

***Over-uh-load.* The occurrence of *uh(m)* between elements of compounds during interpreting.**

Koen Plevoets & Bart Defrancq

Department of Translation, Interpreting and Communication
Ghent University

Cognitive load is a major source of processing difficulties and disfluencies, such as *uh* or *uhm*, in both spontaneous speech (Levelt 1983; Clark & Fox Tree 2002; Bortfeld et al. 2001; Watanabe et al. 2008) and interpreting (Goldman-Eisler 1967; Mead 2000, Tissi 2000; Cecot 2001). Gile (1995) lists the interpreter's lack over the content and his reduced background knowledge as potential obstacles during interpreting, and for simultaneous interpreting he adds the lack of control over the original speech rate as well as the interference of speaking and listening. In Plevoets & Defrancq (2014) and Plevoets & Defrancq (submitted), the occurrence rate of *uh(m)* was analysed with respect to the delivery rate, the lexical density and the proportion of numbers in both interpretations and in non-interpreted speeches. An interesting qualitative finding was that some *uh(m)*'s occurred between the elements of lexical compounds, e.g. *onderzoeks-uh-gelden* ('research-uh-funds'), *voedsel-uh-middelen* ('food-uh-resources'), *uit-uh-breiden* ('ex-uh-pand') or *Cohn-uh-Bendit*. As such constructions can shed more light on cognitive load, this paper will investigate them further.

This paper will make a comparison between interpreted language and spontaneous speech in two corpora. The corpus of interpreted language was compiled at Ghent University between 2010 and 2013. It consists of French, Spanish and Dutch interpreted speeches in the European Parliament from 2006 until 2008, which were transcribed according to the VALIBEL guidelines (Bachy et al. 2007). For the purposes of this analysis, a sub-corpus of French source speeches and their Dutch interpretations was selected (amounting to a total of 140 000 words), which has been annotated for lemmas, parts-of-speech and chunks (Van de Kauter et al. 2013) as well as sentence-aligned with WinAlign (WinAlign 2014). The reference corpus for spontaneous speech is the sub-corpus of political debates of the Spoken Dutch Corpus (Oostdijk 2000), compiled between 1998 and 2003. This sub-corpus contains 220.000 words of Netherlandic Dutch and 140.000 words of Belgian Dutch and is annotated for lemmas and parts-of speech.

In both corpora, the number of *uh(m)*'s between elements of compounds were counted for each sentence. The difference proved to be statistically significant, in that interpreters produce these constructions more often than spontaneous speakers. A subsequent analysis pointed to two interesting determinants. First, the compounds by the spontaneous speakers often involved a self-repair of a mispronunciation, e.g. *fundu-uh-mentalisme* ('fundu-uh-mentalism'), *geïmplende-uh-menteerd* ('implende-uh-plemented'), *oorsprangsk-uh-land* ('*origin-uh-country'), *samenwerkingsfron-uh-verbanden* ('collaboration fron-uh-relations'), *tariefafspra-uh-aanpassing* ('tariff agree-uh-adaptation'), *topcro-uh-crimineel* ('top cro-uh-criminal'), or *voortgaand-uh-voorgaand* ('persistent-uh-previous'). Second, the compounds by the spontaneous speakers frequently consisted of a combinations of a word with non-word such as an acronym, a proper noun or a number, e.g. *CBS-uh-cijfers* ('CBS-uh-figures'), *E-uh-313* ('E-uh-313'), *SG-uh-niveau* ('SG-uh-

level'), *vrijstellings-uh-AMVB* ('license-uh-AMVB'), *Millinx-uh-buurt* ('Millinx-uh-neighbourhood'), *Polaris-uh-systeem* ('Polaris-uh-system'), *post-94-uh-traject* ('post-94-uh-trajectory'), and *15-procent-uh-eis* ('15 percent-uh-demand'). These results suggest that the occurrence of *uh(m)* with spontaneous speakers is more related to issues of phonetic realisation of the message than to the memory limitations which are at stake with interpreters.

- Bachy, S., Dister, A., Francard, M., Geron, G., Giroul, V., Hambye, P., Simon, A.C. and Wilmet, R. (2007). *Conventions de transcription régissant les corpus de la banque de données VALIBEL*. University of Louvain-la-Neuve. Available online at http://www.uclouvain.be/cps/ucl/doc/valibel/documents/conventions_valibel_2004.PDF.
- Bortfeld, H., S.D. Leon, J.E. Bloom, M.F. Schober & S.E. Brennan (2001). "Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender". *Language and Speech* 44: 123-147.
- Cecot, M. (2001). "Pauses in simultaneous interpretation: a contrastive analysis of professional interpreters' performances". *The interpreters' newsletter* 11: 63-85.
- Clark, H.H. and Fox Tree, J.E. (2002). "Using uh and um in spontaneous speaking". *Cognition* 84: 73-111.
- Gile, D. (1995). *Basic concepts and models for interpreter and translator training*. Amsterdam: John Benjamins.
- Goldman-Eisler, F. (1967). "Sequential temporal patterns and cognitive processes in speech". *Language and Speech* 10 (3): 122-132.
- Levelt, W. (1983). "Monitoring and self-repair in speech". *Cognition* 14: 41-104.
- Mead, P. (2000). "Control of pauses by trainee interpreters in their A and B languages". *The Interpreters' Newsletter* 10: 89-102.
- Oostdijk, N. (2000). "The Spoken Dutch Corpus: Overview and first evaluation". *Proceedings of the Second International Conference on Language Resources and Evaluation*: 887-894.
- Plevoets, K. & B. Defrancq (2014). "Informational load as a trigger for disfluencies in interpreting". Paper presented at the *UCCTS4* conference, Lancaster, UK.
- Plevoets, K. & B. Defrancq (submitted). "The effect of informational load on disfluencies in interpreting". *Translation and Interpreting Studies*.
- Tissi, B. (2000). "Silent pauses and disfluencies in simultaneous interpretation: A descriptive analysis". *The Interpreters' Newsletter* 10: 103-127.
- Van de Kauter, M., Coorman, G., Lefever, E., Desmet, B., Macken, L. and Hoste, V. (2013). "LeTs Preprocess: The multilingual LT3 linguistic preprocessing toolkit". *Computational Linguistics in the Netherlands Journal* 3: 103-120.
- Watanabe, M., Hirose, K. Den, Y. and Minematsu, N. (2008). "Filled pauses as cues to the complexity of up-coming phrases for native and non-native listeners". *Speech Communications* 50: 81-94.
- WinAlign (2014). *SDL Trados WinAlign Tutorial*. Available at: <http://www.translationzone.com/resources/downloads/winalign-tutorial.html>.