PLENARIES:

1.   Making complexity simpler: a theoretical evaluation of some unfounded ideas and unnecessary confusions

Wouter Kusters, University of Leiden, The Netherlands

In this talk I will discuss several theoretical issues concerning complexity. First I will argue that complexity can only be a *relational* concept. That is, a concept that depends on the relationship between aspects of language on the one hand and aspects of an evaluator on the other hand. This evaluator may be a linguist (or linguistic theory), a hearer, a speaker, a first or second language learner and so on. Much of the confusion around complexity can be cleared when we make explicit what kind of evaluator we have in mind when using the word 'complexity'.

Next I will propose a definition of complexity which is based on the concept of a 'generalised outsider', which is an idealisation of a second language learner who is primarily interested in a particular language for purposes of communication instead of expressing identity or other non-referential meanings.

With this operationalisation of complexity I will strongly argue against the unfounded claim that 'complexity in one domain would correlate with lack of complexity in another'. Nevertheless, some domains of language, like morphology, are more suitable for research in complexity than others, which, however does not mean that the complexity to find complexity would imply that there would be no complexity differences. Next, I will describe what morphological phenomena can be named complex.

Finally I will bring the concept of complexity from the individual to the social level. In other words, will the number and influence of 'generalised outsiders' in a speech community have a simplifying effect on a language? In my second talk I will discuss this question with an example of language change in several Quechua varieties.

2.   Making complexity complexer: a discussion of the complexities of simplification in four Quechua varieties

Wouter Kusters, University of Leiden, The Netherlands

In this talk I will show how the notion of complexity can be used in language change research. I will discuss one of the four language groups of my dissertation, namely Quechua. Older and more conservative Quechua varieties as spoken in Peru (Cuzco and Ayacucho Quechua) have a quite abundant and complex inflectional verbal morphology. Inflectional categories in these Quechua varieties are among others, object and subject agreement, tense, number, and inclusiveness. Although Quechua

morphology has a quite agglutinative outlook, there are several affixes that that are fused, split, and/or display allomorphy and homonymy.

Modern Quechua varieties vary with respect to the extent that they have been in contact during the last five hundred years with ' generalised outsiders', as defined earlier. We may plot the various Quechua varieties along this dimension and make a prediction about which varieties should have been simplified most in comparison with the original older Quechua variety. I will examine first whether these predictions are borne out for the Quechua varieties of Bolivia, Argentina and Ecuador. We will see that this is partially the case, but that the exact path of change depends also on other factors like substrate influence, and particular properties of Quechua structure itself.

After discussing this particular example, I will compare Quechua simplification with other kinds of simplification in Arabic, Scandinavian and Swahili, in order to arrive at a tentative typology of simplification paths.

When there is time left, I will spend a few words on the time-bopundedness of the peculiar circumstances that seem to foster simplification, like rapid language expansion, in combination with limited access to native varieties of the languages. It remains to be seen whether such kinds of limited access will occur in the future of the globalised world.

3. TBA

Ritva Laury California State University, Fresno, USA; University of Helsinki, Finland

4. TBA

John McWhorter, Manhattan Institute; University of California Berkeley, USA

5. TBA

John McWhorter, Manhattan Institute; University of California Berkeley, USA

SECTION PAPERS:

6.   Complex phonotactics: two-element consonant clusters in Slovak

Zsuzsanna Bárkányi, Hungarian Academy of Sciences, Research Institute for Linguistics, Hungary

CANCELLED

7.   "Redundant" marking more or less complex?

Angela Bartens & Niclas Sandström, University of Helsinki, Finland

In order to address this question, posed for instance in McWhorter (2005:44), we will apply the morphosyntactic 4-M model developed by Myers-Scotton & Jake (2000a,b, 2001) to data from Cape Verdean Creole Portuguese and Popular Brazilian Portuguese. Cape Verdean Creole Portuguese has been less strongly restructured than so-called prototypical creoles. Popular Brazilian Portuguese, far from being a creole, has nevertheless undergone some morphosyntactic restructuring as well.

We have taken our example from the domain of nominal plural marking where the two varieties present similar morphosyntactic configurations. Previous accounts of Kabuverdianu and Popular Brazilian Portuguese nominal plural marking mention the occurrence of (at least) one inflectional marker per NP. The 4-M approach enables us to account more precisely for the occurrence of this inflection. Definiteness, an "early system morpheme" according to the 4-M model, is in Portuguese linked to the expression of the notion of plurality which is marked on the first element of the NP to the left of the nucleus. All other plural marking, required by nominal congruence in Standard Portuguese, is lost during the kind of restructuring Kabuverdianu and PBP have undergone, as we are dealing with "late system morphemes" in Standard Portuguese.

We argue that in spite of being strictly rule-governed, the reduction of inflectional plural marking in Cape Verdean Creole Portuguese and in Popular Brazilian Portugues vis-à-vis Standard Portuguese constitutes a case of overall loss of morphosyntactic complexity which can be partly explained by the shallow time-dephth of the existence of the varieties in question. "Redundant" marking is therefore found to be morphosyntactically more complex. Whether this is true on the cognitive level as well is quite another issue (we suggest it is not).

References:

McWhorter, John H. 2005. Defining Creole. Oxford: Oxford University Press.
Myers-Scotton, Carol & Janice L. Jake. (Eds.) 2000a. Testing the 4-M Model.
    International Journal of Bilingualism 4.

Myers-Scotton, Carol & Janice L. Jake. 2000b. Four types of morphemes: Evidence from aphasia, codeswitching, and second language acquisition. Linguistics 38:6, 1053-1100.
Myers-Scotton, Carol & Janice L. Jake. 2001. Explaining aspects of codeswitching and their implications. In: J. Nicol (ed.). One mind, two languages: Bilingual language processing. Oxford: Blackwell, 84-116.

8. Segmental and Syllabic Complexity: The case of R in Haitian Creole

Parth Bhatt & Emmanuel Nikiema, University of Toronto, Canada

Creole languages. McWhorter (2001) has for example proposed that Creole grammars are the world's simplest grammars with the implication that such language should contain few, if any, highly marked structures. The purpose of this paper is to examine a particular case of a highly marked phonological structure – the behaviour of the segment /R/ in French based Creoles. In this paper we will outline the specific case of /R/ in Haitian Creole (henceforth HC).

The segment /R/ in HC exhibits an interesting distributional pattern. /R/ is systematically pronounced in prevocalic position (1), but is systematically "deleted/ in post-vocalic position word-finally (2a) or word-internally (2b).

(1)	HC	Gloss	Etymon
	[ʀɛv]	"dream"	(rêve)
	[ʀeta]	"late"	(retard)
	[ʀivjɛ]	"river"	(rivière)
	[maʀi]	"husband"	(mari)
	[tʀavaj]	"work"	(travail)

(2)	a. word-final position			b. word-internal position		
	HC	Gloss	Fr. Etymon	HC	Gloss	Fr. Etymon
	[bɛ]	"butter"	(beurre)	[tɔti]	"turtle"	(tortue)
	[flɛ]	"flower"	(fleur)	[ʒunal]	"journal"	(journal)
	[fɔ]	"strong"	(fort)	[dɔmi]	"to sleep"	(dormir)
	[ʃalɛ]	"heat"	(chaleur)	[abit]	"referee"	(arbitre)
	[riʃa]	"wealthy"	(richard)	[bab]	"beard"	(barbe)

For words in (2), however, if a vowel occurs after the post-vocalic /R/, as is the case in morphologically related forms, the segment is pronounced as in (3).

4

(3)   a. non-complex forms   b. morphologically related forms
      bè [bɛ]   "butter"   bere [bɛwe]   "to butter, to baste"
      flè [flɛ]   "flower"   fleri [flɛwi]   "to blossom"
      fè [fɛ]   "iron"   fere [fɛwe]   "to shoe (a horse)"

The reappearance of post-vocalic /R/ has lead previous researchers such as d'Ans
(1968:53, 76), Tinelli (1981) and Cadely (1994) to propose that /R/ is present
underlyingly in HC lexical forms. Further evidence of the underlying presence of /R/
is provided by compensatory vowel lengthening in words with (deleted) post-vocalic
/R/ as in (4) (cf. Valdman 1978 :61) and the blocking of regressive vowel nasalization
(5) for words which contain /R/ in the French etymon (Hazaël-Massieux 1972 ; Bhatt
& Nikiema 2000).

(4)   Vowel   length   contrast   (5) Blocking of regressive nasalization
corresponding to a lexical contrast

a. short vowel		b. long vowel		Underlying form	HC Pronunciation (oral vowel)	French Etymon
[nɛf]	"nine"	[nɛːf]	"nerve"	/kɔRnø/	[kɔn] /*[kɔ̃n]	"corne"
[sɔt]	"stupid"	[sɔːt]	"to go out"	/aRmwaR/	[amwa] /*[ãmwa]	"armoire"
[pɔt]	"pot"	[pɔːt]	"door"	/l'aRme/	[lame] /*[lãme]	"armée"
[pak]	"Easter"	[paːk]	"parc"	/ʃaRmø/	[ʃam] /*[ʃãm]	"charme"
[pɛ]	"peace"	[pɛː]	"father"	/fɔRmø/	[fɔm] /*[fɔ̃m]	"forme"

In this paper, we address aspects that have not been addressed in previous analyses,
namely how to account for the phonological effects produced by this deleted segment
(vowel length contrasts and blocking of regressive nasalization). Are these effects of
the syllabic representation of this segment? Or are they attributable to the internal
structure of the segment? Nikiema and Bhatt (2003) propose that post-vocalic /R/ is
syllabified in the nucleus with the preceding vowel, forming a light diphthong as in
(6) below. This constitutes, however, a highly complex syllabic structure and
constitutes an apparent counter-example to the Creole simplicity hypothesis.

```
(6a) [flɛ] "flower"                    (6b) [bab] "beard"
     O     R                               O   R   O   R
    /\     |                               |   |   |   |
   / \     N                               |   N   |   N
  /   \    |                               |   |   |   |
  x  x     x                               x   x   x   x
  |  |    /\                               |  /\   |
  f  l   ɛ R                               b  a R  b
```

This also raises the question of whether other HC segments also show a similar
complex syllabic representation or whether this is a property of /R/ only. Bhatt and

Nikiema (2000) have argued that nasal consonants in St. Lucia Creole have similar properties and are to be represented in the same manner. This suggests that in French-based Creoles such as HC and St. Lucia Creole, sonorants create complex structures in post-vocalic position. This in turn raises the issue as to why these segments are syllabified in this manner and why they are open to a wide variety of phonological realizations. Is this a result of the internal structure of the segment or the syllabic structure?

In short, the behaviour of /R/ in HC requires a complex syllabic representation which is apparently atypical of Creole languages. This in turn suggests that languages that have an apparently simple segment inventory or morphology may in fact show considerable complexity at the syllabic and morphophonological levels.

References

Bhatt, Parth & Emmanuel Nikiema (2000): Le statut de la nasalité en créole de Ste-Lucie in Revue québécoise de linguistique, 28, 23-45.

Cadely, Jean-Robert (1988): Représentations syllabiques et distribution des diphtongues en créole haïtien, in Études créoles , Vol. XI (1):9-40.

Cadely, Jean-Robert (1994): Aspects de la phonologie du créole haïtien, Unpublished Ph. D. Dissertation, Université du Québec à Montréal.

Cadely, Jean-Robert (1996): On Nasality in Haitian Creole: the Case of the Definite Determiner, paper delivered at the11th Biennial Conference, Society for Caribbean Linguistics.

D'Ans, André-Marcel (1968) Le créole français d'Haiti, The Hague : Mouton.

Hazaël-Massieux, Guy (1972): Phonologie et phonétique du créole de Guadeloupe, Unpublished Doctoral Dissertation, Université de Paris, Sorbonne Nouvelle.

McWhorter, John (2001) The world's simplest grammars are Creole grammars, Linguisic Typology, 5, 388-412.

Nikiema, Emmanuel (1999b): Government-Licensing and Cluster Simplification in Quebec French in Canadian Journal of Linguistics, Vol . 44(4), 327-357.

Nikiema, Emmanuel & Parth Bhatt (2003) "Two types of R deletion in Haitian Creole", The Phonology and Morphology of Creole Languages, Ingo Plag (ed.), Tubingen: Niemeyer, pp. 43-69.

Tinelli, Henri (1981) Creole phonology, The Hague: Mouton.

Valdman, Albert (1978a): Le créole: statut et origine. Paris: Klincksieck.

9.   The complexity of grammatical metaphors

Aleth Bolt, Vrije Universiteit Amsterdam, The Netherlands

This paper focuses on the psychological effects of grammatical metaphors, which have not been studied in depth as yet. Grammatical metaphor subsumes a number of non-congruent structures which result from shifts between functional categories in the lexico-grammar, by which, for instance, processes (congruently worded as verbs) are reworded as nouns (Halliday 1994: 341).

Several suggestions have been made in the literature that support the idea that a text becomes more difficult because of metaphorical style. Halliday (1994: 350), for instance, suggests that grammatical metaphors may be intrinsically more complex than their congruent equivalents and that "the least metaphorical wording will always

be the one that is maximally simple." This hypothesised relationship between metaphoricity and complexity has been supported in several studies on language acquisition (Derewianka 2003; Youping 2001).

In the literature on text composition, grammatical metaphor has been assigned a very negative value. It is often argued that metaphorical expressions such as nominalisations render a text inaccessible, dull and abstract and, related to this, adversely influence comprehension. However, textual advice on metaphoricity seems to be predominantly based on intuitions and experience rather than on psycholinguistic research.

In this paper, I will present the results of two reading experiments that seem to provide psycholinguistic evidence for the assumption that at least one type of grammatical metaphor, nominalised infinitives, may be inherently more difficult than their congruent equivalents. An example of a nominalised infinitive (1) and of its congruent counterpart (2) would be:

1.   De Britse regering heeft zich verantwoordelijk gedragen door het jarenlang vrij laag houden van de buitenlandse schulden.
        [The British government has itself responsibly behaved by the for many years relatively low keep_INF of the foreign debts.]
2.   De Britse regering heeft zich verantwoordelijk gedragen door jarenlang de buitenlandse schulden vrij laag te houden.
        [The British government has itself responsibly behaved by for many years the foreign debts relatively low to keep]

It was investigated whether there would be differences in processing ease, comprehension and recall between nominalised infinitives and their congruent equivalents. Both construction types could occur naturally, without there being a clear difference in interpretation. In the experiments, journalistic texts and sentences were presented on a computer screen in a self-paced reading paradigm. The data showed that the nominalised infinitives were processed significantly slower than their congruent equivalents. Contrary to expectations, the data do not suggest a decline in comprehension and recall performance.

The results show a need to further look into the role of genre expectations in the processing and representation of grammatical metaphors. I hypothesise that nominalised infinitives and so-called boxed constructions are comprehended better and processed with greater ease in policy papers than in newspaper articles.

References:

Derewianka, B. (2003). Grammatical metaphor in the transition to adolescence. In: Van den Bergen et al. (eds.) Grammatical metaphor. Amsterdam: Benjamins, 185-219

Halliday, M.A.K. (1994). Introduction to functional grammar. London: Edward Arnold.

Youping, C. (2001). The use of grammatical metaphor by EFL learners and their language proficiency. Unpublished PhD thesis, National University of Singapore.

10. Complex patterns in phonological systems

Christophe Coupé, Egidio Marsico, François Pellegrino, Laboratoire Dynamique Du Langage, Lyon, France

The notion of complexity is intrinsically relative; we always define an object as more or less complex than another according to one or several properties of these objects. Thus, the main question is twofold: how to identify the relevant properties at the right level of description and how to calculate their complexity.

In phonology, both aspects of this question are still a challenge. First, what is the correct level at which complexity should be measured: features, segments or systems? And second, are the relevant properties to be considered from articulatory or perceptual perspectives, or even both?

To illustrate this point, typological studies (Liljencrants and Lindblom, 1972, Maddieson, 1984, Vallée & al., 2002) have shown for example that phonological systems tend to be organized in such a way to reach a balance between ease of articulation and perceptual salience. Therefore, any measure of complexity of phonological systems would have to consider both dimensions. However, we still don't have any objective and quantitative way to calculate complexity on either articulatory or perceptual dimensions, let alone when considered together.

Previous work by Lindblom and Maddieson, (1988) has tried to define a complexity scale for obstruent consonants, based mainly on intuition. The goal of this paper is to address their suggestion to find a way to define this scale of complexity from "quantitative phonetic considerations". Furthermore, we will not restraint our investigations to obstruents, but consider complete phonological systems.

Our data come from the UPSID database (Maddieson, 1984, Maddieson and Precoda, 1990), which contains the phonological systems of a representative sample of 451 languages. These systems are described by means of 100 features and 900 segments. We consider these three levels (features, segments, systems) as a good approximation of the extent of possible phonological elements. Our approach tries to capture the intra and inter levels constraints by defining a set of indices (Marsico & al., 2004) and different measures of phonological distance and complexity (Marsico& al., 2002).

Our previous results show that phonological systems tend to be organized around basic (neutral) segments. When the number of segments increases systems tend to saturate the existing dimensions (economy). Finally, it seems important that a system keeps an important adaptive potential by "recruiting" highly generative basic segments.

Our recent developments relies on the construction of different networks of features or segments, and the application of refined complexity measurements to these networks, in order to characterize the internal structure of phonological systems. More precisely, these networks capture in which way systems make use of features to constitute segments and to generate oppositions between them. To measure the networks complexity (one network per language), we rely on a recent proposal found in Claussen, (2004), called "offdiagonal complexity" based on node-node link correlations.

Our preliminary results show that the complexity of a system is not a function of the number of segments contrary to what could be intuitively expected. Our presentation will include more quantitative and qualitative analyses as well as a possible phonological distance derived from graph complexity measurements.

References

Claussen, Jens Christian. 2004. Offdiagonal complexity: a computationally quick complexity measure for graphs and networks. q-bio.MN/0410024

Liljencrants, Johan and Bjorn Lindblom. 1972. Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language* 48: 839-862.

Lindblom, Björn and Ian Maddieson. 1988. Phonetic universals in consonant systems. In C. N. Li & L. H. Hyman (eds.) *Language, Speech and mind: Studies in Honor of Victoria A. Fromkin*, 62-78. Beckenham: Croom Helm.

Maddieson, Ian and Kristin Precoda. 1990. Updating UPSID. *UCLA Working Papers in Phonetics* 74: 104-111.

Maddieson, Ian. 1984. *Patterns of sounds.* Cambridge, MA: Cambridge University Press.

Marsico Egidio, Christophe Coupé, François Pellegrino & Jean-Marie Hombert. 2002. *Vers une mesure de la complexité des espaces phonologiques.* Journées d'Etudes Linguistiques. Nantes, France.

Marsico Egidio, Ian Maddieson, Christophe Coupé, & François Pellegrino. 2004. Investigating the 'hidden' structure of phonological systems. Proceedings of the 30th Berkeley Linguistics Society Meeting. Berkeley, USA.

Vallée, N., L.-J. Boë, J.-L. Schwartz, P. Badin, & C. Abry. 2002. The Weight of Phonetic Substance in the Structure of Sound Inventories. *ZAS Papers in Linguistics* 28: 145-168.

11. Changes in structural complexity as a diachronic phenomenon

Östen Dahl

In Dahl (2004), I make a number of distinctions pertaining to the notion of complexity as applied to language. One fundamental distinction is that between "resources" and "regulations". A speaker of a language has an inventory of morphemes, words, phrases and constructions from which he or she can choose in building an utterance. But the process of building utterances is also regulated in various ways largely by what would traditionally be called "rules of grammar". For instance, in English, we may combine a noun phrase and an intransitive verb to form a sentence, but English grammar demands that we put the noun phrase before the verb and that the verb agree with the (subject) noun phrase (with certain exceptions, as this clause shows). In Dahl (2004), I try to show how diachronic processes such as those usually called "grammaticalization" increase the complexity of linguistic regulations, mainly by introducing non-linearity into the expression. I do not have very much to say about the possible application of complexity considerations to the resource side of language. In many case, an increase in resources, such as additions to the lexicon, would seem to affect the complexity of a language in a somewhat trivial way. On the other hand, grammatical constructions may vary with respect to structural complexity a transitive verb phrase is more complex than a transitive one etc. Also, even if they involve the same number of elements, one construction may have the potential of introducing more complex structure than another thus constructions that involve recursivity could

be said to be more complex than those that do not. It is relatively trivial to state that human languages exhibit more structural complexity in this sense than most other communication systems. What is more contentious is whether diachronic developments can entail changes in the complexity of languages in this sense. In the paper, I will discuss how the rise of "tight" grammatical constructions may involve changes in complexity of various kinds, in particular increase in structural complexity, and how this entails replacing implicit discourse-level links with explicit grammatical ones.

Reference

Dahl, Östen. 2004. The Growth and Maintenance of Linguistic Complexity. Amsterdam: Benjamins.

12. Complexity in nominal plural allomorphy – a contrastive survey of ten Germanic languages

Antje Dammel (Mainz), Sebastian Kürschner (Freiburg i. Br.)

The Germanic languages are characterized by a broad range of morphological systems with different degrees of complexity. In nominal allomorphy the most extreme cases can be observed in Icelandic and English. Icelandic has a highly inflectional system involving a large number of noun classes, strong use of stem alternations and a high degree of idiosyncratic allomorphy. English is at the opposite end of the spectrum, with only one productive noun class characterized by the -*s*-plural, and only a small amount of nouns with idiosyncratic behavior. The other languages we examined can be ordered between these two poles, with Faroese, German, Luxemburgish, and Swedish closer to the complex Icelandic pole, and Afrikaans, Dutch, Frisian, and Danish closer to the simpler English pole.

In a contrastive analysis of these ten Germanic standard languages, we have accounted for the complexity of plural systems. For this reason, we identified the factors responsible for morphological complexity in this domain and ordered them in relation to our intuitive sense of a complexity scale involving the languages examined. The factors for complexity can be identified on formal grounds and then involve

a) the techniques for plural marking (cf. suffixation, stem alternation, subtraction, zero marking and combinations of these) and their amount of use in single languages,
b) the number of plural markers within one noun (single vs. multiple exponents) and its amount of use,
c) the basic systems of base form inflection vs. stem inflection,
d) the formal system of noun classes (cf. plural-classes vs. case-number-classes),
e) the amount of (ir-)regularity in plural marking, i.e. the number of nouns with idiosyncratic plural formation.

In addition to these formal features, we found assignment principles – which account for the use of specific allomorphs – highly relevant. These involve

a) formal principles as
   (i) the stem's final sounds,
   (ii) prosodic assignment (e.g. syllable number, accentuation criteria),
b) derivational suffixes,
c) semantic features (e.g. animacy),
d) gender.

With all plural systems considered, we accounted for the mentioned principles in relation to the degree of complexity. The formal criteria and assignment principles were arranged on a scale to account for their impact on morphological complexity. Basic technical considerations for the measurement of complexity in noun systems were thereby identified. Our presentation will deal with the identification process and the theoretical groundings for these considerations. Their use in complexity measurement will be the topic of future research.

13. Morphological complexity and language typology. The case of language change in Hungarian by language contact.

Casper de Groot, University of Amsterdam, The Netherlands

Studies on Hungarian spoken outside Hungary (Austria, Slovakia, Ukraine, Romania, Vojvodina, Prekmurje, Slovenia, United States and Australia as presented in Fenyvesi ed. 2005) reveal interesting information on the change of Hungarian as a minority language. Most differences – if not all – between standard Hungarian (HH) and Hungarian outside Hungary (HO) can be explained in terms of language contact where HO takes over features from adjacent languages, which all happen to belong to the Indo-European language family. De Groot (2005) argues that a number of the differences observed follow linguistic universals and implicational hierarchies, and that the co-occurrences of changes can actually be explained in terms of universals or hierarchies.

However, for the systematic change from synthetic to analytical expressions in the varieties of HO there are no typological studies available that could form the basis for an explanation. Where HH uses morphologically complex, synthetic forms, HO shows a preference for the use of morphologically simplex, analytical forms, as in table 1. Note that the preference for analytical forms applies in all three major components of morphology, i.e. inflection (modality), derivation (reflexive and causative), and compounding.

Table 1. Synthetic versus analytical expressions in HH and HO.

	HH (synthetic)	HO (analytic)
Modality	Ki-me-*het*-ek? out-go-MOD-1SG 'May I go out?'	Ki *tud*-ok men-ni? out be.able-1SG go-INF 'May I go out?'
Reflexive	Szépít-*kez*-ett beautify-REFL-PAST.3SG.INDEF 'She beautified herself.'	Szépít-ette *magá*-t beautify-PAST.3SG.DEF oneself-ACC 'She beautified herself.'

Causative	Meg-rajzol-*tat*-ta	*Hagy*-ta	a szék-et
	PVB-draw-CAUS-PAST.3SG.DEF	permit-PAST.3SG.DEF the chair-ACC	
	a szék-et.	rajzol-ni.	
	the chair-ACC	draw-INF	
	'S/he had the chair designed.'	'S/he had the chair designed.'	
Compounding	tag-létszám	tag-ok	létszám-a
	member-number	member-PL	number-3SG.POSS
	'number of members'	'number of members'	

As for the general view on morphological change, consider Hock and Joseph (1996:183): "The fate of morphology from Sanskrit to its modern descendants gives credence to the common belief that languages tend to develop in cycles: from isolating to agglutinating, from agglutinating to inflectional (through amalgamation of different affixes into one), from inflectional to isolating (through sound change and analogy), and so on."

HO still contains the morphological properties of an agglutinating language (cf. *szépítette, széket,* and *létszáma* in Table 1. The change from HH to HO then does not follow the chain of gradual changes attested in many languages, but rather seems to be a change from morphologically complex to morphologically simplex, i.e. from an agglutinating language with multi-morphemic words to an agglutinating language with less-multi-morphemic words.

If we take complex and simplex to be a typological parameter as well, the two types of development are two different realizations:

(1)   complex   →   simplex
   (i)   agglutinating → inflectional → isolating
   (ii)   multi-morphemic → less-multi-morphemic

A second point I would like to make is the following. At the same time that HO seems to develop morphological simplex forms, the syntax of the varieties gets more complex. Moreover from a linguistic typological point of view, HO develops into a type of language which is rather the exception than the standard as HH.

14. Optional Determiners in Child Language and Headlines

Joke de Lange, Utrecht University, the Netherlands, Sergey Avrutin, Utrecht University, the Netherlands, Maria Teresa Guasti, University Biccocca, Italy

Children acquiring Dutch omit determiners more often, and for a longer period of time, than Italian children of the same age although both languages require overt determiners in normal adult speech (e.g. Guasti et al 2003). Various existing accounts suggest that the difference is based on differences in the morphosyntactic system (Roeper and de Villiers, 1995) or on the semantic properties of the nominal system (Chierchia, 1998). At the same time, Stowell (1999), Avrutin (1999) and Schutze (1997) pointed out that such omissions are allowed in special registers, such as Headlines:

Kamer Verlengt Fiscale Regeling Voor Film Met Een Jaar   [Dutch]
Government extends tax arrangement for film with a year   (3determiners omitted)

Questions:

- Do Dutch and Italian headlines exhibit the same differences as observed in child Dutch and Italian?
- What are the language specific properties that are reflected in the cross-linguistic differences both in child speech and the headlines?
- Why does it take longer for a Dutch child to reach the adult stage w.r.t. determiner use?

Method:

CHILDES database (MacWhinney and Snow, 1985). 3 Italian, 3 Dutch children. Headlines: 3 Dutch, 2 Italian newspapers, 2000 headlines for each language.

Results:

Language	Determiners in Headlines % produced	Determiners in child speech % produced (example: Stage 2 from Guasti et al 2004)	Structurally correct use of determiners, %
Dutch	12.9	46	100
Italian	72.1	83	100

Analyses:

1. We argue that in both languages children, at an early age, possess grammatical knowledge of the determiner system, which is reflected in their correct use, whenever a determiner is produced (e.g. correct structural position, presence of a nominal complement).

2. We propose a processing theory of complexity that is based on the recent work on the application of the information theory to lexical access (Baayen et al. 2004, Kostic 2003, Moscoso del Prado Martin, 2003). We show how the rich morphosyntacic paradigm of Italian determiners increases their information load compared to Dutch (in the technical sense proposed in Shannon and Weaver 1948, Baayen et al 2004, see also Caramazza and Schiller 2003, among others). As the child processing system takes some time to mature, elements with a higher information load become available before less informative elements, as confirmed by recent results (e.g. Baayen et al 2004). Hence the difference between Dutch and Italian child speech production.

3. The same explanation holds for special registers (e.g. headlines) where the amount of information determines which elements are more likely to be omitted in order to facilitate processing of the entire utterance (e.g. Kostic et al, 2003).

4. In our presentation we further discuss the role of the structural position (e.g. subject vs. object, the role of the left periphery (Rizzi 1997, Cinque 1990) for the possibility of omissions, both for child language and headlines, as also observed in Stowell (1999).

Figure 1

p<.0001   p<.0001

Determiner Production in Dutch and Italian Headlines and Child Speech
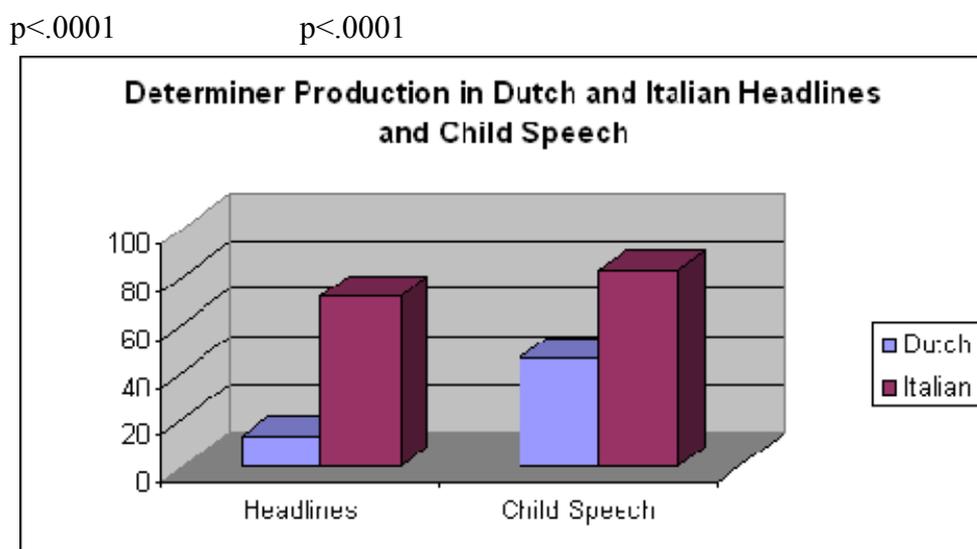
15. To what extent does the difficulty of processing relative clauses parallel typological complexity?

Claire DELLE LUCHE, Frédérique GAYRAUD, Bruno MARTINIE, Fanny MEUNIER-HOEN, Laboratoire Dynamique du Langage UMR5596 CNRS & Université Lumière-LYON2, France

Introduction

Center-embedded object RCs have proved to be more difficult to process than their subject counterparts. Typical explanations accounting for this difference include the following:

(i)   Canonical order: According to Bever (1970) and Caplan & Waters (1999) the reason why Object RCs are more difficult to process is that object extraction results in a non canonical OSV word order.
(ii)   Functions parallelism: Object RCs are more difficult to process because of the difference of functions of the first NP, which is subject of the matrix clause but object of the embedded clause (e.g. Sheldon 1974).).
(iii)   The depth of embedding explanation predicts that the greater distance between the filler and the gap, the more difficult the relative is to parse (Fodor 1977, Hawkins 2004).
(iv)   Working memory cost: Gibson (1998) claims that the complexity difference between Object and Subject RCs can be explained by the amount of memory and integration resources required by these structures.

These psycholinguistic explanations are mirrored by a typological scale proposed by Keenan & Comrie (1977), the Accessibility Hierarchy (AH) setting the different relativizable positions as Subject < Object < Indirect Object < Oblique < Genitive < Comparative. Hawkins (2004) extends the AH's predictions by identifying further cognitive implications.

While many studies investigate the processing of Subject and Object RCs, little work is done concerning the other RCs on the AH (except Keenan & Hawkins 1987). The goal of this study is two-fold:

to investigate the processing of other functions than Subject and Object RCs and to test the cognitive foundations of the AH by adding object genitive RCs (henceforth OG);
to disentangle between the different explanations (i)-(iv).

Hypotheses

The different theories presented above predict different results:

(i) and (ii) predict that Subject RCs are easier to process than any others;
(iii) and (iv) predict the following hierarchy : S < O < OG

Method

To test these hypotheses we conducted a self-paced moving window experiment on 30 French speaking participants. They were presented three types of RCs (1-3), followed by a comprehension question. We collected error rates and reading times per word.

(1)   S: Le barman qui plaint le client poivre la viande. (The bartender that pities the client peppers the meat)
(2)   O: Le barman que le client plaint poivre la viande. (The bartender that the client pities...)
(3)   OG: Le barman dont le client plaint la situation poivre la viande. (The bartender of which the client pities the situation...)

Results

Our results suggest that hierarchical predictions are not validated: O, and OG RCs show no difference, neither in terms of the number of errors, nor in terms of time of

processing. Hence it is a dichotomic explanation that fits the data best: subject RCs are easier to process, they generate less errors, and their verbs are shorter to read, contrary to Gibson's model, to the predictions related to the depth of embedding and to the AH.

Bibliography

Bever, T. G. 1970. The Cognitive Basis for Linguistic Structures. In Cognition and the Development of Language, ed. J.R. Hayes, 279-362. New York: Wiley.

Caplan, D., and Waters, G. S. 1999. Verbal working memory and sentence comprehension. Behavioral and Brain Sciences: 77-126.

Gibson, E. 1998. Linguistic complexity: locality of syntactic dependencies. Cognition 68: 1-94.

Hawkins, J. A. 2004. Efficiency and Complexity in Grammars. Oxford: Oxford University Press.

Just, M. A., and Carpenter, P. A. 1992. A capacity theory of comprehension: Individual differences in working memory. Psychological Review: 122-149.

Keenan, E. L., and Comrie, B. 1977. Noun Phrase accessibility and Universal Grammar. Linguistic Inquiry: 63-99.

Keenan, E. L., and Hawkin, S. 1987. The Psychological Validity of the Accessibility Hierarchy. In Universal Grammar, ed. E. Keenan, 60-85. London: Croom Helm. Sheldon, A. 1974. On the Role of Parallel Function in the Acquisition of Relative Clauses in English. Journal of Verbal Learning and Verbal Behaviour: 272-281.

16. The balance between word complexity and syllable complexity

August Fenk, Gertraud Fenk-Oczlon, University of Klagenfurt, Lisa Fenk, University of Vienna, Austria

The most unassuming "definition" of the complexity of a certain system or unit refers to the number of elements or components of this unit (a); more "complex" conceptions may also take into account the complexity of the components (b), the number of different component types (Changizi 2001), the number of possible interactions between the parts (Simon 1996), the number of different rules determining these interactions, etc. This paper confines itself to the relatively simple levels (a) and (b) * higher complexity means a higher number of (more complex) components * and analyses the complexity of syllables as a function of the complexity of words. The relevant complexity measures are three metric variables: the size of syllables in terms of the number of phonemes, and the size of words in terms of the number of syllables and the number of phonemes. Such data concerning the variation within a single language (German) are for instance available in Menzerath (1954); crosslinguistic data are available in previous studies by the authors (Table 4 in Fenk & Fenk-Oczlon 1993 and its continuation in Fenk-Oczlon & Fenk 1999, Table 1) and, for very few languages, also in Fucks (1964).

In the 1999-study mentioned Menzerath's data for German as well as our crosslinguistic data were analysed in the sense of a linear regression (syllable complexity Y as a function of word complexity X) and, in addition, as a quadratic and as a logarithmic function which resulted in somewhat higher determination coefficients. Present study uses another statistical program in the search for the "best fitting" function. Regarding the intra-German data, this search revealed an exponential decay as a function with an extremely precise fit: a determination coefficient of 0.995 when word complexity was defined in n of syllables, and 0.994 when defined in n of phonemes. In crosslinguistic data sets the fits obtained are of course lower, but the model of an exponential decay seems to be a very good approximation as well.

The results are discussed within the framework of a correlational model of language variation and especially with respect to Menzerath's first law (I: the bigger the whole, the smaller its parts) and his second law (II) saying that the variability of the size of components is lower within bigger units. An integrating view seems possible from the perspective of a relatively constant flow of linguistic information, as indicated by an almost ideal proportionality between informational content in bits (Y) and the number of syllables per word (X): A doubling of the number of syllables, for instance, means a doubling not only of the information transmitted, but also of the time for processing.

References

Changizi, M.A. (2001). Universal Scaling laws for Hierarchical Complexity in Languages, Organisms, Behaviors and other Combinatorial Systems. J. theor. Biol. 211, 277-295.

Fenk, A. & Fenk-Oczlon, G. (1993). Menzerath's Law and the Constant Flow of Linguistic Information. In R.Köhler & B.Rieger (Eds.) Contributions to Quantitative Linguistics. Dordrecht: Kluwer.

Fenk-Oczlon, G. & Fenk, A. (1999). Cognition, Quantitative Linguistics, and Systemic Typology. Linguistic Typology 3, 151-177.

Fucks,W. (1956). Die mathematischen Gesetze der Bildung von Sprachelementen aus ihren Bestandteilen. Nachrichtentechnische Fachberichte 3, 7-21.

Menzerath, P. (1954). Die Architektonik des deutschen Wortschatzes. Bonn: Dümmler.

Simon, H.A. (1996 [1962]). The Architecture of Complexity: Hierarchic Systems. Proceedings of the American Philosophical Society 106 (Dec.1962), 467-482.

17. Linguistic complexity from a systemic perspective

Gertraud Fenk-Oczlon & August Fenk, University of Klagenfurt, Austria

Combinatorial systems change as they become more expressive. An increase in expression complexity is achieved by increasing the number of component types, or, in addition, by increasing the number of components per expression (Changizi 2001). We will analyse crosslinguistic patterns of language variation from this perspective and will put stress on possible compensatory processes: An above average complexity of a certain language in a certain domain or property may go hand in hand with economies in other domains or properties.

The following set of mutually dependent significant crosslinguistic correlations (a sample of 34 languages in Fenk-Oczlon & Fenk 1999) points to such trade-offs, probably due to time limits in the mechanisms involved in language processing:

a)   The more syllables per clause, the fewer phonemes per syllable.
b)   The more syllables per word, the fewer phonemes per syllable.
c)   The more syllables per clause, the more syllables per word.
d)   The more words per clause, the fewer syllables per word.

Taking correlation (b) as a starting point we assume that a high average syllable complexity, i.e. a high mean number of phonemes per syllable, is a presupposition for a high intra-language variability of syllable complexity, and that such a high intra-language variability is a presupposition for a large inventory of monosyllabic words. This last inferential step in other words: the higher a language's number of syllable types (X), the higher the inventory of monosyllabic words (Y). In order to test this assumption, we collected relevant data from Menzerath's (1954) statistical descriptions of 8 languages. Within this small sample we found that an extreme position in both variables (X, Y) was occupied by English (with 43 syllable types and 6856 monosyllabic words) and the opposite extreme by Italian (with 8 syllable types and 93 monosyllabic words). A crosslinguistic correlation between X and Y revealed a coefficient of $r = +.895$ (sign., $p<0.01$).

This new result regards the level of phonology. The following chain of (in part rather hypothetical) arguments starts from this level and takes English as an example. English has a high number of monosyllabic words; the construction of so many different monosyllabic words requires a high number of syllable types (our correlation above), and the construction of many different syllable types also requires a big phonemic inventory. All that means that English is very complex in the domain of phonology. Therefore it can "afford" lower levels of complexity in other domains: It has relatively short words (one of our previous results), and short words are not apt to encode too many grammatical categories (a positive crosslinguistic correlation between n of syllables per word and n of cases?); instead, the relevant information is encoded in - a rather rigid - word order. This lower complexity in morphology and syntax may again be associated with higher complexity * and higher cognitive costs * at the level of semantics: Languages with a high proportion of monosyllabic words will show a strong tendency to homonymy, etc. The discussion of these results and assumptions refers to mechanisms and strategies * e.g. "chunking" (Simon 1974) * of our information processing system.

References

Changizi, M.A. (2001). Universal Scaling laws for Hierarchical Complexity in Languages, Organisms, Behaviors and other Combinatorial Systems. J. theor. Biol. 211, 277-295.
Fenk-Oczlon, G. & Fenk, A. (1999). Cognition, Quantitative Linguistics, and Systemic Typology. Linguistic Typology 3, 151-177.
Menzerath, P. (1954). Die Architektonik des deutschen Wortschatzes. Bonn: Dümmler.
Simon, H.A. (1974). How big is a chunk? Science 183, 482-488.

18. Recent advances in information and network theory about the complexity of human language

Ramon Ferrer i Cancho, Università 'La Sapienza', Italy

Here we review recent advances in information (Ferrer i Cancho, 2005b-d, Ferrer i Cancho & Solé; 2003) and network theory (Ferrer i Cancho et al., 2004; Ferrer i Cancho et al., 2005) that have shed new light on what makes human language complex. We may study the complexity of human language from two points of view: (a) what makes human language special with regard to simpler communication systems? (b) what kind of solutions human language may have adopted when communication principles are in conflict? As for (a), there is a long tradition arguing that the crux of human language is our ability to construct an infinite amount of sentences by combining words from a finite lexicon (Hauser et al., 2002). If one tries to be mathematically precise, the fact is that real sentences have finite length. Thus, a system with recursion, i.e. a system capable of constructing an infinite amount of sentences (Hauser et al., 2002), may not be distinguishable, in practice, from a system capable of constructing a finite but astronomically large amount (e.g. the number of atoms of the universe) of sentences. Recently, it has been proposed that (almost) connectedness, a property of the network of word syntactic associations provides, immediately, an astronomically large (but finite) amount of sentences and the skeleton for recursion (Ferrer i Cancho et al., 2005). Connectedness is a prerequisite for recursion and constitutes an intermediate level of complexity. We propose four levels of language complexity based on the presence of combinatorics, connectedness and recursion.

As for b), recent models have shown that the organization of word frequencies in human language is consistent with a particular balance between maximizing the information transfer and saving the cost of communication, which are principles in conflict. Far from that particular balance, the system cannot communicate at all or it communicates at the maximum cost (Ferrer i Cancho & Solé, 2003; Ferrer i Cancho, 2005d). Word frequencies organize according to Zipf's law (Zipf, 1949) in that particular balance. It has been shown that Zipf's law may lead to a rudimentary form of language with connectedness in a wide set of conditions (Ferrer i Cancho et al., 2005a). If a system following Zipf's law favours maximizing the information transfer too much, it becomes the most expensive communication suddenly (Ferrer i Cancho, 2005b). It can be shown that connectedness is destroyed at the same time. Those models suggest that a fundamental ingredient of the complexity of human language is achieving a proper balance between conflicting communication principles. The alternatives are no communication, a very expensive communication system and the loss of connectedness. We should not complain about word ambiguity and be very grateful to our limited brain resources.

References:

Ferrer i Cancho & Solé, R. V. (2003). Least effort and the origins of scaling in human language. Proc. Nat. Acad. Sci. USA. 100, 788-791.

Ferrer i Cancho, R., Solé, R. & Köhler, R. (2004). Patterns in syntactic dependency networks. Phys. Rev. E. 69, 051915.

Ferrer i Cancho, R., Riordan, O. & Bollobás, B. (2005). The consequences of Zipf's law for syntax and symbolic reference. Proc. R. Soc. B 272, 561-565.

Ferrer i Cancho, R. (2005b). Decoding least effort and scaling in signal frequency distributions. Physica A. 345, 275-284.

Ferrer i Cancho, R. (2005c). The variation of Zipf's law in human language. Eur. Phys. J. B (in press).

Ferrer i Cancho, R. (2005d). Zipf's law from a communicative phase transition. Submitted to Eur. Phys. J. B.

Hauser, M. D., Chomsky, N. and Fitch, W. T. (2002). The faculty of language: what it is, who has it, and how did it evolve? Science 298, 1569-1579.

Zipf, G. K. (1949). Human behavior and the principle of least effort. Reading: Addison-Wesley.

19. The effect of inferential complexity on the processing and appreciation of humorous utterances

Kurt Feyaerts & Geert Brône, University of Leuven, Belgium

This paper reports on the aims and setup of a research project (start September 2005), which investigates the effect of complexity-related variables on the cognitive processing and appreciation of one-liners and one-panel cartoons. The project revolves around two major research questions. First, are variants of humorous utterances, which by specific parameters differ in their inferential complexity, processed in ways consistent with this complexity difference? Second, are utterances that differ in their inferential complexity appreciated in ways consistent with this complexity difference?

Of particular interest for this project are conceptually-related parameters such as (inferred) event structure (McKoon & Macfarland 2000; Gennari & Poeppel 2003), but also conceptual *effects* such as plausibility, imageability, and novelty (Giora 2003; Giora et al. 2004). Characterizing the processing of humor as 'problem solving activity' (Attardo et al. 2002; Giora 1991) motivates the interpretation of complexity in terms of 'inferential complexity'. This paper focuses on relevant parameters which define this notion: inferential distance, scene plausibility, imageability and contextual presence.

With regard to the first parameter, **inferential distance**, Feyaerts & Brône (2004) demonstrate that extensions and modifications of expressive utterances are often created by increasing the inferential distance between the linguistic/pictorial stimulus and the envisaged interpretation. In terms of the underlying event structure, an increasing inferential distance is apparent in the sentences from 1a) to 1e) (Keenan et al. 1984).

1a)   Joey's big brother punched him again and again.
        The next day his body was covered in bruises.
    1b)   Racing down the hill Joey fell off his bike.
        The next day his body was covered in bruises.
    1c)   Joey's crazy mother became furiously angry with him.
        The next day his body was covered in bruises.
    1d)   Joey went to a neighbor's house to play.

      The next day his body was covered in bruises.
1e)   Joey told his girl friend that he wanted to marry her.
      The next day his body was covered in bruises.

The impact of the variable **scene plausibility** (also Gruber & Gibson 2004) can be illustrated by the variants 1d) and 1e), which differ with respect to the plausibility of the second sentence occurring as an effect of what is predicated before. **Scene imageability** reflects the extent to which the referent of a linguistic item evokes a mental image. Compared to 2a), 2b) is hypothesized to be ranked higher because of the specification of the way in which the depicted scene might work out.

2a)   You can break down doors with this man (target: stupid)
      2b)   You can break down doors with his head (target: stupid)

The impact of **contextual presence** is investigated by looking at priming effects caused by 'conceptual markers' in the stimulus: lexical elements belonging to the domain matrix of the target concept. In 3a) the item *fell* triggers the target concept "physical injuries" whereas 3b) lacks such an element and can therefore be considered to be inferentially more complex.

3a)   He fell while racing down the hill
      The next day his body was covered in bruises.
      3b)   He kept on racing down the hill:
      The next day his body was covered in bruises.

References

Attardo, Salvatore, Christian F. Hempelmann and Sara Di Maio 2002 Script oppositions and logical mechanisms: Modeling incongruities and their resolutions. Humor. International Journal of Humor Research 15-1, 3-46.

Feyaerts, Kurt and Geert Brône 2004 Expressivity and metonymic inferencing: stylistic variation in non-literary language use. Style 38-2.

Gennari, Silvia and David Poeppel 2003 Processing correlates of lexical semantic complexity. *Cognition* 89, 27-41.

Giora, Rachel 1991 On the cognitive aspects of the joke. *Journal of Pragmatics* 16, 465-485.

Giora, Rachel 2003 *On Our Mind. Salience, Context, and Figurative Language.* Oxford: Oxford University Press.

Giora, Rachel, Ofer Fein, Ann Kronrod, Idit Elnatan, Noa Shuval and Adi Zur. 2004 Weapons of mass distraction: Optimal innovation and pleasureratings. *Metaphor and Symbol* 19-2, 115-141.

Gruber, Jeff and Edward Gibson 2004 Measuring linguistic complexity independent of plausibility. *Language* 80, 583-590.

McKoon, Gail and Talke Macfarland 2000 Externally and internally caused change of state verbs. *Language* 76, 833-858.

20. Are "complex sentences" really difficult to process?

Frédérique GAYRAUD & Bruno MARTINIE, Laboratoire Dynamique du Langage
UMR 5596, CNRS & Université Lumière-Lyon2, France

Introduction

The well-known survey of Goldman-Eisler (1972) established the link between syntactic properties and temporal organization of speech production. It was shown that the higher the degree of integration between clauses, the less pauses are encountered at the junction.

Besides it has commonly been admitted that subordination equals syntactic complexity (Kroll 1977, Halliday, 1989, Beaman, 1984Š). According to a traditional conception of subordination, (1), (2) and (3) are equally complex.

In our view subordination has to be considered as a continuum reflecting various degrees of syntactic integration (Koch, 1995). Thus it is easy to demonstrate that examples (1) to (3) display an increasing degree of integration.

It is raining since she is taking her umbrella
Peter is happy because it is raining
he wants you to come

Hypotheses

H1. The more subcategorized a clause is, the less pauses should occur between the matrix and the subordinate clause since the latter is required by the lexical properties of verbs. By contrast, pseudo-embedded clauses are the less integrated. Thus we predict the following scale: subcategorized < modifier < pseudo-embedding.
H2. if H1 stands, then, pseudo-embedding clauses should exhibit planning characteristics of simple sentences.

Method

Twenty French speaking adults were asked to produce spoken texts and recorded. Pause were extracted and two locations were identified: matrix clause initial and embedded clause initial.

The degree of syntactic embedding was characterized as Subcategorization (3), Modification (2), and Pseudo-embedding (1).

Results

If the (pseudo)-embedded clause initial position is considered, we find an effect of the degree of integration on the duration and number of pauses (respectively $F(2,214 = 9,464$; $p =.0001$ and $F(2,214 = 12,170$; $p<.0001$). Pseudo-embedding clauses display more and longer pauses around the grammatical delimitator than the others types of clauses.

If we compare subordinate clause initial location and simple sentence initial location we observe an overall effect of the type of sentence on the number and duration of pauses (respectively $F(3,277) = 13,217$; $p<.0001$ and $F(3,277) = 22,338$; $p<.0001$). Furthermore, subcategorized display shorter pauses than pseudo-embedding

(p<.0001) and than simple sentences (p<.0001). Similarly, they show less pauses than pseudo-embeddings (p<.0001) and than simple sentences (p<.0001).

Discussion

The first Hypothesis is globally confirmed except for sentence initial position suggesting that other locations should be investigated. The second result confirms globally H2 : pseudo-embedding indeed resembles simple sentences, which reflects the lack of syntactic integration in both cases.

In conclusion, supposedly complex sentences are not necessarily complex to process. For instance, subcategorized clauses give rise to less and shorter pauses. According to our criteria, it means that they are less complex to process. Our interpretation is that on the linguistic ground, it is a consequence of valence which specifies both syntactic and semantic selections. The cognitive implication is that this choice reduction lightens planning work.

References

Beaman, Karen. 1984. Coordination and subordination revisited: Syntactic complexity in spoken and written narrative discourse. In *Coherence in spoken and written discourse*, ed. Deborah Tannen, 45-80. Norwood, N.J.: Ablex.

Goldman-Eisler, Frieda. 1972. pauses, clauses, sentences. *Language and Speech* 14:103-113.

Halliday, Michael A.K. 1989. *Spoken and written language*. Oxford: Oxford University Press.

Koch, Peter. 1995. Subordination, intégration syntaxique et "oralité". *Etudes romanes* 34:13-43.

Kroll, Barbara. 1977. Combining ideas in written and spoken English: A look at subordination and coordination. In *Discourse across time and space. Southern California Occasional papers in linguistics n° 5*, eds. E. Keenan and T. Bennett, 69-108. Los Angeles, CA University of Southern California.

21. How Complex Are Isolating Languages?

David Gil, Max Planck Institute for Evolutionary Anthropology, Germany

For those who adhere to the position that all languages are of equal overall complexity, a particular challenge is presented by languages of the isolating type, since, by definition, such languages are characterized by substantially reduced complexity in one major structural domain, that of morphology. If indeed all languages are of equal overall complexity, isolating languages must somehow compensate for their simpler morphology by being more complex in other structural domains, presumably pertaining to syntax and semantics. One proposed compensatory mechanism would entail that languages without nominal case-marking systems may have more grammatical constraints on the linear order of words; however, counterarguments have been made against the existence of such a purported correlation. More generally, it is hard to envisage the kind of regulatory mechanism

that would compensate for the morphological simplicity of isolating languages by increasing their syntactic and semantic complexity, given that most or all theories of grammar are endowed with an essentially modular architecture, distinguishing between different domains that are subject to different kinds of rules and principles.

In order to test the equal-overall-complexity hypothesis, this paper presents some results of a cross-linguistic experiment designed to gauge one central aspect of syntactic and semantic complexity in isolating languages, pertaining to the association operator. In previous work, I have argued that in at least one isolating language, Riau Indonesian, the compositional semantics is based on a single simple rule of association, in which when two expressions X and Y with meanings P and Q respectively are combined, the meaning of the collocation X Y is derived from that of its constituent parts by means of the association operator, A (P, Q), which says that the meaning of X Y is associated with the meanings of X and Y in a maximally underspecified manner. In Riau Indonesian, the association rule takes the place of a wide range of other construction-specific rules generally posited for other languages, including, in particular, rules making reference to thematic roles such as agent, patient and so forth; the prominence of associational semantics thus contributes to the characterization of Riau Indonesian syntax and semantics as being simpler than that of other languages.

The association experiment is designed to measure the availability of apparently associational interpretations: interpretations that appear to be obtainable from the association operator without reference to thematic roles or other semantic categories. Two kinds of apparently associational interpretations are sought: (a) those in which what looks like a bare noun preceding a bare verb is interpreted as the patient (rather than the agent); and (b) those in which what looks like a bare noun in construction with a bare verb is interpreted as an oblique argument or even a non-argument (in the absence of prepositions or other such markings). The experiment is ongoing; as of April 2005, over 1000 subjects had been tested in over a dozen languages.

The results of the experiment are as follows. Whereas in non-isolating languages such as English and Hebrew, the availability of apparently associational interpretations is near zero, in isolating languages the availability of such interpretations varies considerably from language to language: in creole languages it ranges from 10% in Papiamentu to 34% in Bislama; in colloquial varieties of Malay/Indonesian it ranges from 25% in Kuala Lumpur Malay through 43% in Riau Indonesian to 56% in Jakarta Indonesian; and among other isolating languages of Southeast Asia it ranges from 41% in Vietnamese to 66% in Minangkabau. The existence of isolating languages such as Minangkabau with high availability of apparently associational interpretations shows that such languages do not necessarily compensate for their impoverished morphology by means of increased syntactic and semantic complexity.

Such languages would thus appear to warrant the global characterization as simpler languages; in doing so, they cast doubt on the naive belief that all languages are of equal overall complexity.

22. Applying Concept Map to develop a new Technical Writing technique for enhancing Reading comprehension performance

Hung Ha, Wollongong University, Australia

Key words: concept map, technical writing, spatial text technical writing, reading comprehension, language ambiguity.

In today dynamic information technology era, technical documents are becoming bigger and are updated frequently more than ever. As a result, people have to spend a huge mount of time and efforts to digest these technical documents. Hence, to help people to read effectively and understand quickly these technical documents, a technical writing technique was devised for technical writers to produce good technical documents. Technical writing forces writers to use simple and accurate words to produce cohesive and easy-to-be-understood technical documents.

However, because technical writing uses the prose word-after-word text to produce technical documents, the final technical documents still suffer the essence problems of the prose text. These problems are language ambiguity and ineffective key concept manipulation. Nevertheless, interestingly, these problems can be solved effectively by concept map technique. Concept map is a graphical representation of a network of the interrelationship concepts. Because of its 2-dimension spatial concept representation, concept map can help to limit the language ambiguity problem, to manipulate the concepts effectively, and to perceive the information quickly. Thus, concept map has a potential to replace the technical writing technique and to produce efficient technical documents.

The research described in this thesis aimed to propose a new more effective technical writing technique by applying concept map to map the whole technical document. The value of mapping the whole technical document is that it can help readers to perceive the knowledge in that technical document faster. However, the current concept map technique got many problems such as time consuming, effort consuming, and diagrammatic messiness. These problems are the big hurdles for mapping the whole technical documents. So, to be able to map the whole document, a new technical writing technique should be devised to limit these problems. The new technical writing technique should also retain the goodness of linear sentence-after-sentence prose technical document.

This thesis suggested a new technical writing technique called temporarily as "spatial text technical writing" (STTW) that can solve these problems. Basically, STTW is the combination of concept map and technical writing. STTW is used to produce "spatial text technical document" (STTD).

The research used a variety of examples to demonstrate the ability of concept map as well as STTW in reducing the language ambiguity problem and facilitating concept manipulation. Besides, the research also used some very basic experiments in psychology to demonstrate the ability of concept map in helping human to perceive the information quicker.

As mentioned above, the amount of technical knowledge ever increase, sooner or later human has to solve the problem how to consume effectively this huge knowledge amount. If we do not change the way we represent the text in technical document, it will be very hard for people to enhance their reading comprehension

performance of technical documents. So, a new type of technical writing that can represent spatially the whole technical document is very valuable for learners to understand the technical knowledge quickly.

23. Modeling the Effects of Sublexical Features and Letter Order in Transposition Priming Experiments

Tarik Hadzibeganovic, Karl-Franzens-University Graz, Austria

To successfully recognize words during reading, proper encoding of the position/order in which the letters occur must be established. Such an encoded input is then matched against stored representations in memory in order to recognize a given word. The present study opens up new possibilities of theorizing by taking into account the importance of and the difference between a preattentional and an attentional stage in n-letter word processing. The difficulty/easiness of processing n-letter words with permuted/re-arranged inner letters in a typical transposition priming paradigm (TPP) is addressed. A procedure for obtaining estimates of letter order is introduced, and then the effects of relevant independent variables such as phonotactic probability (positional segment probability and biphone probability), orthographic and phonological neighborhood density and similarity in languages with different degrees of orthographic depth (varying orthography-phonology consistency) are explored. One non-classical, holistic quantum-like information processing system and one classical psychophysical framework for modeling these effects are developed. It is shown that words in transposition priming tasks can be treated as sums of letter-vectors, with higher weights given to the first and last letters (the initial letter receives the highest level of bottom-up input and the final letter is not inhibited by a subsequent letter during the induction of serial firing), and if permutation of inner letters (up to a given recognition threshold defined by the stimulus order intensity) has little effect on meaning realization (and final recognition), the word vectors differing by inner letters should be almost parallel. Quick associative, input-triggered, pattern identification and retrieval processes leave the outer letters - due to their previously mentioned advantages - unchanged, but the inner ones become symmetrized, indicating that words are representable as objects subject to symmetry processing. Additionally, it is demonstrated that both conscious (CP) and automatic processing (AP) contribute to performance in TPP experiments, with attention having a selective effect on CP and lexicality influencing AP. Time permitting, we will discuss the status of an experiment designed to test the physical validity of these conclusions.

24. Creoles and Complexity in Numeral Systems

Harald Hammarström, Chalmers University, Sweden

Numerals provide an ideal testing bed for patterns across languages given their comparatively clear semantics and modularity. In a survey of the numerals systems of over 2000 languages from all over the world, we have come across the following kinds of complexity:

Complexity as Irregularity The form of a numeral is not predictable from the forms of its mathematical parts and knowledge of the rest of the language. Plenty of languages have irregularities specific to their numeral forms that can be captured by a system of rules. Even less regular are languages such as Hindi [1] where the numbers 1-99 (although easily etymologizable) show no consistent pattern in formation. They have to be learned more or less by rote.

Complexity as Global Ordering Constraints A small number of languages, such as e.g Guajiro [16], Kikongo [13], Breton [11], have discontinuous numerals. As high numerals often behave like nouns we and forms like, e.g. Breton:

    pevar   mil   ha   tri-ugent
    four   thousand   on   three-twenty
    '64 000'

Some languages (e.g. [2, 9]) that have N Num order order while still having bigger-larger order between additive constituents get a potential ambiguity in phrases like 1000 8 which can then mean either 8000 or 1008.[1] Perhaps from pressure to resolve the ambiguity more clearly than intonation, some languages have evolved reordingering possibilities that depend on the diagnosed ambiguity of the whole resulting expression. The potential implications of such global parameters on parsing and production justify looking at them separately of everyday parametres.

Complexity as Circumlocution Perhaps in response to a sudden demand for an expanded numeral system, there are languages which have evolved very morpheme-promiscuous number expressions:

e.g Murinypata [14]:

    ma-enumima-enumimenumimenumima-enuminumi
    'hand-one-hand-one-foot-one-foot-one-hand-one-one'
    '26'

or Maipure [15]:

    papeta   janà   pauria   capi-ti   purenà
    one+CL:HUMAN   take/follow(?) other   hand-NPOSS   relative
    'one takes one relative from the other hand'
    '6'

The forms are highly regular yet even grammar writers tend to label them cumbersome in comparison to the more familiar expressions in European languages.

A language with lots of irregularity, infixing and/or circumlocution specific to its numerals will be said to have a complex numeral system. An example that will be familiar to many is (not-too-progressive) Welsh. The opposite, a numeral system with almost a mininum number of lexemes and completely regular formation of higher

---

[1] N Num order is quite common [4] whereas Malagasy [3] seems to be the only modern language with consistent smaller-higher ordering.

number from them, is non-complex one (e.g. Sango [12] or many Austronesian languages).

Not counted are some numeral systems with ridiculous amounts of (quite idiosyncratic) forms because the difficulty invariably comes from the classifier fusing with the numeral (e.g Loniu [8], Upriver Halkomelem [6], Ket [5], Palikúr [7]) rather than irregularity in forming the numerals themselves.

Now, if the characterization of complexity make sense, it would be highly interesting to see to what extent creoles have the same amount or less complexity than their lexifier language. A tabulation of a sample of 20 creoles is given below. Unfortunately the lexifier languages (Portuguese, French, English, Spanish, Dutch, Arabic) showed only complexity of the first type (and none of the creoles introduced more complexity of any kind), so the investigation results in one of analyticity in formation of teens and tens:

# of Creoles	Irregularity	
	teens	tens
6	-	-
3	-	x
11	x	x

So, whereas some creoles do tend to analyticity, a majority don't. Parkvall [10] shows similar findings for African creoles specifically. We will discuss some possible interpretations of this fact.

References

William Bright. Language Variation in South Asia, chapter 7: Hindi Numerals, pages 74_85. Oxford University Press, 1990.

Terry Crowley. An Erromangan (Sye) Grammar, volume 27 of Oceanic Linguistics Special Publication. University of Hawaii Press, Honolulu, 1998.

Jacques Dez. Cheminements Linguistiques Malgaches: au-delà des grammaires usuelles , volume 20 of Asie et Monde Insulindien. Peeters/SELAF, Paris, 1990.

Matthew S. Dryer. Order of numeral and noun. In David Gil Bernard Comrie, Matthew S. Dryer and Martin Haspelmath, editors, World Atlas of Language Structures. Oxford University Press, 2005.

A. P. Dul'zon. Ketskij Jazyk. Izdatel'stvo Tomskogo Universiteta, Tomsk, 1968.

Brent D. Galloway. A Grammar of Upriver Halkomelem, volume 96 of University of California Publications in Linguistics. University of California Press, Berkeley and Los Angeles, 1993.

Diana Green. O sistema numérico da língua palikúr. Boletim do Museu Paraense Emílio Goeldi, Série Antropologia, 10(2):261_303, 1994. Excellent previously not exposed data.

Patricia J. Hamel. A Grammar and Lexicon of Loniu, Papua New Guinea, volume 103 of Pacific Linguistics: Series C. Research School of Pacific and Asian Studies, The Australian National University, 1994.

Abdulaziz Y. Lodhi. Jifunze kusema kiswahili. Nordiska Afrika Institutet, Uppsala, 1974.

Mikael Parkvall. Out of Africa: African In_uences in Atlantic Creoles. Battlebridge
    Publications, London, 2000. Ian Press. A Grammar of Modern Breton, volume 2
    of Mouton Grammar Library. Mouton de Gruyter, 1986.
William J. Samarin. A Grammar of Sango, volume 38 of Janua Linguarum: Series
    Practica. Mouton de Gruyter, 1967.
Bertil Söderberg and Ragnar Widman. Kikongo. Svenska Bokförlaget, Bonniers,
    Stockholm, 1966.
Michael J. Walsh. The Murinypata Language of North-West Australia. PhD thesis,
    Australian National University, Canberra, 1976.
Raoul Zamponi. Maipure, volume 192 of Languages of the World/Materials. Lincom
    GmbH, München, 2003.
Jesús Olza Zubiri and Miguel Angel Jusayú. Gramática de la Lengua Guajira
    (morphosintaxis). Universidad Católica del Tachira, San Cristóbal, 1986.

25. Compression-based analysis of language complexity

Patrick Juola, Duquesne University, USA

Even in a well-defined domain like mathematics, the definition of "complexity" can be problematic. In linguistics, it is almost impossible to get a single acceptable definition, which has not, of course, stopped people from trying (Berwick and Weinberg, 1984; Nichols, 1992; Perkins, 1992; Chater and Hahn, 1997). Previous research (Juola, 1998, 2000; Juola and Bailey, 1999) has suggested that a useful definition of "complexity" is simply the minimal amount of (cognitive) machinery necessary to reconstruct the object of study. In the case of computable objects, the minimal amount is (obviously?) the minimal computer program to do the computation, also known as the Kolmogorov or Chaitin complexity (Li and Vitanyi, 1997).

Although Kolmogorov complexity cannot be measured exactly, it can be approximated by standard compression tools such as gzip or bzip2. The compressed file constitutes a shorter "program" to reconstruct the original input. A more complex data set would have a large residual file after compression. I have further argued (2004) that systematic distortion, (for example, by replacing lexical types with random markers) followed by compression, can provide a method of isolating linguistic aspects such as the complexity of different domains.

To briefly recap, studies of the Bible in various languages (Juola, 1998) have shown clear evidence of differences in morphological and syntactic complexity between Dutch, English, Finnish, French, Maori, and Russian. I extend that analysis here using a slightly different framework and an enlarged data set.

As before, the analysis is performed on 25 Bible translations representing a variety of languages. Each sample is systematically distorted in one of three ways. With fixed probability (in the experiments reported, 10%) either a) each non-whitespace letter is removed (distorting the systematic subword regularities of morphology), b) each "word" [maximal non-whitespace sequence of characters] is removed (distorting the intrasentential syntactic regularities), or c) each "line" [Bible verse, typically one or more complete sentences expressing a single thought] is removed and replaced by its whitespace elements (distorting the intersentential

pragmatic regularities such as topic introduction). It is easily provable that these three methods are expected to remove the same amount of source text, but the resulting files can be expected to respond differently to compression. The resulting texts are then compressed using one of two standard compression technologies, and the residual size calculated as a percentage of the compressed size of the undistorted file.

The extended results largely replicate those of previous studies on compression-based morphological and syntactic complexity. As before, languages differ reliably on their measured morphological and syntactic complexity, in an intuitive continuum that puts (morphologically impoverished, but syntactically rich) English at one end and Finnish at the other. Also as before, the correlation between high morphological complexity and low syntactic complexity (or vice versa) is extremely good, suggesting a strong domain tradeoff in complexity. On the other hand, variation in pragmatic complexity among languages is very small (almost nonexistent), suggesting that "pragmatic" constraints and complexity may be crosslinguistic universals.

One counterintuitive finding was that the residual percentage of information after 10% deletion was not usually 90%. Deleting 10% of the Bible verses typically resulted in 91.4-91.5% of the original data remaining (reasonably close to the expected 90.0%), but word deletion typically left 95-98% of the original file size. Even more surprising, deletion of 10% of the characters actually increased the residual file size, typically to 110-120% of the original! Although not difficult to explain mathematically, this may have broader implications in terms of the primacy of the lexicon in the "algorithms" of language processing, a point also discussed in (Hutchens and Adler, 1998; Juola and Bailey, 1999; Juola, 2000). By treating lexical items as computational and representational primitives, it may be more "efficient" to describe a nonce sequence such as *"pscholgy" as a deletion from an established (primitive) lexical item than building it from scratch as a composition of primitive letters. This has implications not only for human cognition, but may also have general application in evaluation of text-based AI systems or text compression systems.

References

Berwick, Robert C. and Amy S. Weinberg. (1984). The grammatical basis of linguistic performance : Language acquisition and use. Cambridge, Mass.: MIT Press.

Chater, Nick and Ulrike Hahn. (1997). "Representational distortion, similarity, and the Universal Law of Generalization." In Proceedings of the Interdisciplinary Workshop on Similarity and Categorization (SimCat 97), Edinburgh.

Hutchens, Jason L. and Michael D. Alder. (1998). "Finding Structure via Compression." In Proceedings of New Methods of Language Processing 3 (NeMLaP 3), Sydney.

Juola, Patrick. (1998). Measuring linguistic complexity : The morphological tier. Journal of Quantitative Linguistics, 5(3):206--13.

Juola, Patrick. (2000). "A Linear Model of Complexity (And Its Flaws)." At the Fourth International Conference on Quantitative Linguistics (QUALICO-2000), Prague, Czech Republic.

Juola, Patrick. (2004). "On Composership Attribution." Proceedings of the 2004 Joint Annual Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities (ALLC/ACH 2004), Gothenborg, Sweden.

Juola, Patrick and Todd M. Bailey. (1999). "Quantifying Quasi-Regularity."
       Proceedings of LSA 1999, Los Angeles.
Juola, Patrick, Todd M. Bailey, and Emmanuel M. Pothos. (1998). "Theory-Neutral
       System Regularity Measurements." In Proceedings of the Twentieth Annual
       Conference of the Cognitive Science Society (CogSci-98), Madison, WI.
Lempel, Abraham. and Jakob Ziv. (1976). On the complexity of finite sequences.
       IEEE Transactions on Information Theory, IT-22(1):75--81.
Li, M. and P. Vitanyi. (1997). An Introduction to Kolmogorov Complexity and Its
       Applications. Graduate Texts in Computer Science. Springer, New York.
Nichols, Johanna. (1992). Linguistic diversity in space and time. Chicago, IL:
       University of Chicago Press.
Perkins, Revere D. (1992). Deixis, grammar and culture. Amsterdam: John
       Benjamins.

26. Compexity and simplicity in minimal lexica: on multifunctional lexical items and compositionality in pidgin lexica

Päivi Juvonen, Stockholm University, Sweden

If the size of lexica is taken as a measure of lexical complexity, pidgin lexica are propably the simplest lexical systems in any human language: the number of lexical items listed for such well documented pidgin languages as Chinook Jargon, Mobilian Jargon and Lingua Franca range from around 500 to just over 2000. If, however, we by lexical complexity mean e.g. the number of argument structures connected to a lexeme and their position within that structure or, the number of derivational rules in a lexicon, pidgin lexica do show signs of complexity.

The present paper exemplifies and discusses the use of both multifunctional lexical items (i.e. lexemes with several possible syntactic functions within a clause) and productive derivational processes in pidgin lexica in terms of complexity.

27. From complexity to reduction: The case of Central African Contact Languages

Valeri Khabirov, Urals Pedagogical University, Russia

Contact linguists have always considered complexity to be an important issue for characterizing the structure of the creoles and simplicity – as a characteristic feature of pidgins. McWharter (2000:115) defined creolization as radical reduction by non-natives followed by reconstitution into a natural language. So, what underwent the process of reduction was first complex. Being a creolized language Sango (CrS) underwent a certain structural change which makes it different from the substrate language characterized by structural complexity belonging to the dialect continuum of Ubangian Niger-Congo languages. The loss of grammatical tone brought about changes in the verbal system of sango and some lexical items were grammaticalized

as tense and aspect markers. The processes of creolization affect the lexicon of all languages, but the extent to which they are evidenced in creolized languages suggests that they are accelerated by restructuring. The process of creolization in CrS brought about the reduction of the lexicon (down to 300 words) and as a result the reduction of the derivational morphemes existing in the source language - the ethnic sango (EthS). The latter has in its lexicon verbs with general meaning along with words having specialized meaning (process akin to semantic narrowing) produced by derivational morphemes which were lost during creolization: *-rV* (iterative meaning), *-ngbì* (meaning of insistence), *-ngà* (resultant action), *-sà* (meaning of removal), *-ndò* (meaning of accumulation) and some others. But 4 bound morphemes have been retained: the subject marker *à-*, the plural marker *á-*, derivational morphemes *wà,* marker of agent or possessor and *-ngo,* marker or deverbal nouns which denote the process of the action. This proves the fact that creolized languages as compared with creoles are structurally more close to their substrate. As far as the tonal system is concerned the CrS has retained only lexical tones – grammatical tones (with very few exceptions) were lost and grammaticalized lexical items took their place to express tense-aspect relations. A telling example is the formation of the Future Tense with the help of the predicative element *yèkè* (a borrowing from kikongo) and *fàdē* – probably a grammaticalized content word meaning "quickly": 1) (CrS) <u>fàdē</u> kóli à gā : 2) (EthS) kóli <u>á</u> gā "A man will come"; 3) (CrS) kóli à <u>yèkè</u> gā ándè : 4) (EthS) kóli á gā bíàní : "A man will come" (by all means). In sentence 3 the aspectual significance of the predicative element *yèkè* is Imperfective Obligatory. The verb form refers to a future action with the help of *ándè* meaning "later". In sentence 4 in riverine ethnic sango the obligatory action is expressed by the notion word *bíàní* meaning "for sure" and the future tense – by the high tone of the Subject Marker *á*. The answer to the question of whether sango must be called a pidgin, a creole or a koiné is that it is a creolized language like Congolese languages munukutuba and lingala because pidgins and creoles do not have direct link to their lexifiers unlike CrS. There have also been great changes in the grammatical structure of another creolized language of Central Afrika – Munukutuba. Creolized Munukutuba is characterized by the reduced system of noun classes (down to [from?] 12) as compared to its source language, KiKongo. In the process of creolization, Munukutuba's lexifier underwent the restructuring of its grammatical system that can be seen by comparing, for example, agreement phenomena. An example of full agreement is found in KiKongo: ki-lunzi (cl. 6) ki-a:-ni ki-a ki-bote ki:wi:di//life/my/det./beautiful/finished//; bi-lunzi (cl. 7) bi-a-ni bi-a bi-bote bi-wi:du//lives/our/det./beautiful/finished//. In Munukutuba there are no possessive prefixes and the adjective does not agree with the noun. There are no concords on verbs, either, and the meanings of kilunzi kia:ni and bilunzi biani are give by the determinative syntagma. The source language, KiKongo, has the same kind of syntagma, but the original paradigm preserves the necessary concords, for example: mwa:na mwa mbote/child/det./goodness ("good child") (cl. 1) and ba:na ba bibote – plural (cl. 2). Of 48 demonstrative pronouns in KiKongo indicating distance form the speaker, only three remain in Munukutuba. There were large changes in other aspects of Munukuta as well.

28. Second Language Text Processing and Induced Schema

Jung-Tae Kim, University of Incheon, South Korea

Reading involves the ability to read not only from the language of the text, but also from the reader's own prior knowledge and perspective. One related question is what takes place when the schema that the reader possesses is informationally appropriate for the target text that the reader will read, but structurally different from the text. The structural difference may result when the perspective or macrostructure of the context from which the knowledge was first acquired is different from that of the target text. The current study questioned what happens in L2 reading comprehension of the expository text, especially in recall and inference-making abilities, when a L2 reader was induced to develop a content schema about the topic of a target text, but the structure of that schema departs from the structure of the target text. Two English expository texts were used in the current study. Both texts share the same information about the violin but were organized according to different themes. One text described some aspects of the violin from a historical perspective (historical version). The other text provided the same information about those aspects of the violin, but from a structural and functional perspective (functional version). Seventy-four Korean university students read either the same version text twice (consistent condition group) or two different version texts (inconsistent condition group) with a three-day interval between the two readings. The results of a comprehension test indicate that, for the subjects with high L2 English reading proficiency, those who were in the inconsistent condition group outperformed those in the consistent condition group in the inference-making task, while no such difference was found in the recall task between the two groups. For the subjects with low-level L2 English proficiency, on the other hand, there was no significant difference between the two groups in the results of the inference-making task while the consistent condition group performed better in the recall-task. These results were interpreted as showing that inducing a structurally inconsistent schema through an L2 pre-reading would be beneficial only when the reader's L2 linguistic ability is proficient enough to produce sufficient propositions from the pre-reading. Research on reading process suggests that the inference-making process, compared to the recall process, depends more on rich interconnections between information in the reader's knowledge base because the construction of interconnection means the construction of new relation that has not been presented explicitly in the texts. These richer interconnections would be produced in the inconsistent reading condition, rather than in the consistent reading condition, as more connections are needed to reconcile the target text with the organization of the pre-existing content schema. In the present study, the high level L2 readers might be able to construct content schema about a topic through pre-reading a text, and could use that schema to develop deeper level comprehension of a target text. The low level L2 readers did not have such ability. Instead, the study showed that pre-reading the same text is more beneficial to the lower-level L2 readers' recall ability, suggesting the positive re-reading effect for the text recall.

29. Interlinguistic difficulty: one cup complexity and two cups linguistic baggage?

Eva Lindström, Stockholm University, Sweden

Kuot is spoken by some 1500 people on New Ireland, Papua New Guinea. It is the only language in the province which does not belong to the Austronesian family. Both speakers of Kuot and speakers of neighbouring languages maintain that Kuot is extremely difficult and the languages around it are really easy. But why would it be easier for a speaker of Kuot to learn, say, Nochi than vice versa?

In this talk I will show how it can be argued that it would indeed be harder for an adult speaker of Nochi to acquire Kuot than the other way around. The reasoning is that a measure of difficulty could be the number of categories that are obligatorily or typically marked, combined with the number of morphs used to express those categories.

I will focus on the morphological paradigms of the verb phrase in Kuot and the three neighbouring Austronesian languages Nalik, Nochi and Madak, simply moving left-to-right through the phrase and counting the number of choices encountered along the way.

I am restricting my claims to the concept of "difficulty". While systemic complexity is certainly a factor in the sort of difficulty considered here, it should be noted that difficulty is anchored in the person for whom something is difficult. That is, the match or mismatch with particular categories and sub-systems in the learner's own language will factor into the level of difficulty experienced.

30. Approaching grammatical complexity from a cross-linguistic point of view

Matti Miestamo, University of Helsinki, Finland

Recently grammatical complexity has received more and more attention among language typologists. In this context questions of the comparability of languages in terms of complexity have naturally become important. Linguists often take it for granted that all languages are equally complex - in different languages complexity may be found in different areas of grammar, but complexity in one area is usually thought to be compensated by simplicity in another. This so-called equi-complexity hypothesis is extremely difficult to test empirically. One attempt can be seen in the metric proposed by McWhorter (2001), which contains criteria for measuring complexity in different areas of grammar. But like any attempt to measure overall grammatical complexity, the metric faces the following fundamental problems: the problem of representativity (it is not possible to exhaustively take into account all aspects of grammar) and the problem of comparability (the contributions of different areas of grammar to overall complexity are incommensurable) (Miestamo, in press). These problems are encountered in all levels of language and they make the comparison of overall complexity extremely difficult if not impossible. When approaching complexity from a cross-linguistic point of view, we must therefore concentrate on the complexity of specific grammatical domains. We can then make cross-linguistic generalizations about the complexity of these specific areas and gain some understanding about possible compensatory tendencies between the complexities of the areas studied.

A usable basis of comparison can be found in how languages grammaticalize different functional domains. It can be argued that of two languages that grammaticalize a given domain, one that makes a larger number of distinctions within the domain is more complex in this respect than the other. Thus, the functional

domain of tense, for example, is more complex in a language that makes six tense distinctions than in one that makes only three. The number of grammaticalized distinctions figures as one of the criteria of complexity in McWhorter's metric as well. In addition to the investigation of the distinctions made within functional domains on the paradigmatic axis, we can also pay attention to how these distinctions are realized on the syntagmatic axis: how densely different functional domains find grammatical expression in running text and how these findings are related to the number and nature of the distinctions available on the paradigmatic axis. I will investigate these questions empirically on the basis of a representative language sample, trying to find statistical correlations between the complexity and simplicity of different domains. In this paper I intend to pay special attention to the expression of the grammaticalized distinctions on the syntagmatic axis, examining short texts available in the sample languages.

References

McWhorter, John H. 2001. The World's Simplest Grammars are Creole Grammars. Linguistic Typology 5: 125-66.

Miestamo, Matti. In press. On the Feasibility of Complexity Metrics. In: Krista Kerge and Maria-Maren Sepper, ed., FinEst Linguistics, Proceedings of the Annual Finnish and Estonian Conference of Linguistics. Tallinn, May 6-7, 2004. Tallinn: TPÜ/EKI.

31. The simplicity of creoles in a typological perspective

Mikael Parkvall, Stockholm University, Sweden

McWhorter's suggestion that "world's simplest grammars are creole grammars" has generated considerable debate, both within and outside creolistics. While typologists seem to regard this claim as interesting, though controversial, there is a near concensus among creolists that it is not, and could not be correct. Contemporary creolists generally assume that the label "creole" is one which is based exclusively on socio-historical grounds, and which has no structural/typological correlate whatsoever.

Meanwhile, little work has been done to empirically asses the issue. The recently compiled *World Atlas of Linguistic Structures* (WALS) provides a unique possibility of doing so. Including nearly 60 000 data points, it is unparalleled by any other typological database.

Examining creoles in the light if this wealth of data, it turns out creoles do indeed form a typological group, and furthermore one that is characterised by simplicity. Interestingly, given the idea that creoles only share socio-historical properties, it is even impossible to *invent* a grouping that matches the creole profile.

32. "Subjects and complexity in the recent history of English"

Javier Perez-Guerra, University of Vigo, Spain

CANCELLED

33. Complexity of Verbal Morphology in Finnish Romani

Helena Pirttisaari, University of Helsinki, Finland

CANCELLED

34. Language Complexity and Multilingual Learning

Françoise Ratzlaff Ph. D, Osaka University, Osaka, Japan, Gordon Ratzlaff Ph.D, Ritsumeikan University, Kyoto, Japan

In simultaneous multilingual learning, language complexity does not necessarily increase with the number of languages learned. The more languages learned the less complex the languages seem to be. A global look on many languages brings a feeling of unity rather than complexity. If the combinations of difficulties present in these languages make learning difficult, they also can encourage the learner to see more similarities than differences among these languages.

Speakers of Turkish, Corean, Mongolian, Hindi or Tamil living in Japan are known to learn Japanese more easily than speakers of, for example, Swedish, Russian, English or French, who have to adopt a new pattern with SOV, postpositions and adjectival sentences, which slows down their learning process. But how can so many speakers of languages very different from Japanese still manage to master Japanese?

Examples of the complexity of interrelated rules are definite conjugations in Hungarian where the verb ending matches the definition of the direct object, verb pairs in Russian with accomplished/ non accomplished aspects, the system of the presentative in Wolof, the passé composé in French and Italian which requires either the auxiliary to have or to be, making the use of object personal pronouns difficult. To a motivated multilingual learner who is able to focus further ahead and patiently advance, such complexity is felt as a source of difficulties necessary to progress.

The difficulty to embrace a complex system may be greater at the onset of the learning process. From the start, a multilingual learner opts for active learning, trying to remain highly conscious of L2 differences with L1 and by practicing L2 word order in L1. These learning techniques are all part of what we call the "unlearning process" of L1.

We would like to put forward the following hypothesis: complexity experienced at the beginning of the learning process is part of a network which reaches the structure of other languages, later facilitating their study.

35. Complexity in Isolating Languages: Lexical Elaboration vs. Grammatical Economy

Elizabeth M. Riddle, Ball State University, USA

Many East and Southeast Asian languages such as Mandarin, Thai and Hmong are highly analytic or isolating, with little affixation and minimal overt marking of syntactic relations other than by word order and independent morphemes. As compared to more synthetic languages such as Finnish, Polish, or even English, their surface structures are relatively paratactic in multiple ways. I argue that this tendency toward economy of surface grammatical marking, which can be considered a type of simplicity, is balanced by a strong tendency toward lexical elaboration, a form of complexity. Thus Mandarin, Thai, and Hmong, among others, have numerous ways of elaborating structures lexically.

I claim that the characteristic verb serialization and other forms of verbal concatenation of such languages (which have a corresponding relative paucity of subordination markers and other signals of grammatical hierarchy) serve an elaboration function by providing a richness of detail about subparts of a situation with a string of verbs. In contrast, languages such as English or Polish often leave those subparts unsaid, (but elaborate the syntax with overt hierarchical marking of subordination). Compare the three verbs in Hmong "sawv nres tos bus" [stand stop wait bus] with the single verb "wait" in English "wait for the bus".

Another case is the common use of four-part "elaborate expressions" (term from Haas 1964) in these languages, ranging from everyday vocabulary to elegant style. Typically, these comprise four compounded morphemes, where the first and third are identical or similar on a semantic and/or phonological level, and the second and fourth likewise share semantic and/or phonological properties. Two examples from Hmong are the very common "cua daj cua dub" [wind yellow wind black] 'storm' and the elegant "nroo tshaib nroo nqhis" [sigh hunger sigh thirst] 'starve.' Also common are Mandarin (numbers indicate tones) "jin33 yin213 cai45 bao213" [gold silver wealth treasure] 'a lot of money, and Thai "khit45 pay33 khit45 maa33" [think go think come] 'think back and forth until reaching a decision.'

Elaborated two-syllable compounds of semantically similar morphemes are also frequent, as in Thai "baan41 ryan213" [home house] 'house' where both "baan41" and "ryan213" can be used independently, likewise Mandarin "jin33 qian213" [gold money] 'money.'

Another type involves sound symbolism. Hmong has dozens of lexicalized and highly specialized partial reduplications to indicate particular sounds, such as that of tigers crunching bones: "nkig nkog," which elaborate on particular aspects of a situation.

I discuss the formal and functional interrelations of these and other structures, and conclude that there is a correlation between complexity in terms of lexical elaboration and surface simplicity in terms of economy of grammatical marking in the analytic (vs. synthetic) languages examined here.

Reference

Haas, Mary. 1964. Thai-English Student's Dictionary. Stanford: Stanford University Press.

36. An Information-Theoretic Approach to Complexity in Pidgin and Creoles

Stuart Robinson, Fermin Moscoso del Prado Martin, Max Planck Instituut, The Netherlands

CANCELLED

37. Complexity and productivity in word formation

Carmen Scherer, University of Mainz, Germany

The German language is famous for its possibility to coin long and highly complex words such as *Fußballweltmeisterschaftsqualifikationsspiel* (lit. 'football world cup qualification match') or the famous *Donaudampschifffahrtsgesellschaftkapitän* (lit. 'Danube steamboat company captain'). However, the option to create words of such a complexity is rather new. Compounds containing three – or even more – roots, such as *Schnell₁koch₂topf₃* 'pressure cooker' (lit. 'fast cook pot') or *Tief₁druck₂gebiet₃* 'low-pressure area', are very common in modern German, whereas in Old and Middle High German compounds usually contained no more than two roots (Splett 2000, Zutt 2000). The rise of morphologically complex words starts in Early High German. In this period, new types of compounds arise due to the loss of internal structure in syntactic units, e.g. so-called genitive compounds (Demske 1999), such as *Lehrersgattin < des Lehrers Gattin* (lit. 'teacher-wife' < 'the teacher's wife'), and derivational processes become more complex, e.g. simultanuos pre- and suffixation as well as double-prefixation of verbs is first attested (Wegera/Prell 2000, Heinle 2000).

In my talk I will address the question of changes in morphological complexity by means of morphological productivity. I will present data from the Mainz Newspaper Corpus (1609-2000, 9 measuring points, 1 million word forms; Meibauer et al. 2004, Scherer 2005) and trace the development of complexity for one particular word formation pattern, the German *-er*-nominalization, over the last 400 years. Three different measures of productivity will be evaluated: the number of types in the corpus ('extent of use', cf. table 1), the proportion of hapax legomena ('productivity in the narrow sense', cf. tables 2, 4) and the combination of both ('global productivity', cf. tables 3, 5). I will aim at showing in how far these three notions of productivity are compatible and can help to understand the increase of *-er* nominals with complex bases as compared with *-er* nominals formed out of simplex bases. Depending on the morphological complexity of the base, my analyses will show significant variation of morphological productivity within the period investigated. I will finally illustrate that the increase of morphologically more complex *-er* nominals does not restrain the productivity of the less complex ones (cf. tables 2-4).

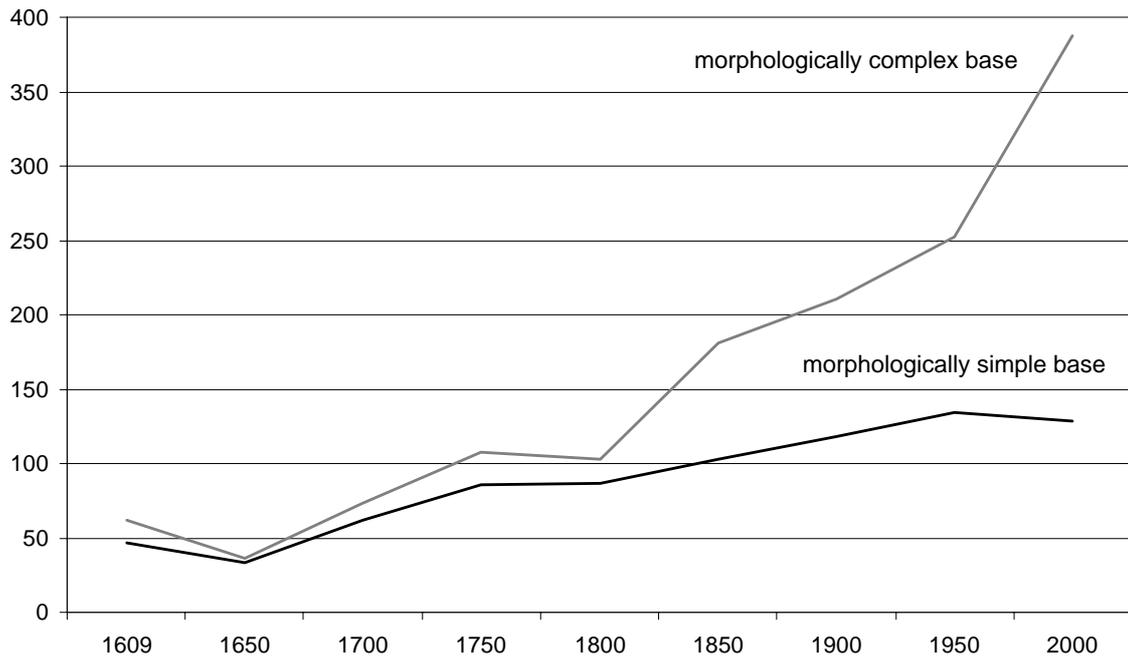Table 1:   types per 100,000 word forms

Table 2:   morphologically simple base: productivity in the narrow sense

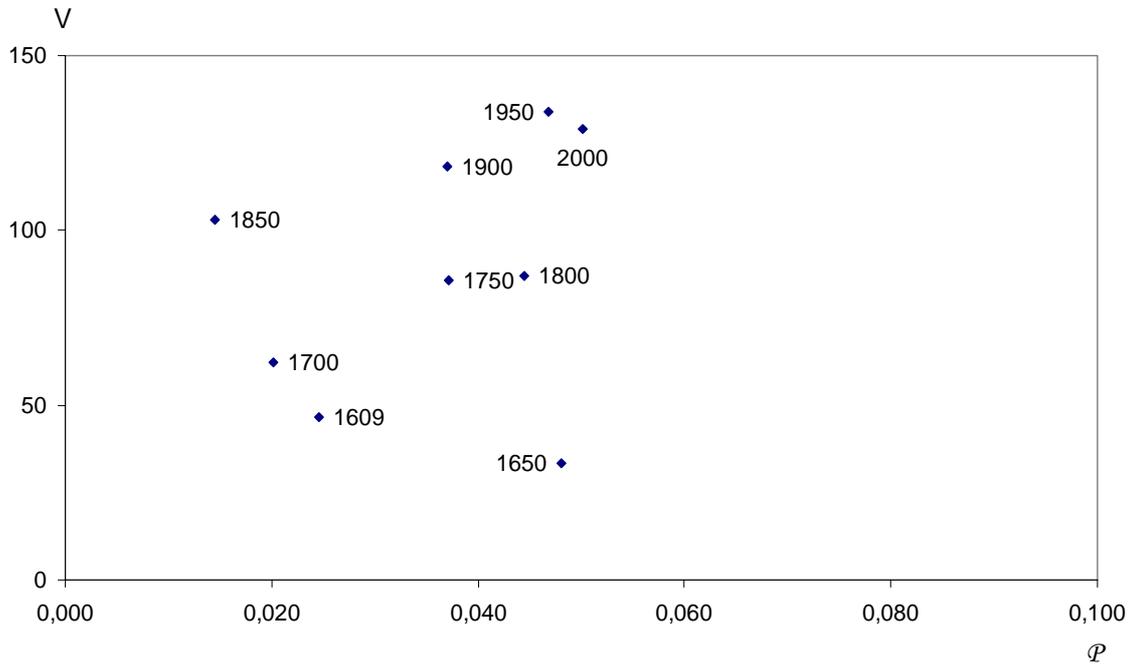Table 3: morphologically simple base: global productivity

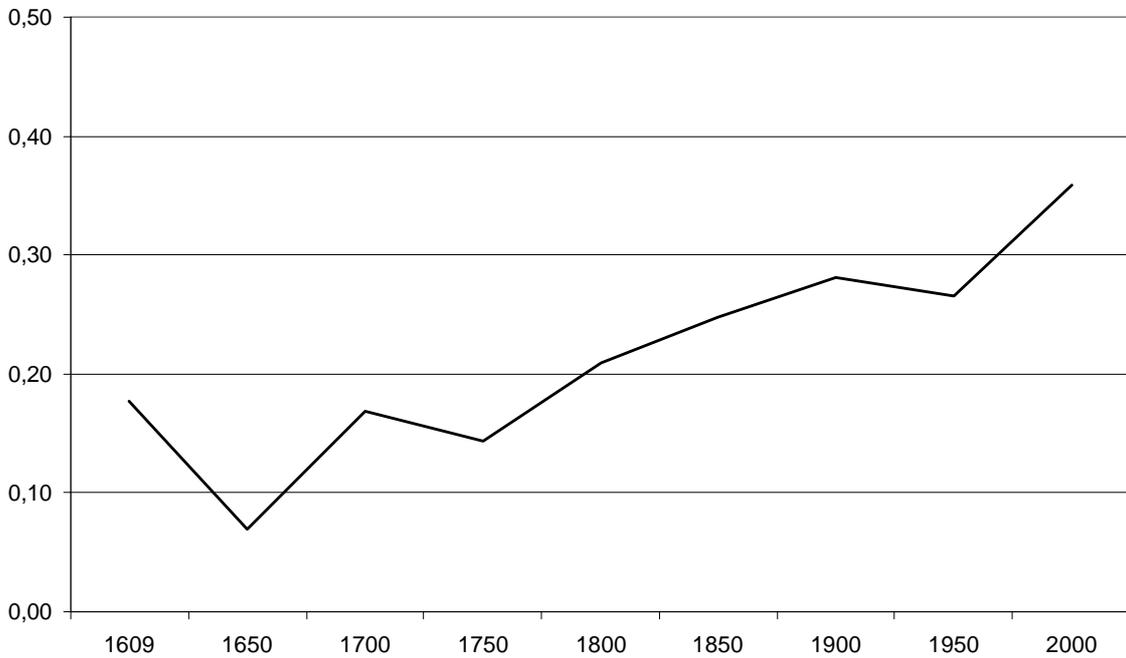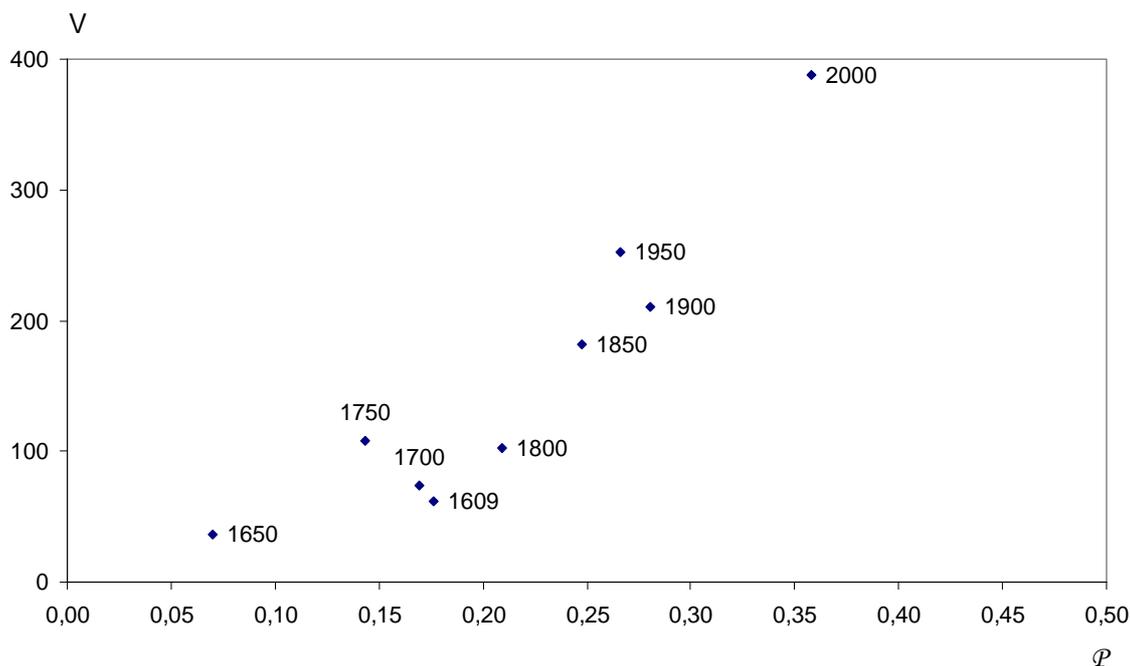Table 4: morphologically complex base: productivity in the narrow sense

Table 5:   morphologically complex base: global productivity

References:

Besch, Werner et al. (eds.) (2000): Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung. Vol. 2. 2nd ed. Berlin: de Gruyter.

Demske, Ulrike (1999): Case Compounds in the History of German. In: Butt, Matthias/Fuhrhop, Nanna (Hgg.): Variation und Stabilität in der Wortstruktur. Untersuchungen zu Entwicklung, Erwerb und Varietäten des Deutschen und anderer Sprachen. Hildesheim: Olms, 150-176.

Heinle, Eva-Maria (2000): Wortbildung des Neuhochdeutschen bis zur Mitte des 20. Jahrhunderts. In: Besch et al. (eds.) (2000), 1911-1917.

Meibauer, Jörg/Guttropf, Anja/Scherer, Carmen (2004): Dynamic aspects of German -er-nominals: a probe into the interrelation of language change and language acquisition. In: Lingustics 42, 155-193.

Scherer, Carmen (2005): Wortbildungswandel und Produktivität. Eine empirische Studie zur nominalen -er-Derivation im Deutschen. Tübingen: Niemeyer. (in press).

Splett, Jochen (2000): Wortbildung des Althochdeutschen. In: Besch et al. (eds.) (2000), 1213-1222.

Wegera, Klaus-Peter/Prell, Heinz-Peter (2000): Wortbildung des Frühneuhochdeutschen. In: Besch et al. (eds.) (2000), 1594-1605.

Zutt, Herta (2000): Wortbildung des Mittelhochdeutschen. In: Besch et al. (eds.) (2000), 1358-1365.

38. Complex combinations of case marking, agreement and word order

Kaius Sinnemäki, University of Helsinki, Finland

The two most thoroughly investigated areas in linguistic typology are the order of constituents and the morphological marking of the primary verbal roles (whether subject and object, agent and patient, or topic and focus). It has often been claimed that languages with no morphological marking of the primary roles would disallow flexible word order and prodrop of pronouns. Conversely, languages with case marking and/or agreement would warrant flexible word order and prodrop of pronouns. These claims reflect the assumption that languages "balance out" complexity in one area with simplicity in another, e.g. when languages with little or no (inflectional) morphology are said to compensate this by elaborate syntax (e.g. Hockett 1958: 180,181).

Recently, e.g. Siewierska (1998) has tested these hypotheses with a sample of 171 languages. According to her results, the absence of either agreement or case marking is a good indicator of rigid word order, whereas flexible word order presumes the presence of either of the two. Consequently, formal subsystems of languages seem to adhere to certain level of compensation. However, Siewierska did not test how the presence or absence of BOTH case marking and agreement would correlate with word order flexibility. This is rather unfortunate, since case marking and agreement tread the same functional domain (in the sense of Givón 1981), thus being prone to compensating one another. For instance, Siewierska's sample consisted of seven languages which had no agreement but exhibited three or more word order variants. It is possible that the languages which lacked agreement but exhibited flexible word order employed case marking for distinguishing the primary verbal roles. In fact, this is exactly what happens in Yidiɲ (a PamaNyungan language from Australia) and some other languages in her sample. This suggests that both case marking and agreement can by themselves warrant flexible word order.

In this presentation, I will present a methodology for discussing certain complexities of morphological and syntactic marking. I argue that both agreement and case marking should be considered simultaneously when compared to word order flexibility. I also discuss some ways of measuring syntactic complexities in the marking of the primary verbal roles. The method will pay attention to the presence vs. absence of agreement and case marking (and their qualities), to the flexibility and complexity of word order, and to the possibility of prodrop. The method will be applied to several languages, some of which come from Siewierska's sample.

References

Givón, Talmy 1981. Typology and Functional Domains. *Studies in Language* 5 (2): 163193.
Hockett, Charles F. 1958. *A Course in Modern Linguistics.* New York: Macmillan.
Siewierska, Anna 1998. Variation in major constituent order: a global and a European perspective. In A. Siewierska (ed.), *Constituent Order in the Languages of Europe*, pp. 475551. Berlin / New York: Mouton de Gruyter.

39. Nominal predicate constructions in Erzya

Rigina Turunen, University of Helsinki, Finland

This paper focuses on the morphosyntax of Erzya-Mordvin nominal, adjectival and locational predications. One special feature of both Mordvin languages is that it is possible to inflect predicate nouns in person, number and tense using the same suffixes as for verbal inflection. This facilitates highly synthetic as well as morphologically and semantically complex productive predicative constructions. Interestingly, analytic predicative constructions are used side by side with the synthetic constructions.

The corpus of this survey consists of Erzya texts of variable ages and genre, the oldest dating back to the beginning of the last century. The method of the study is more qualitative than quantitative, although frequency observations have been included in order to provide a more comprehensive view of the usage of different constructions.

In the synthetic construction (Type 1), the nominal predicate agrees in number and person with the subject. The nominal inflection paradigm is shorter than the verbal: it is found only in the indicative and present and perfect forms of the tenses. The predicate and possessive suffixes are not identical. Synthetic constructions seem to have arisen in two ways. Firstly, in the perfect the copula 'be' has fused with the non-verbal predicate. Secondly, in the present tense the agreement of predicate with subject gives an impression of inflection.

1. *Min' son-ze   marto   paro   jalga-tano.*
   We   (s)he-SG3 with   good   friend-PL1
   'We are good friends with her/him.'

2. *Mon anok-o-l-i-n'   risht'-e-ms   son-ze.*
   I   ready-PERF-SG1   hit-INF   (s)he-SG3
   'I was going to hit her/him.'

Analytic constructions are formed both with and without a copula. It is worth noting that also in analytic constructions the nominal predicate agrees in number with the subject, except in the locational predicate construction of the 1st and 2nd persons plural. The zero-copula construction (Type 2) is usually used in the third person of the present tense but it can be used with other persons, too.

3. *Min' Misha   marto   vadr'a   oja-t.*
   we   Misha   with   good   friend-PL
   'We are good friends with Misha.'

Also, an analytic construction with the copula ul'ems, ul'n'ems 'be' (Type 3) can be used in the present/future.

4. Sin-st   vechkema sazoro-st   mon ul'-a-n.
   they-PL3   beloved   sister-PL3 I   be-PRES-SG1
   'I am their beloved sister.'

In past tense analytic constructions, the copula must always be used.

5. Min-s'   ul'-n'-i-n'ek   azor-t.
    we-REFL   be-FREQ-IMP-PL1   landlord-PL
    'We were the landlords.'

Even though the results are not yet complete, according to my data it seems that there has been a change towards less complex constructions. Type 2 is almost absent from the older data in the 1. and 2. person, but is found in the newer texts taking space from the synthetic Type 1. Furthermore, in the present Type 3 seems to have been more frequent earlier, suggesting that copula constructions have become rarer during the past century.

40. The Complexity of Valence Morphemes in Afrikaans

Gerhard B van Huyssteen, North-West University, South Africa

CANCELLED

41. Sentence complexity and contextual knowledge. Is structural ambiguity immediately solved in context?

Nina Versteeg, University of Utrecht. The Netherlands

CANCELLED

42. On Measuring Language Complexity as Relative to the Conveyed Linguistic Information

Relja Vulanovic, Kent State University, USA

This paper deals with language complexity in a formal, mathematical way. Language complexity is measured relatively to the information conveyed, in the sense of the following example. Consider three languages which have almost identical grammatical structures, the only difference being that languages A and B have one more verbal form than language C. Speaking absolutely, both A and B are more complex than C. Suppose now that the extra verbal form is used in A to convey a tense which does not exist in C, whereas in B it merely duplicates the usage of another verbal form. In other words, the need for another verbal form can be justified in A by the need to convey one more tense, but this justification cannot be extended to B. Thus, relative to the information that the languages have to convey, A is less complex than B. However, if this information is a factor of language complexity, then speaking relatively, A is not necessarily more complex than C. When finding relative complexity, the main question is how complex the language is in comparison to the simplest possible structure that conveys the same amount of linguistically relevant information. This question has been considered in (R. Vulanovic, Grammar efficiency

and complexity, Grammars 6 (2003), 127-144), where a formula is proposed for measuring grammar complexity. A somewhat more general formula is derived in the present paper. The starting point of the derivation is the paper (J. H. McWhorter, The world's simplest grammars are creole grammars, Linguistic Typology 5 (2001), 125-166), which lists four criteria for comparing language complexity. Only three of those are considered here. They concern syntax, grammaticalization, and morphology and are represented in the formula k/r, where k is the number of different grammatical categories and r is the number of the permissible word orders. Thus, language is more complex if it makes use of more grammatical categories and if its word order is more rigid (i.e. if r is less). The formula k/r is just the first iteration in the process of finding a reliable measure of language complexity. A sequence of examples shows the need for a more sophisticated formula. The examples are simple sentences from various languages, but the approach can be applied in principle to more complicated structures as well. The formula is modified step by step in order to represent syntactic complexity better, take ambiguity into account, and ultimately include the amount and relevance of the syntactic, semantic, and pragmatic functions that are to be conveyed. The final formula therefore measures language complexity with respect to the conveyed linguistic information. It is argued that this relative measure is more suitable for measuring language complexity than an absolute one.

43. "A Measure for Syntactic Complexity of Non-Projective Dependency Trees"

Anssi Yli-Jyrä, University of Helsinki, Finland

Since Victor Yngve's depth hypothesis in 1960, resource-sensitive models have provided us an approach to complexity in syntax. In this talk, such a model is discussed. While many earlier models are based on phrase-structure trees, the discussed model captures dependency graphs, being applicable to a considerable variety of different languages.

The discussed model simulates a left-to-right process that is capable to produces any dependency graphs for sentences. The process allocates two resources: (i) pushdown stacks, and (ii) stack cells. The more crossing dependencies it allocates, the more stacks we need, and the more embeddings there are, the more cells we need in each stack. We are particularly interested in the number of pushdown stacks as they can indirectly measure the complexity of movements, scrambling and discontinuity.

The effects of syntactic complexity to language understanding has been studied earlier through limits of short-term memory, narrowness and similarity-based inferences. These phenomena are motivations for a performance degradation when the number of parallel dependencies (and pushdown stacks) is very high. Furthermore, there is a cross-linguistic tendency towards projective dependency trees that would be captured using one pushdown stack only.

A limit for a number of stacks would alone be a decent hypothesis for acceptable complexity in syntax, but we have made the hypothesis even stronger: our hypothesis is that even if the stack operations are regulated with three carefully selected principles, the number of pushdowns needed to capture crossing dependencies is still limited and, in particular, that it is not sensitive to the length of the sentence. The principles regulate the pushdown operations in such a way that

there a one-to-one correspondence between dependency graphs and left-to-right processes that manipulate the pushdowns.

The model that involves the carefully selected principles has several favourable properties. First, we can mechanically and efficiently find, for any dependency graph, a process that produces it (the process is unique). Second, any specified limit for the number of pushdowns corresponds to a class of mildly context-sensitive (MCS) grammars. Third, since the steps of the process generating a graph correspond to word tokens, a MCS grammar can be learned from a dependency treebank.

Although the current study is limited to a relatively rigid word order language -- Danish --, we encountered some interesting combinations of discontinuous phrases and movements. Based on the Danish Dependency Treebank, our experiments suggest a fuzzy limit of 4-5 pushdowns. Our hypothesis should be tested against a language with a very free word order. Sporadic examples from Swiss German, Latin, and Czech have already been processed without any troubles.

Workshop

44. Consequences of informational complexity for human language processing

ORGANIZER: Dr. Fermin Moscoso del Prado Martin, Medical Research Council - Cognition and Brain Sciences Unit Cambridge, United Kingdom

CANCELLED