



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
Main Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2014

---

## **Comment on Klaas Willems' contribution**

Kabatek, Johannes

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-105372>

Journal Article

Originally published at:

Kabatek, Johannes (2014). Comment on Klaas Willems' contribution. *Energieia*, 5:113-114.

### Comment on Klaas Willems' contribution

I completely agree and think it is obvious that frequency data must not be denied since they are important indicators for language dynamics. But frequency data, as well as introspection data, are more a starting point than an end point of a possible explanation of linguistic dynamics (as to the concept of explanation, see Esa Itkonen's paper). I would like to add three further points to Klaas Willem's reflections. The first refers to entrenchment, the second to "corpus effects" and the third to the notion of discourse traditions.

- obviously, what we find in a corpus is something else than what we find in linguistic competence. We find a collection of utterances that might be more or less "representative" (whatever that means: we are on the level of utterances and maybe utterances can never be "representative" for something on another level).<sup>1</sup> In practice, we find what we get easily (newspapers and internet data in current research, much of literature and official documentation in the past). And there are lots of texts we will never get (a fact that might be appreciated for ethical reasons, but not for linguistic research). And there are even contrasts between competence and utterances: if we look at certain taboo words, we see that they appear with low frequency in corpora but we know at the same time that they might be transmitted from one generation to another with a certain stability. Research on frequency claimed that high frequency is an indicator for high "entrenchment", i.e. cognitive anchorage of linguistic forms. Even if this seems to be roughly true, there is no one-to-one causal relationship for that. Forms are present with a certain independence of their overt realization. Forms are virtually ready to be pronounced, but they are actually pronounced due to communicative purposes. When we observe everyday language production statistically (like in the lexical observatory projects that exist currently for several languages), we can see cyclic effects like the high frequency of month names every twelve months (April is very frequent in April, but does this say something about language dynamics?).

- "corpus effects" are those effects produced due to the selection of texts in a corpus but that are no real indicators of language change. We can observe, to put an example, that in the huge historical Spanish corpus CORDE, some colloquial forms appear for the first time in the 18<sup>th</sup> century. We can infer that these forms exist from the 18<sup>th</sup> century onwards, but not that they had not existed before (as Klaas Willems states: "corpora cannot be used to prove what is *not* possible in a language"). In fact, in many cases, it seems rather to be the fact that a certain tendency of introducing colloquial forms especially in private correspondence makes appear forms that probably have existed long before. I have called this tendency "oralization".<sup>2</sup> Oralization is of course an indicator of a dynamic process, that of writing down something that used to be only a fact of the spoken language, but the spread of a form from spoken to written language should be separated from the innovation of a form. Another relevant notion

---

<sup>1</sup> Cf. J. K. (2013): "¿Es posible una lingüística histórica basada en un corpus representativo?", *Iberoromania* 77 DOI 10.1515

<sup>2</sup> Cf. J. K. (2013): "Corpus histórico, oralidad y oralización", in: Béguelin-Argimón, Victoria, Cordone, Gabriela, de La Torre, Mariela: *En pos de la palabra viva: huellas de la oralidad en textos antiguos. Estudios en honor al profesor Rolf Eberenz*, Bern: Lang, 37-50.

---

is that of *hapax*, not only as a lexical but also as a morphological or syntactic phenomenon. A historical corpus might show unique forms that centuries later in history appear as features of that language, and the low frequency or the inexistence in the corpus during that time can be, but not necessarily is, an indicator of inexistence in the language.

- the previous remark in a way refers also to the notion of discourse traditions, traditions of texts that have to be considered when studying language dynamics. Sometimes forms are very frequent in a particular DT or *genre*, but not in others, and frequency changes can be due to increases of productions of a certain genre or to a higher presence of that genre in a corpus as well as due to increase in several genres or in the language as such (if we believe, as we do, that the notion of DT should not replace that of a language and language should not be reduced just to a collection of discourse traditions).