

Ana Monika Habjan

*Chunking, Emergence, and Online Production
as Theoretical Concepts in ELF*

ABSTRACT:

Located at the intersection of applied linguistics and more formal language theory, this paper draws a parallel between concepts applied to grasp ELF and increasingly influential *usage-based approaches* to grammar. More precisely, I compare and discuss notions such as *chunking*, *intuition*, *emergent grammar* and *ad hoc constructions*. The discussion is based on chosen texts from the respective fields of study. Basically, that is Sinclair and Mauranen's book on Linear Unit Grammar and, the work of Joan Bybee.

Introduction

The creative and heterogeneous language use, a traditional object of sociolinguistics, is increasingly becoming the focus of scientific research in more formal linguistics as well. In particular, the so called *usage-based approach* seems to be suitable for the description of language acquisition and development in its various kinds of discourse. From this perspective, language is not innate, and grammar is 'distilled' out of language experience (cf. Kaltenböck, 2011). Hence, linguistic structures are not seen as fix and stable, but dependent on concrete interactions and thus dynamic and in constant change. Thus, moving away from the idealization of grammatical sentences of *native speakers*, the usage-based approach should be able to account for *non-native discourse* as well.

Traditionally, theoretical linguistics has been based mostly on inquiries of idealized *native speaker discourse*, more specifically, of *grammatical sentences* of idealized *native speakers*. The abstraction and idealization is, of course, necessary if one wants to come up with (more or less) strict regularities and a well-defined system with clear characteristics. But this limited definition is often questionable because language in its actual performance is in constant change and variation. My main goal is to

stress that the idealization of the *native speaker* is closely related to the assumption of innateness of language and to the project of identifying and describing the universal rules of language, i.e. *Universal Grammar*. Hence, my approach to *non-native discourse* is closely connected with the notion of rules, going hand in hand with the notion of grammaticality and therefore with the notion of *native speaker*, who is traditionally assumed as being the arbiter – by means of his intuition – of grammaticality and appropriateness. The aim of this paper is, therefore, to look beyond this established boundary in linguistics, determined by *grammaticality*, by investigating *non-native discourse*, which is commonly assumed as being ‘ungrammatical’ by definition.

1. The ‘non-native speaker’ in sociolinguistics and formal linguistics

The limits of the traditional ‘science of language’ have during the last decades been indicated by extensive studies in the fields of sociolinguistics and applied linguistics. The concept of *native speaker* itself has been called into question by numerous scholars and although it plays a central role in general theoretical linguistics, it is still «fugitive and subtle» (Davies, 2003: 47-49). Research on English as a Lingua Franca, which not only reflects on the question of *native speakerness* and ownership of language(s), but also provides invaluable *non-native speaker* databases, contributes in an important degree to this research field. For my investigation, therefore, ELF is interesting especially because it touches very fundamental questions of *language* that most linguists in more conventional linguistic theories take for granted.

This paper raises above all the question, how insights into the characteristics of *non-native discourse* achieved by sociolinguistics and applied linguistics, i.e. in this case by ELF studies, can be connected with theoretical concepts developed by general, more formal theoretical linguistics. As dynamic language use and irregularities play a central role in this type of discourse, I want to examine some of the alternatives to the stable and regular grammatical view, which have been developed by chosen theories.

I will therefore basically draw from the increasingly influential *usage-based model*. The striking parallels between their basic tenets and key notions of the descriptions of ELF enable a comparison, which, on the one hand, sheds light on the very ontology of ELF and, on the other, on the linguistic model as such. As Anna Mauranen claims, ELF is supposed to be «a good testbed for models of language aspiring to generality» (Mauranen, 2009: 231).

The discussion will be based on chosen texts from the respective fields

of study. Basically, that is Sinclair and Mauranen's (2006) book on Linear Unit Grammar as well as Anna Mauranen's article on chunking in ELF (2009). From the more general theoretical field, I will focus in particular on the work of Joan Bybee (2010, 2013), one of the most influential proponents of the usage-based approach. In order to draw concrete parallels between the fields of research, I will present and comment on some key notions used in both kinds of texts (i.e. chunking, emergence, on-line production of forms, *ad hoc* construction).

2. Usage-based model and Linear Unit Grammar: main tenets

The term *usage-based model* has first been used by Langacker (1987) and is since then related to cognitive linguistics. In recent years it has, however, evolved into a separate coherent linguistic theory, with an explanatory power in line with those of other important theories of the last decades. Since the model proposes that structure and usage are not separated from each other, it convincingly deals with some phenomena that 'mainstream' linguistics of the last decades has avoided. In this approach, «grammar is seen as an emergent system consisting of fluid categories and dynamic constraints that are in principle always changing under the influence of general cognitive and communicative pressures of language use» (Diessel, 2011: 830).

Consequently, context plays a crucial role in the creation of linguistic forms and the semantic interpretation of these forms, which are seen as variable and dynamic. Special emphasis is put on the functions and the interactive accounts of communication as well as the interaction with the physical world. Grammar is 'distilled' out of language experience (cf. Kaltenböck, 2011: 96), or 'sedimented' out of usage¹. The domain-general processes, responsible for the sedimentation of linguistic forms, are: categorization, chunking, rich memory, analogy and cross-modal association (cf. Bybee and Beckner, 2010; Bybee, 2010, 2013; Diessel, 2011). In short, it is a dynamic model, in which the usage influences the linguistic forms and at the same time the (already sedimented) forms shape usage. So there is no real distinction between competence and performance².

In order to compare some of the key concepts, I have chosen one of the theories, which have already been applied to ELF. The *Linear Unit Grammar* (Sinclair and Mauranen, 2006) seems particularly suitable for drawing parallels, since it also makes use of notions such as chunking, linearity, on-line processing and emergence of linguistic structures.

In Linear Unit Grammar (LUG) they follow the work by Brazil (1995) and Hunston and Francis (2000). The approach tries to overcome the limitations of hierarchical structures in the description of utterances, in particular spoken utterances. Instead of the tree diagrams, utterances are viewed as a linear sequence of chunks, therefore *chunking* as «an intuitive perceptual response to the incoming speech stream» (Mauranen, 2009: 220) is the central concept of the theory. Special importance is ascribed to the world of shared experience, which gets co-constructed by the participants. Interestingly, this shared world is essentially a ‘virtual world’, but interaction takes place in the ‘real world’ (cf. Mauranen, 2009: 223). Even though Sinclair and Mauranen call LUG «a coherent theoretical stance» (Sinclair and Mauranen, 2006: 23), it is presented more as a descriptive apparatus and method than a real detailed model (of analysis). For this paper, it is of prime importance that LUG takes into account all kinds of discourse, not only grammatical sentences of *native speakers*, and that it has already been applied to ELF (cf. Mason, 2007; Mauranen, 2009; Sinclair and Mauranen, 2006)³.

Valuable insights into both research directions can be gained if we select and compare the key similarities and differences between these theoretical approaches to ELF and the *usage-based model*.

3. *Chunking*

As stated above, chunking is one of the central concepts in both grammatical approaches. I will first list the defining characteristics of chunking in the usage-based accounts and then compare these theoretical statements with the so-called ‘pretheoretical term’ from Sinclair and Mauranen’s Linear Unit Grammar.

In usage-based accounts, chunking

«is the process by which sequences of units that are used together cohere to form more complex units. [...] In language, chunking is basic to the formation of sequential units expressed as constructions, constituents and formulaic expressions. Repeated sequences of words (or morphemes) are packaged together in cognition so that the sequence can be accessed as a single unit» (Bybee, 2010: 7).

«[T]he formation of chunks is a continuous process, [therefore] the emerging phrases exhibit varying degrees of cohesion. Other things being equal, smaller chunks (e.g. *the dog*) tend to be more tightly

organized than larger ones (e.g. *the old dog over there that is barking*) because they are more frequent, suggesting that constituency is a gradient concept just like any other grammatical category» (Diessel, 2011: 836f).

From the perspective of LUG, «[c]hunking is a natural and unavoidable way of perceiving language text as it is encountered» (Sinclair and Mauranen, 2006: 6).

In the text on Chunking in ELF by Anna Mauranen chunking is characterized as one of the fundamental features of LUG. She connects chunking to linearity and temporality of speech processing and states that «[c]hopping up the incoming speech stream into chunks seems an efficient way of coping with it, and in line with our other perceptual processes» (Mauranen, 2009: 220). She further suggests that these chunks of up to five words are formed 'naturally'.

«In LUG, we take chunking to be an intuitive perceptual response to the incoming speech stream. There is no reason therefore to expect it to differ in L1 and L2 speech in principle; [...] LUG takes chunking as a pretheoretical term, and we use our own intuitive capacity to perform chunking on stretches of transcribed speech» (Mauranen, 2009: 220).

In the following comparison I will state some characteristics and, basically, weak points of the LUG-approach, and then try to show how, in my view, they can easily be overcome, for instance by introducing some principles taken from usage-based approaches to grammar.

3.1 Intuition

In LUG chunking is an intuitive process, it is natural and unavoidable.

In the usage-based model chunking is not innate, but a domain-general process based on experience, repetition, and conventions.

Considering a broader perspective on cognitive processes in linguistic production, it is questionable and perhaps superfluous to insist, in LUG, on the pre-theoretical, 'intuitive' nature of chunking. In theoretical terms, the term 'intuitive and pre-theoretical' could be replaced, for instance, by 'domain-general cognitive processes, based on experience' as defined by usage-based approaches to grammar. In this way, Mauranen's claim about linguistic chunking as being «in line with other perceptual processes» (Mauranen, 2009: 220) can be reinterpreted in a theoretically more appropriate way.

From this starting point we can replace the view of chunking as a ‘natural’ process that takes place ‘intuitively’ with a more scientific description.

Sinclair and Mauranen also insist that LUG does not highlight the recurrent aspect of chunking, but instead its on-line, linear aspect, where the chunks are not (necessarily) conventionalised because «we use our own [more general, intuitive] capacity for chunking up language» (Sinclair and Mauranen, 2006: 40). But in grammar, the least one would expect is to get an explanation of its central concept, in this case ‘intuition’: how it is formed, who has this intuition, what influences it etc. Another question is whether external factors like conventions are perhaps not one of the pre-requisites for the existence of this kind of capacity. Indeed, the usage-based accounts can provide some plausible answers to these questions, simply because for them the ‘intuition’ of how to divide a text into chunks is shaped by experience. The experience comes with usage, that is, with everything we encounter both linguistically and extra-linguistically. This does not mean that intuition should be entirely banned from linguistic investigation or description. It does imply, however, that ‘intuition’ is a very vague concept and – if used in a theory – should be defined more thoroughly.

3.2 Variability

Furthermore, chunking varies between speakers, according to LUG. As obvious as this sounds, the theory does not systematize, explain or comment on it, which is what a robust model would demand. To solve this problem we can, again, turn to the usage-based accounts: in their theory, chunking differs according to experience, i.e. the speaker’s previous usage. The more similar the experience, the more similar and cognitively entrenched (i.e. present in long-term memory), and the more similar the process of chunking between speakers. In *Linear Unit Grammar* (2006) chunking is unfortunately only very briefly associated with perception, learning and entrenchment (Sinclair and Mauranen, 2006: 37). It is interesting that Sinclair and Mauranen even name Bybee as one of the scholars who «emphasize social interaction in the shaping of grammar» and are «compatible with our approach» (Sinclair and Mauranen, 2006: 38). One can hence only regret that this direction is not developed further in LUG.

3.3 Examples

In order to draw a more concrete comparison, it seems also very productive to see if the chunks from LUG can be interpreted as (the conventionalised)

chunks in usage-based accounts. Indeed, most of the chunks from the corpus of English as a lingua franca in academic settings (ELFA) in Sinclair and Mauranen (2006: 57; see (1) below) would also be identified as chunks in Bybee's sense. (1) and (2) below draw a parallel between a text from ELFA divided into chunks using LUG and individual chunks or types of chunks from Bybee (2010):

1. the Estonian	(2) small, frequent chunk (cf. Bybee, 2010: 35)
2. it was an article	larger, composed chunk (cf.: 25)
3. i read	small, frequent chunk (cf.: 35)
4. it was a famous Estonian tele-	larger, composed chunk (cf.: 25)
5. television	small chunk: word (cf.: 35)
6. i don't know	«I don't know» (cf.: 5)
7. reporter	small chunk: word (cf.: 35)
8. or something	prefabricated expressions (cf.: 35)
9. he went on strike	formulaic or prefabricated sequences of words (cf.: 34)
10. on the hunger strike	prefabricated unit (cf.: 28)

To sum up this section, as Bybee notes in her article from this year (Bybee, 2013: 68), the usage-based approaches provide «a linguistic theory with powerful explanatory possibilities», because they take into consideration the change and dynamics of representations and are based on domain-general processes. Unlike these approaches, LUG is a powerful tool for handling linguistic data, but has, in my opinion, little explanatory power when it comes to the functioning and generation of linguistic communication. Therefore it would be very productive to combine the principles of both kinds of approaches and show in practice the relevance of LUG as a model.

4. *Emergence, on-line production of forms and 'ad hoc' constructions*

The so-called emergence of constructions is another central point in the usage-based model, in LUG and other recent grammar theories: linguistic structures can emerge constantly, so they can be *emergent* and not (necessarily) stable and well-defined (in advance). According to Hopper (1998), there is actually no stable, definite state of an adult grammar: language acquisition is never fully completed, grammar is constantly *emerging*, even a competent speaker can modify, extend and change it. Performance is shaped

by social, cultural and discursive forces: «Structure, or regularity, comes out of discourse and is shaped by discourse in an ongoing process. Grammar is, in this view, simply the name for certain categories of observed repetitions in discourse» (Hopper, 1998: 156). In his more recent work Hopper sees grammar as ephemeral and passing (cf. Hopper, 2011: 26), and therefore also language acquisition can never be considered as finished. The form of the structure can be either conventional or formed *ad hoc*.

In the same sense as Hopper, also Barbara Seidlhofer (quoting Cameron and Larsen-Freeman) claims that «concepts such as “end-state” grammars become anomalous» (Cameron and Larsen-Freeman, 2007: 230, cited in Seidlhofer, 2011: 99). In ELF-studies, the emergence and *ad hoc* production is in general repeatedly emphasized. Seidlhofer stresses that ELF-speakers use all the linguistic features they know in order to achieve a communicative goal and therefore a lot happens *ad hoc*. The negotiation of meaning is at work in the concrete situation and is influenced by a specific context. Similarly, Jenkins, Cogo and Dewey point out that «[s]peakers routinely – but not unvaryingly – exploit the language systems of English to the extent that we can identify EMERGING PATTERNS of lexical and grammatical forms» (Jenkins *et al.*, 2011: 288-289; emphasis in original). Finally in LUG, as its name suggests, linearity, emerging utterances and the on-line dealing with the speech stream are, of course, key principles of the approach (cf. Sinclair and Mauranen, 2006: 88f., 136f.).

Regarding the on-line processing, Holger Diessel observes that there should be more emphasis put on this phenomenon in the usage-based theory as well: «The sequential decision-making process is at the heart of language use; it determines the language users’ linguistic behavior and the development of linguistic structure over time» (Diessel, 2011: 841).

The notion of *ad hoc* constructions and the on-line production of forms is a matter of constant debate among linguists and is very problematic. On the one hand, something that is produced *ad hoc*, is dependent on the very situation, the context, the speakers in the concrete interaction etc. In this respect, an analysis of this kind of data goes beyond the scope of possible scientific analysis. On the other hand, *ad hoc* structures definitely constitute a key concept in linguistic production and should therefore be accounted for in some way. Theoretical approaches to these phenomena try to deal with it by, for instance, transforming *ad hoc* structures into well-formed sentences (the so-called ‘Step 5’ in LUG, cf. Sinclair and Mauranen, 2006: 96ff.) or locating the investigation on some kind of ‘local micro-level’ (cf. Zima and Brône, 2011: 266) for specific speech situations. Both of them are, however, very problematic for linguistic theories aspiring to generality.

5. *Pairings of form and function*

In this final section of the paper, I will just briefly draw a last parallel between usage-based accounts of grammar and ELF-studies, i.e. by comparing constructions as form-function mappings with the unusual and dynamic pairings of form and function that are indicated by the special, recently developed POS-tagging of VOICE Corpus. In the *Part-of-Speech Tagging and Lemmatization Manual* (VOICE Project, 2014: 11) these tags are characterized in the following way (emphasis in original):

«For all tokens in the corpus, separate tags for paradigmatic form and syntagmatic function are assigned. The tag for form is indicated first, followed by a tag for function, given in brackets.

Format: **FORM-tag(FUNCTION-tag)**

There are **2 options** of this format:

OPTION 1: form and function converge → identical form (function) tag is assigned, e.g. a **house_NN(NN)**

OPTION 2: form and function do not converge → different tags for form and (function) are assigned, e.g. **two house_NN(NNS)**».

I would suggest that this can be directly connected to the form-function mapping, which the usage-based model (and especially Construction Grammar as one of the best known manifestations of it) attempts to represent in its theory (cf. Bybee, 2010: 9-10), as illustrated by the following schematic representation:

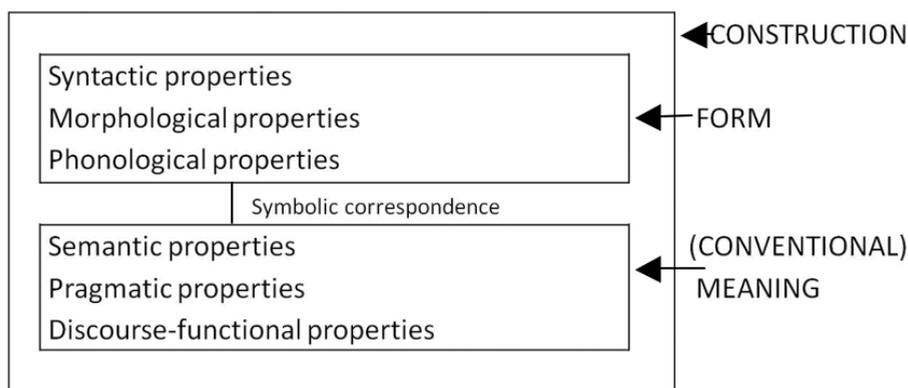


Fig. 1 – The structure of a construction according to Croft (2001: 18)

Without making a thorough analysis and comparison, it is evident that in both approaches the two levels are essential and inseparable, and are taken into consideration in every description. So it seems that the concrete device for the POS-annotation of *non-native discourse* directly resembles a coherent theoretical principle from recent approaches to grammar.

6. Conclusions

The present discussion of some of the key concepts in usage-based accounts of language has shown important convergences with some conceptualisations and descriptions of ELF. The online production of forms, emergent structures, chunks as central units, co-constructing language, *ad hoc* constructions and form-function units play a central role in both research fields. It seems particularly interesting to compare the notion of chunking in the usage-based accounts and in Linear Unit Grammar. But it has also become clear that the formalisation and concrete modelling of ‘ungrammatical’ data (for instance, *non-native discourse*) is very complicated and has not yet been taken into account sufficiently. In many respects it is still debatable whether grammatical approaches to *non-native discourse* are possible at all: the question, how to model, systematize and conceptualize the dynamic, the unsystematic and the irregular therefore remains open.

¹ See also the passages in Seidlhofer (2011) about sedimentation (2011: 114) and the respective passages in Pennycook about «sedimented products of repeated acts of identity» (Pennycook, 2007: 73).

² This is one of the points of difference in relation to Chomskyan generative grammar. In ELF-publications the distinction competence/performance is, however, maintained. This is especially striking in the term of *virtual language*, which is supposed to be an «underlying abstract set of rules» (Seidlhofer, 2011: 112), common to ELF and English as a Native Language.

³ In her attempt to apply LUG to ELF Mauranen points out that ELF is «fundamentally normal language use despite some surface deviations from Standard English» (Mauranen, 2009: 218).

REFERENCES

- Brazil, D. 1995, *A grammar of speech*. Oxford: Oxford University Press.
- Bybee, J. 2010, *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, J. 2013, Usage-based theory and exemplar representation of constructions. In Hoffmann, T. and Trousdale, G. (eds.), *The Oxford handbook of construction grammar*. Oxford: Oxford University Press, 49-69.
- Bybee, J. and Beckner, C. 2010, Usage-based theory. In Heine, B. and Narrog, H. (eds.), *The Oxford handbook of linguistic analysis*. Oxford: Oxford University Press, 827-855.
- Cameron, L. and Larsen-Freeman, D. 2007, Complex systems and applied linguistics. *International journal of applied linguistics*, 17(2), 226-239.
- Croft, W. 2001, *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Davies, A. 2003, *The native speaker: Myth and reality*. Clevedon: Multilingual Matters.
- Diessel, H. 2011, Language, usage and cognition. By Joan Bybee. *Language*, 87, 830-844.
- English as a Lingua Franca in Academic Settings (2008). The corpus of English as a lingua franca in academic settings. Director: Anna Mauranen. Retrieved from <<http://www.helsinki.fi/elfa/elfacorporus>> (last access 15.01.2014).
- Hopper, P. 1998, Emergent grammar. In Tomasello, M. (ed.), *The new psychology of language: Cognitive and functional approaches to language structure*. Mahwah, N.J.: Laurence Erlbaum Associates, 155-176.
- Hopper, P. 2011, Emergent grammar and temporality in interactional linguistics. In Auer, P. and Pfänder, S. (eds.), *Constructions: Emerging and emergent*. Berlin-Boston: De Gruyter, 22-44.
- Hunston, S. and Francis, G. 2000, *Pattern grammar*. Amsterdam: John Benjamins.
- Jenkins, J., Cogo, A. and Dewey, M. 2011, Review of developments in research into English as a lingua franca. *Language teaching*, 44(3), 281-315.
- Kaltenböck, G. 2011, Explaining diverging evidence: The case of clause-initial *I think*. In Schönefeld, D. (ed.), *Converging evidence: Methodological and theoretical issues for linguistic research*. Amsterdam-Philadelphia: John Benjamins, 81-112.
- Langacker, R. 1987, *Foundations of cognitive grammar, vol. 1: Theoretical prerequisites*. Stanford, CA: Stanford University Press.
- Mason, O. 2007, Review of Linear unit grammar: Integrating speech and writing. By John McH. Sinclair and Anna Mauranen (2006), *The*

- Electronic Journal for English as a Second Language (TESL-EJ)*, 11(2). Retrieved from <<http://www.cc.kyoto-su.ac.jp/information/tesl-ej/ej42/r3.html>> (last access 12.09.2014).
- Mauranen, A. 2009, Chunking in ELF: Expressions for managing interaction. *Journal of intercultural pragmatics*, 6(2), 217-233.
- Pennycook, A. 2007, *Global Englishes and transcultural flows*. London-New York: Routledge.
- Seidlhofer, B. 2011, *Understanding English as a lingua franca*. Oxford: Oxford University Press.
- Sinclair, J. and Mauranen, A. 2006, *Linear unit grammar*. Amsterdam-Philadelphia: John Benjamins.
- Vienna Oxford International Corpus of English Project 2014, *VOICE part-of-speech tagging and lemmatization manual*. Retrieved from <http://www.univie.ac.at/voice/documents/VOICE_tagging_manual.pdf> (last access 12.09.2014).
- Zima, E. and Brône, G. 2011, Ad-hoc-Konstruktionen in der Interaktion: eine korpusbasierte Studie dialogischer Resonanzzeugung. In Lasch, A. and Ziem, A. (eds.), *Konstruktionsgrammatik III. Aktuelle Fragen und Lösungsansätze*. Tübingen: Stauffenburg, 255-271.