



CogniTExtes

Revue de l'Association française de linguistique cognitive

Volume 19 | 2019

Corpora and Representativeness

A rejoinder to Maarten Lemmens's paper 'In defence of frequency generalisations and usage-based linguistics. An answer to Frederick Newmeyer's "Conversational corpora : when big is beautiful"'

Frederick J. Newmeyer



Electronic version

URL: <http://journals.openedition.org/cognitextes/1657>

DOI: 10.4000/cognitextes.1657

ISSN: 1958-5322

Publisher

Association française de linguistique cognitive

Electronic reference

Frederick J. Newmeyer, « A rejoinder to Maarten Lemmens's paper 'In defence of frequency generalisations and usage-based linguistics. An answer to Frederick Newmeyer's "Conversational corpora : when big is beautiful"' », *CogniTExtes* [Online], Volume 19 | 2019, Online since 17 June 2019, connection on 19 June 2019. URL : <http://journals.openedition.org/cognitextes/1657> ; DOI : 10.4000/cognitextes.1657

This text was automatically generated on 19 June 2019.



CogniTExtes est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

A rejoinder to Maarten Lemmens's paper 'In defence of frequency generalisations and usage-based linguistics. An answer to Frederick Newmeyer's "Conversational corpora : when big is beautiful"'

Frederick J. Newmeyer

EDITOR'S NOTE

Please also read Frederick Newmeyer's position paper (<http://journals.openedition.org/cognitextes/1584>) and Maarten Lemmens' response paper in this issue (<http://journals.openedition.org/cognitextes/1616>).

1. Some introductory remarks

- 1 First, I must express my heart-felt gratitude to Maarten Lemmens for writing a collegial, thought-provoking, and informative reply to my position paper. I learned a great deal from it and can honestly say that if I pursue the issues that I brought up in my piece I will have much recourse to Lemmens's reply.
- 2 That said, I found this an extremely difficult rejoinder to write. The reason for my difficulty derives from the great disconnect between the content and claims of my position paper and those of the reply to it. My piece was, I have to say, very modest in its conception and realization. It was in no sense either a defence of generative grammar nor a critique of usage-based linguistics. As far as the former is concerned, I avoided

discussion of most of the topics central to the generative approach: innateness and the poverty of the stimulus (though Lemmens thought it appropriate to bring them up), the autonomy of syntax, rule formalisms, extraction constraints, binding conditions, parameter settings, and other notions explored by the theory. As far as the latter is concerned, I did not even begin to scratch the surface of usage-based approaches. How could I have done so in a 20-odd page paper, when the various usage-based approaches, of which Lemmens's model is only one, differ more from each other than do generative approaches? (One thinks of models as different as the Emergent Grammar (Hopper 1987), which sees structure as ephemeral, models drawing from work in information theory and AI (Jaeger 2010), Radical Construction Grammar (Croft 2001), models giving centre-stage to grammaticalisation (Heine and Kuteva 2007), parsing-based approaches (Hawkins 2009), and models of Construction Grammar like that of Fillmore and Kay (1993) that have more in common with the formal theory HPSG than they do with Cognitive Construction Grammar (Goldberg 2003).)

- 3 No, the goals of my position paper were extremely modest. I merely wished to show that some of the more extreme claims made by some usage-based grammarians were empirically false and that their falsity derived from confining themselves to a too-small data base. I certainly did not say (or wish to imply) that once this fact is acknowledged, differences between the two camps would evaporate. I merely wanted to point out that one (and only one) major area of dispute between the two camps might in part be neutralised. I found Lemmens's explication and defence of (his version of) usage-based linguistics interesting and potentially useful to me should I pursue investigation of these matters, although often disconnected from the major points defended in my position paper.
- 4 Hence, in the remainder of this rejoinder I will confine my remarks to the extent possible to those issues that I discussed in more than superficial detail in my position paper.

2. Some more on the importance of corpus size

- 5 Lemmens appears skeptical that the position of Sandra Thompson (one of the two authors of one of the two papers that I critique) is that 'what we think of as grammar may be best understood as combinations of reusable fragments' (Thompson 2002: 141), even though I provided a direct quote. In that same paper, she endorses the idea that grammar is 'constituted of actual bits of texts which are remembered, more or less, and then retrieved to be reshaped to new contexts' (Becker 1984: 435) and that 'everyday language is built up out of combinations of ... prefabricated parts. ... a kind of pastiche, pasted together in an improvised way out of ready-made elements' (Hopper 1987: 144). Given those quotes, I felt that I was totally within my rights to assert that (some) usage-based grammarians see grammar as a collection of stock memorized phrases and very simple constructions. I even cited work — un-commented upon by Lemmens — from Altenberg (1998) that found 80% of the words in the London-Lund Corpus to be formulaic and from Erman and Warren (2000) that estimated that 58.6% of spoken texts are filled with what they call 'prefabs'. Both estimates are fully compatible with my quote above.
- 6 Perhaps elsewhere Thompson takes a more moderate position, but whether Thompson's views on the nature of grammar are consistent or not is quite irrelevant. The point is that such thinking derives in large part from looking at a minuscule amount of extremely

informal conversational data among intimates and then generalising that to the data upon which one might construct a grammatical theory.

- 7 Thompson and Hopper claim that 'Get is a prime example of a verb with no easily imagined argument structures, precisely because it is used in so many lexicalized "dispersed" predicates and specific constructions' (Thompson and Hopper 2001: 49). I demonstrated that that claim is false. Thompson and Hopper noted correctly that 'discussions for argument structure have to date been based on fabricated examples rather than on corpora of ordinary everyday talk'. So I took a *much larger* corpus than theirs and demonstrated that large corpora refute the claims in the quoted sentence above. Interestingly, Lemmens does not dispute my argumentation here, which might (or might not) suggest that he accepts it.
- 8 Let us turn to my critique of Miller and Weinert (1998). Lemmens is correct to point out that their corpora contain utterances of Scottish adolescents and mine of American adults. If the sentence-types that I used as exemplars are more likely to be uttered by the latter than by the former, then my point about the variety of construction types mastered by speakers holds only for American adults — but it still holds! However, I have always found the full quotation from Miller and Weinert (cited by Lemmens) to be borderline incomprehensible, even after several readings. I did not understand what they mean by 'boundary conditions' or by the claim that 'sentences breaking through the boundaries occur neither in speech nor in writing'. If these 'boundary conditions' (whatever they are) are meant to be relevant only to the speech of Scottish adults, then what they write is in all probability false, since it seems unlikely to me that there would be any significant difference in syntactic competence between adult Scots and adult Americans. If they are making claims about the speech of Scottish adolescents, then given their minuscule corpora, we have no idea if their claim is true or false.
- 9 Lemmens makes the following remark with respect to Miller and Weinert's book:

In other words, if one reads their work carefully, one sees that their point is not so much to present a comprehensive account of all English constructions possible in conversation, but rather to show that the constructions that they find in their spoken data differ quite strongly from those discussed in the generative literature which quite legitimately leads one to question the validity of the latter, in particular in view of language acquisition.
- 10 Some sentence-types differ from genre to genre? I would say 'So what?' My primary goal is to capture the grammatical competence of a native speaker (that is, constructing a formal account of what is grammatical in their speech variety). It is interesting to observe that some sentence-types are more common in some genres than others and to develop models that explain why that might be, but doing so is irrelevant to my primary goal.
- 11 I went on, citing Biber (1988), to demonstrate that the differences in construction types that appear from genre to genre are much smaller than what one might be led to believe, given some writings of usage-based linguists. Lemmens interprets my remarks here 'as criticism of Biber's work on linguistic variation in genre'. But the opposite is true. My reference to Biber was intended as *support* for my position. The discussion around Table 1 indicates my delight that differences in frequency type from genre to genre are very small. Given that the data gleaned from face-to-face conversations and academic prose are so similar, we have one more reason to question the bold assumption of many usage-based linguists that certain types of genres are more central than others for theory construction.

3. Frequency and related issues

- 12 My paper ended with a few sketchy suggestive remarks on the limitations of conversational corpora. In retrospect that may have been a mistake, as it led Lemmens to largely focus his critique on this specific issue, rather than engage with the contentful material earlier in my paper. Any one of those subsections could have been developed into a full-length article. It would perhaps have been wiser of me to end my paper with section 6 and to have left discussion of corpora limitations for future work.
- 13 Nevertheless, there are aspects of these remarks that need to be brought up here, in particular around the issue of 'frequency'. Lemmens turns to discussions of frequency over and over again in his reply, using the words *frequent* and *frequency* 60 times (as opposed to 21 in my position paper). For the most part, what he writes on this topic is relevant to a broad discussion of the pros and cons of generative and usage-based linguistics, but tangential to the limited claims of my position paper. As a result, I find astonishing both the title of his reply, 'In Defence of Frequency Generalisations' and the first paragraph of his conclusion. Here he writes:
- Newmeyer is quite sharp in his rejection of the studies by usage-based linguists who argue for the importance of frequency-based generalizations to the structure of grammar.
- 14 Lemmens's account of my paper is erroneous, and needs to be qualified. In fact, here are the first two sentences of my own slightly-over-one-page discussion of frequency:
- Frequency of use, as calculated on the basis of evidence from conversational (and other) corpora is uncontroversially an important factor in directing grammatical change. Frequency drives the grammaticalization of locative nouns to adpositions, pronouns to person markers, auxiliaries to tense and aspect particles, and much much more.
- 15 I go on to offer a 'word of caution' about overreliance on frequency (giving a few innocuous examples) and conclude the section with the statement:
- In sum, frequency is an important factor leading to the shaping and reshaping of grammar. But appeals to frequency should never be used as a substitute for careful grammatical analysis. Frequency generalizations derived from conversational corpora do not challenge theories constructed on the basis of introspective judgments.
- 16 Let me hence reassert that I do not reject frequency-based generalisations, but rather encourage linguists not to rely exclusively on them. Indeed, I myself appeal to frequency-derived data in my critique of Thompson and Hopper and in my discussion of Bresnan et al. Lemmens disagrees with the third (final) sentence above, but only because he misinterprets it. I stand by my claim that conversational corpora do not challenge *theories* constructed on the basis of introspective judgments, but they most certainly contradict particular *analyses* so constructed. As I wrote, conversational corpora are 'quite useful for grammatical theorists'. Let me here take the opportunity to provide more details on the Fisher corpora, as requested by Lemmens in his reply. The corpora include conversations between participants who represent a cross-section of American English speakers, in terms of gender, social class, region, and other variables. Subjects were given pre-determined topics to discuss with total strangers over the phone. For most the topic was terrorism, though they were encouraged to digress from that topic as they wished. I would argue that one can learn more about the linguistic resources of a speaker from a

task like these participants were given than from a back-and-forth between intimates in a social situation, but, alas, that is a topic for future discussion.

- 17 As I mentioned, I had hoped in my position paper to avoid 'big' questions of linguistic theory, focussing to the extent possible on questions of the relevance of corpus size. Lemmens evidently believes that one cannot discuss one without the other, so it is obvious that I need to make my broader position as clear as possible. I am not a 'classical' generative grammarian. Although, I continue to advocate and argue for the autonomy of syntax, for some years I have expressed skepticism about the innateness of purely grammatical constructs. As a consequence, Lemmens' paragraphs in §2.4 on usage-based theories of acquisition have no relevance to my paper. My overall position on grammatical theory, expressed in such publications as Newmeyer (1998, 2003, 2005), is that the 'Saussurean' distinction between grammatical knowledge and language use is well motivated, and that (perhaps surprisingly) such a distinction is wholly compatible with functional / usage-based explanation of the properties of grammatical systems. Indeed, the paper cited by Lemmens (Newmeyer 2003) endorses functional explanation throughout. It cites the work of many more functionalists than formalists and most of these citations are favourable (pp. 683-687, 694-695, 700-701). Martin Haspelmath (2000: 235), on the basis of my 1998 book, has gone so far as to describe me as 'the functionalist Chomskyan', a characterization that I feel to be right on the mark. For what it's worth, I take the compatibility of formalism and functionalism to be the mainstream tradition within world linguistics for most of the past century, a tradition which is represented most notably by the work of such individuals as Roman Jakobson and the syntacticians of the Prague School (for discussion, see Newmeyer 2001) and linguists intellectually indebted to them such as Simon Dik and André Martinet.
- 18 I do not reject the generalisations that led Lemmens to write that 'usage-based linguists see grammar as an adaptive, co-evolving system of mappings of patterns with communicative intent' and I would also agree with Joan Bybee's comment that 'usage feeds into the creation of grammar just as much as grammar determines the shape of usage' (2006:730). The big question for me is what is 'in the grammar itself' and what pertains to systems affecting the grammar. Here I follow staunch generativists like Peter Culicover and Ray Jackendoff (2005) in advocating formal treatments of such constructions as 'to X one's head off' without denying the acquisitional and usage-based matrices that give rise to such constructions (for example, the idea, discussed by Lemmens, that 'high type frequency leads to abstraction which increases productivity'). But I believe that Lemmens goes too far in writing 'that grammar is learned from, and subsequently shaped by usage ... Logically then, one cannot omit usage from grammar'. I would reply to that statement also in appealing to a general point of logic: From the fact that Y is learned from X (shaped by X, beholden to X, etc.), it does not follow logically that X is part of the same system that characterizes Y. But there is hardly the space in this short rejoinder to delve more deeply into these foundational questions, which remain open for scientific discussion.

BIBLIOGRAPHY

- Altenberg, Bengt. 1998. On the phraseology of spoken English : The evidence of recurrent word combinations. In A. P. Cowie (ed.), *Phraseology : Theory, analysis and applications*, 101-122. Oxford : Clarendon Press.
- Becker, A. L. 1984. The linguistics of particularity : Interpreting superordination in a Javanese text. *Berkeley Linguistics Society*, 10, 425-436.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge : Cambridge University Press.
- Bybee, Joan L. 2006. From usage to grammar : The mind's response to repetition. *Language*, 82, 711-733.
- Croft, William. 2001. *Radical construction grammar : Syntactic theory in typological perspective*. New York : Oxford University Press.
- Culicover, Peter W., & Raymond Jackendoff. 2005. *Simpler syntax*. Oxford : Oxford University Press.
- Erman, Britt, & Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text*, 20, 29-62.
- Fillmore, Charles J., & Paul Kay. 1993. *Construction grammar coursebook*. Berkeley : Copy Central.
- Goldberg, Adele E. 2003. Constructions : A new theoretical approach to language. *Trends in Cognitive Science*, 7, 219-224.
- Haspelmath, Martin. 2000. Why can't we talk to each other ? *Lingua*, 110, 235-255.
- Hawkins, John. A. 2009. An efficiency theory of complexity and related phenomena. In G. Sampson, D. Gil & P. Trudgill (eds.), *Language complexity as an evolving variable*, 252-268. Oxford : Oxford University Press.
- Heine, Bernd., & Tania Kuteva. 2007. *The genesis of grammar : A reconstruction*. Oxford : Oxford University Press.
- Hopper, Paul J. 1987. Emergent grammar. *Berkeley Linguistics Society*, 13, 139-157.
- Jaeger, T. Florian. 2010. Redundancy and reduction : Speakers manage syntactic information density. *Cognitive Psychology*, 61, 23-62.
- Miller, Jim, & Regina Weinert. 1998. *Spontaneous spoken language : Syntax and discourse*. Oxford : Clarendon.
- Newmeyer, Frederick J. 1998. *Language form and language function*. Cambridge, MA : MIT Press.
- Newmeyer, Frederick J. 2001. The Prague School and North American functionalist approaches to syntax. *Journal of Linguistics*, 37, 101-126.
- Newmeyer, Frederick J. 2003. Grammar is grammar and usage is usage. *Language*, 79, 682-707.
- Newmeyer, Frederick. J. 2005. *Possible and probable languages : A generative perspective on linguistic typology*. Oxford : Oxford University Press.
- Thompson, Sandra A. 2002. 'Object Complements' and conversation : Towards a realistic account. *Studies in Language*, 26, 125-164.

Thompson, Sandra A., & Paul J. Hopper. 2001. Transitivity, clause structure, and argument structure : Evidence from conversation. In J. L. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 27-60. Amsterdam : John Benjamins.

AUTHOR

FREDERICK J. NEWMAYER

University of Washington, University of British Columbia, and Simon Fraser University