) in §4.4). And analyticization also occurs in languages that do not have a high number of adult second language speakers, so something like extravagance needs to play a role in any event.

If extra transparency is the explanation for the very high degree of analyticization in creoles, then we may expect to find evidence for this also in languages that have undergone less extreme language contact changes, so this is what we briefly consider in the next section.

7. Further examples of increased analyticity in European varieties

Increased analyticity is apparently also found in some varieties of European languages that have undergone more contact influence than closely related varieties. In this section we give a few suggestive examples. Much more in-depth study would be required to really establish this, but we would like to include these examples because we believe that the creoles of *APiCS* are not completely unique but are just more extreme cases of a kind of phenomenon that is also found elsewhere. In particular, increased analyticity is found in a range of constructions in two languages that have been called "semi-creoles", Afrikaans and Colloquial Brazilian Portuguese (cf. Holm 2004).

7.1. Increased analyticity in Afrikaans

In Afrikaans, the old past tense disappeared and was replaced by the 'have' perfect (52), the dative is expressed by the preposition *vir* (53), and the genitive is exclusively expressed by the preposition *van* (54). For discussion, see Holm (2004).

(52) past tense
Ek het geskryf.
'I wrote, I have written'

(53) *vir*-dative

Hy	*het*	*dit*	*gister*	***vir***	*sy*	*broer*	*gewys.*
he	has	this	yesterday	to	his	brother	shown

'He showed it to his brother yesterday.'

(54) possessive *van*
de werken van Vondel 'Vondel's works'
(cf. Dutch *Vondel's werken*)

7.2. Increased analyticity in Brazilian Portuguese

Holm (2004) also discusses Brazilian Verbacular Portuguese, where independent pronouns regularly occur in addition to subject inflectionon verbs, or even replace it (55), independent pronouns replace object clitics (56), and relative clauses are used with resumptive pronouns (57).

(55)

eu parto	*você parte*	*ele parte*	*nós parte*	*eles parte*
'I leave'	'you leave'	'he leaves'	'we leave'	'they leave'

(56)

ela	*chamou*	***eu***
she	called	me

'she called me' (cf. Portuguese *chamou-me*, with object clitic)

(57)

o	*aluno*	*que*	*eu*	*conheço*	*o*	*pai*	***dele***
the	student	that	I	know	the	father	of.him

'the student whose father I know' (lit. '...that I know his father')

7.3. Increased analyticity in Bulgarian

Among the Slavic languages, Bulgarian and Macedonian show the most drastic changes away from the Proto-Slavic patterns. Most strikingly, genitive and dative case are replaced by the preposition *na* (originally 'on') (58-59), and the old comparative degree forms are replaced by the new particle *po-* (60). Hinrichs (2004) even claims that Bulgarian is a creolized form of Old Bulgarian.

(58) *Petar dade kniga-ta* ***na*** *Ivan.*
Petar gave book-DEF on Ivan
'Peter gave book-the to Ivan.' (cf. Russian *Ivan-u* [Ivan-DAT])

(59) *kola-ta* ***na*** *Marija*
car on Marija
'Marija's car' (cf. Russian *Mari-i* [Maria-GEN])

(60) *po-umna*
COMPR-smart
'smarter' (cf. Russian *umn-ee* [smart-COMPR])

There are probably more cases of increased analyticity in European languages. As mentioned briefly above, Carlier et al. (2012) link differences in the pace of grammaticalization in various Romance languages (French, showing more advanced grammaticalization, compared to Italian and Spanish) to language contact. The Eastern Scandinavian languages (notably Swedish and Danish) also show higher degrees of analyticization than western languages (notably Icelandic and Faroese).

8. Conclusion

In this paper, we briefly reviewed the history of the distinction between analytic and synthetic patterns and observed that it cannot be based on a synchronic definition, because we cannot define words and affixes in a consistent way. We therefore proposed a diachronic definition of an *analytic pattern* as a morphosyntactic pattern that was created from lexical or other concrete material and that is in functional competition with (and tends to replace) an older (synthetic) pattern. The main empirical observation is that analyticization is particularly frequent in European-based creole languages, and we proposed an explanation in terms of extra transparency: In social situations with many adult second-language speakers, people need to make an extra effort to make themselves understood, i.e. they need to add extra transparency. This naturally leads to the overuse of content items for grammatical meanings, and thus to analyticization. It remains to be seen to what extent this explanation can account for differences within European languages (some relevant observations were made in §6), and whether it can also account for developments in languages outside Europe.

References

APiCS: Michaelis et al. (2013)

Ansaldo, Umberto, Stephen Matthews & Lisa Lim (eds.). 2007. *Deconstructing creole*. Amsterdam: Benjamins.

Atoyebi, Joseph Dele. 2010. *A reference grammar of Oko: A West Benue-Congo language of north-central Nigeria*. Köln: Köppe.

Bakker, Peter, Aymeric Daval-Markussen, Mikael Parkvall & Ingo Plag. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26(1). 5–42. DOI:10.1075/jpcl.26.1.02bak.

Bickerton, Derek. 1981. *Roots of language*. Ann Arbor: Karoma Publishers.

Bruyn, Adrienne. 2009. Grammaticalization in creoles: Ordinary and not-so-ordinary cases. *Studies in Language* 33(2). 312–337.

Bybee, Joan. 2006. Language change and universals. In Ricardo Mairal & Juana Gil (eds.), *Linguistic universals*, 179–194. Cambridge: Cambridge University Press.

Carlier, Anne, Walter De Mulder & Béatrice Lamiroy. 2012. Introduction: The pace of grammaticalization in a typological perspective. *Folia Linguistica* 46(2). 287–301.

Cristofaro, Sonia. 2012. Cognitive explanations, distributional evidence, and diachrony. *Studies in Language* 36(3). 645–670. DOI:10.1075/sl.36.3.07cri.

Epps, Patience. 2008. *A grammar of Hup*. (Mouton Grammar Library, 43). Berlin: Mouton de Gruyter.

Good, Jeff. 2012. Typologizing grammatical complexities, or: Why creoles may be paradigmatically simple but syntagmatically average. *Journal of Pidgin and Creole Languages* 27(1). 1-47.

Greenberg, Joseph H. 1960. A Quantitative approach to the morphological typology of language. *International Journal of American Linguistics* 26(3). 178–194.

Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6). 1043–1068.

Haspelmath, Martin. 2000. The relevance of extravagance: A reply to Bart Geurts. *Linguistics* 38(4). 789–798.

Haspelmath, Martin. 2004. On directionality in language change with particular reference to grammaticalization. In Olga Fischer, Muriel Norde & Harry Perridon (eds.), *Up and down the cline: The nature of grammaticalization*, 17–44. Amsterdam: Benjamins.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.

Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1). 31–80.

Haspelmath, Martin. 2017. Revisiting the anasynthetic spiral. In Bernd Heine & Heiko Narrog (eds.), *Grammaticalization and language typology*. (to appear)

Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie. 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.

Heine, Bernd & Tania Kuteva. 2006. *The changing languages of Europe*. Oxford: Oxford University Press.

Hinrichs, Uwe. 2004. Ist das Bulgarische kreolisiertes Altbulgarisch? In Uwe Hinrichs (ed.), *Die europäischen Sprachen auf dem Wege zum analytischen Sprachtyp*, 231–242. Wiesbaden: Harrassowitz.

Holm, John. 2004. *Languages in contact: The partial restructuring of vernaculars*. Cambridge: Cambridge University Press.

Hopper, Paul J. & Elizabeth C. Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Jespersen, Otto. 1894. *Progress in language: With special reference to English.*

Kortmann, Bernd & Benedikt Szmrecsanyi. 2011. Parameters of morphosyntactic variation in World Englishes: Prospects and limitations of searching for universals. In Peter Siemund (ed.), *Linguistic universals and language variation*, 264–290. Berlin: De Gruyter Mouton.

Lahiri, Aditi 2000. Hierarchical restructuring in the creation of verbal morphology in Bengali and Germanic: Evidence from phonology. In Aditi Lahiri (ed.), *Analogy, Levelling, Markedness: Principles of Change in Phonology and Morphology*, 71-123. Berlin: Mouton de Gruyter.

Lehmann, Christian. 1985. Grammaticalization: Synchronic variation and diachrinic change. *Lingua e Stile* 20(3). 303–318.

Lehmann, Christian. 2015[1982]. *Thoughts on grammaticalization*. 3rd edition. Berlin: Language Science Press.

Leufkens, Sterre. 2013. The transparency of creoles. *Journal of Pidgin and Creole Languages* 28(2). 323–362.

Lupyan, G. & R. Dale. 2010. Language structure is partly determined by social structure. *PLoS ONE* 5(1). e8559.

Matthews, P. H. 1997. *The concise Oxford dictionary of linguistics*. Oxford: Oxford University Press.

McWhorter, John H. 2001. The worlds simplest grammars are creole grammars. *Linguistic Typology* 5(2-3). 125–166.

McWhorter, John H. 2007. *Language interrupted: Signs of non-native acquisition in standard language grammars*. Oxford: Oxford University Press.

Michaelis, Susanne. 2015. Inflectional complexity in creole languages: Evidence from the Atlas of Pidgin and Creole Language Structures. Paper presented at the SLE conference in Leiden, September 2015.

Michaelis, Susanne, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.). 2013. *The Atlas of Pidgin and Creole Language Structures*. Oxford: Oxford University Press. (apics-inline.info)

Mufwene, Salikoko S. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.

Parkvall, Mikael. 2008. The simplicity of creoles in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 265–285. Amsterdam: Benjamins.

Plag, Ingo. 2002. On the role of grammaticalization in creolization. In Glenn Gilbert (ed.), *Pidgin and creole linguistics in the 21st century*. New York: Peter Lang.

Reinöhl, Uta & Himmelmann, Nikolaus. 2016. 'Renewal': A figure of speech or a process sui generis? Ms., University of Cologne.

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt Brace.

Schiering, René, Balthasar Bickel & Kristine A. Hildebrandt. 2010. The prosodic word is not universal, but emergent. *Journal of Linguistics* 46(3). 657–709.

Schlegel, August Wilhelm von. 1818. *Observations sur la langue et la littérature provençales.*

Schleicher, August. 1860. *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*. Weimar: Böhlau.

Seuren, Pieter & Herman Wekker. 1986. Semantic transparency as a factor in creole genesis. In Pieter Muysken & Norval Smith (eds.), *Substrata versus universals in creole genesis*, 57–70. Amsterdam: Benjamins.

Siegel, Jeff. 2008. *The emergence of pidgin and creole languages*. Oxford: Oxford University Press.

Siegel, Jeff, Benedikt Szmrecsanyi & Bernd Kortmann. 2014. Measuring analyticity and syntheticity in creoles. *Journal of Pidgin and Creole Languages* 29(1). 49–85.

Szmrecsanyi, Benedikt. 2009. Typological parameters of intralingual variability: Grammatical analyticity versus syntheticity in varieties of English. *Language Variation and Change* 21(03). 319–353.

Szmrecsanyi, Benedikt. 2016. An analytic-synthetic spiral in the history of English. In Elly van Gelderen (ed.), *Cyclical change continued*, 93–112. Amsterdam: Benjamins.

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.

WALS: Haspelmath et al. (2005)