

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/324782271>

The cognitive load of interpreters in the European Parliament: A corpus-based study of predictors for the disfluency uh(m)

Article in *Interpreting International Journal of Research and Practice in Interpreting* · April 2018

DOI: 10.1075/intp.00001.ple

CITATIONS

71

READS

1,131

2 authors:



[Koen Plevoets](#)

Ghent University

49 PUBLICATIONS 609 CITATIONS

[SEE PROFILE](#)



[Bart Defrancq](#)

Ghent University

103 PUBLICATIONS 871 CITATIONS

[SEE PROFILE](#)

The cognitive load of interpreters in the European Parliament

A corpus-based study of predictors for the disfluency $uh(m)$

Koen Plevoets and Bart Defrancq
Ghent University

Cognitive load is a major source of processing difficulties in both interpreting and monolingual speech. This article focuses on measurement of cognitive load by examining the occurrence rate of the disfluency $uh(m)$ in two corpora of naturalistic language: the EPICG, with specific reference to Dutch interpretations of French source texts in the European Parliament; and the sub-corpus of non-interpreted parliamentary speeches from the Spoken Dutch Corpus. In both corpora, the frequency per utterance of $uh(m)$ was studied, in relation to delivery rate, lexical density, presence of numbers and formulaicity (i.e. the number of N-grams), as a Generalised Additive Mixed-effects Model: the frequency of $uh(m)$ in interpretations increases with the lexical density of the source text, while it is inversely related to the formulaicity of both the source text and the target text. These findings indicate the maintenance of a cognitive equilibrium between input load and output load.

Keywords: cognitive load, disfluencies, corpus-based interpreting studies, Generalised Additive Mixed-effects Models

1. Introduction

The interpreter's cognitive load has been a long-standing research topic in interpreting studies. Relations between the multiple, concurrent tasks in interpreting were represented in early models such as those of Gerver (1975) and Moser (1978). Further models were in turn elaborated, from neurolinguistic and pragmalinguistic perspectives, by Paradis (1994) and Setton (1999) respectively. The intricate balancing of demands and capacities in the related processing is at the core of Gile's (2009) Effort Models and Seeber's (2011) Cognitive Load Model. Both of these conceptualise interpreting as a dynamic balance between the cognitive challenges involved in language comprehension, language production and memory storage.

In addition, Gile's Effort Models account for management of this balance as a coordination effort in its own right (Gile 1997).

Research on cognitive load in interpreting has typically focussed on cases where processing demands exceed available capacity, leading to deterioration in the content and/or form of the interpretation. Such instances of 'information overflow' are not uncommon, because interpreters generally work at the limits of their cognitive capacities, as stated in Gile's (1999) tightrope hypothesis. Barik (1975) based his elaborate typology of the interpreter's errors and omissions on three main categories (*omissions*, *additions* and *substitutions and errors*, with many subtypes in each category), while both Dillinger (1994) and Tommola and Helevä (1998) found empirical evidence that the translational accuracy of interpreted texts diminishes in relation to increasing propositional density of the source text.

Barik's typology of 'translation departures' was heavily criticised as subjective by Gerver (1976), and the focus in interpreting studies subsequently shifted to disfluencies (mainly false starts, repairs, and silent or filled pauses), which are regarded in the psycholinguistic literature as 'windows on cognitive load' (Arnold et al. 2000, 2003; Clark & Fox Tree 2002; Levelt 1983, 2003; Swerts 1998; Watanabe et al. 2008; see Bortfeld et al. 2001 for a summary).

In the wake of Goldman-Eisler's (1967) 'pausological' studies, most research on fluency in interpreting has focussed on silent and filled pauses. Tissi (2000), for instance, analysed ten students' interpretations of two excerpts from political round tables, observing fewer (but longer) silent pauses in target texts than source texts; filled pauses showed considerable individual variation. The study was replicated among eleven professional interpreters by Cecot (2001), who reached the same conclusions. Finally, Mead (2000) found that fifteen students produced more filled pauses when interpreting into their B language than from B into A (the results were not significant for silent pauses).

Setton's (1999: 247) view of how various types of disfluency relate to the attention requirements of assimilating and reformulating content is summarised below, as Table 1.

Most research on disfluencies in interpreting has been experimental. By contrast, Plevoets and Defrancq (2016) analysed the frequency of the disfluency *uh(m)* in a corpus of interpretations at the European Parliament, the EPICG (Defrancq 2015). Alongside a 'classic' comparison of source speeches with target speeches, the method was also 'comparable' (Baker 1993) in that Dutch interpretations recorded live at the European Parliament were compared to non-interpreted Dutch speeches by members of the Dutch and Belgian parliaments. Based on these corpus data, Plevoets and Defrancq (2016) examined the role of delivery rate, lexical density, frequency of numbers and average sentence length as inde-

Table 1. Types of disfluencies and their relationship to (the degree of) attention (adapted from Setton 1999: 247)

	Attention to input	Attention to formulation
Long silent pause	High	–
Short pausing	Normal listening	Routine planning
Filled pause	Normal listening	Routine planning
Mixed: short & filled pauses & voice effects	Normal listening	Routine planning
Long filled pause	Relaxed or off	Planning/Searching
Fluent unmodulated string	Relaxed or off	Off

pendent variables affecting the occurrence rate of $uh(m)$: delivery rate and lexical density proved to be important predictors. However, delivery rate was significant only for the source texts, whereas lexical density was significant only in the comparison between the interpreted (target) texts and the monolingual texts. The results reported by Plevoets and Defrancq (2016) thus clearly reflect cognitive load, but they also underline the need to differentiate between the cognitive demands of comprehension or listening and those of language production (which we will call ‘input load’ and ‘output load’, respectively).

Corpus-based research is still sporadic in interpreting studies, despite Shlesinger’s (1998) plea for its more widespread adoption. Overviews of corpus-based interpreting studies can be found in Straniero and Falbo (2012), and Bendazzoli (2017).

By contrast, the increasing popularity of corpora in translation studies dates back to Baker’s (1993) defence of the ‘comparable’ approach, based on comparison between translations and texts produced by non-translators. Such an approach, based as it is on output, can be seen as complementary to the classic, input-oriented concern of interpreting studies with the effects of source text features on the interpretation. The distinction between input and output characteristics seems relevant, as interpreters’ strategies differ for listening (shallow parsing, cognate translation, etc. – see Riccardi 1998) and production (chunking of complex sentences, reordering of information, anticipation, etc. – see Chernov 2004). Similarly, numbers may be more difficult to comprehend than to formulate (Alessandrini 1990; Mazza 2001; Pinochi 2009). Cognitive load can thus be expected to differ between input and output: interpreters have no control over what they hear, but they do have considerable control over what they say.

The study described by Plevoets and Defrancq (2016) is not without its limitations. Firstly, the unit of analysis was each text as a whole: this was acknowledged in the study, the authors’ suggestion being a sentence-level analysis. Secondly, the four indicators of informational load (delivery rate, lexical density, propor-

tion of numbers and average sentence length) were treated as *linear* predictors for the frequency of $uh(m)$. However, as all four variables are numerical, their actual effects can be more complex (with peaks and troughs in distribution). By treating the four predictors as straight lines instead of curves, the analysis may have missed potentially interesting trends. Thirdly, syntactic complexity, which is a well-known predictor of cognitive load (Chmiel & Mazur 2013; Gile 2008), was measured rather crudely as average sentence length. Although average sentence length is a traditional measure of text readability (Flesch 1948), there are much more refined operationalisations for the continuum spanning the extremes of linguistic complexity and formulaicity: one interesting measure is the amount of 'N-grams' or 'multi-word units' (Stubbs 2007), 'recurrent word combinations' (Altenberg 1998) or 'lexical bundles' (Biber et al. 1999), reflecting the predictable or formulaic (non-complex) nature of linguistic sequences. N-grams, investigated quite extensively in corpus linguistics, will be taken into account below (see Section 2).

We address the following research questions:

1. Are the effects of the informational load indicators on the related cognitive demand (measured by the frequency of the disfluency $uh(m)$) the same at sentence/utterance level and text level?
2. To what extent do the informational load indicators for the source text and/or target text predict the frequency of $uh(m)$?
3. As interpreting demands comprehension in addition to speech production, does this change the effect of the informational load indicators on the frequency of $uh(m)$ in interpretation by comparison to spontaneous speech?

In the following parts of the text, Section 2 will define and operationalise the various measures of cognitive load. Section 3 will describe the corpus data. Section 4 will outline the statistical methodology. Results will then be presented and discussed, in Section 5 and Section 6 respectively, followed by conclusions in Section 7.

2. Cognitive load: Definition and measures

The concept of cognitive load was introduced in the 1950s as models of working memory developed (Broadbent 1958; Welford 1952). The limited capacity of working memory means that different processing tasks are constantly competing for cognitive resources. Capacity-vs-demand models of working memory have been specifically adapted by interpreting scholars, Gile's Effort Models and Seeber's Cognitive Load Model being cases in point.

Gile's (2009) Effort Models describe cognitive load as the sum of various cognitive capacity-consuming efforts that coincide or overlap. Drawing on information processing models, Gile argues that cognitive capacity is limited by the amount of information interpreters can handle: the sum of the cognitive resources allocated to different efforts should not exceed the available capacity. Cognitive overload occurs when capacity is insufficient, either in terms of the individual efforts or for the cognitive system as a whole. In Gile (1999, 2008) and Kurz (2008), some aspects of the Effort models are tested empirically with both experimental and observational data. An overview of other studies is presented in Gile (2008).

Seeber (2011:187–189) sees cognitive load as the interference between various demands in “a real-time combination of a language comprehension and a language production task”. The level at which these demands occur can be macro-structural (i.e. perceptual, cognitive or response processing) and micro-structural (i.e. processing of linguistic constituents). In addition, cognitive load is generated by memory demands during the language comprehension task. Seeber also discusses two source text features which increase input load in interpreting: delivery rate and verb-final structures (as in German). He lists various strategies for dealing with verb-final structures. Seeber and Kerzel (2012) present an empirical analysis of cognitive load, based on pupillometry, during simultaneous interpreting of German verb-final structures into English: mean pupil dilation is greater than when interpreting segments with the verb in non-final position. The data also lend support to Gile's (2008) concept of *exported load* (i.e. cognitive load manifesting itself downstream of the difficulty), as pupil dilation differences are at their highest just after the verb-final source text structures.

The present study will investigate cognitive load in a similar way to Plevoyets and Defrancq (2016), analysing the frequency of *uh(m)* as a dependent variable of various triggers. On the one hand, triggers known to increase cognitive load (e.g., delivery rate, lexical density and proportion of numbers) will be examined in both the source texts and the interpretations. On the other hand, we will also look at factors believed to ease cognitive load (e.g., hesitations by the speaker, formulaicity of source and target texts). Delivery rate is a well-known predictor of cognitive load in interpreting, clearly affecting the interpretation (Gerver 1969; Pio 2003), though it tends not to affect language comprehension significantly (Voor & Miller 1965). Lexical density, which leads to lower comprehension and retention scores (Gibson 1993; Kintsch et al. 1975), is seen by Gile (2008) as one of the major causes of cognitive load in interpreting. The presence of numbers, widely acknowledged as a prime source of stress for interpreters (Alessandrini 1990; Gile 2009), triggers a high error count in interpreting (Mazza 2001).

On the positive side, hesitations by speakers are sometimes assumed to offer interpreters the opportunity to catch up with the incoming flow of information (Goldman-Eisler 1967) and consequently decrease their cognitive load. In the same vein, the present study will also include a further parameter assumed to decrease cognitive load for interpreters – i.e. formulaicity.

Various authors have found that formulaicity, or the occurrence of formulaic sequences such as *would you mind* or *as can be seen*, decreases the processing demands of language use. Conklin and Schmitt (2012: 50–53) summarise a number of studies showing the relative ease of processing non-idiomatic formulaic sequences (e.g., *in the middle of*). Ease of processing in relation to formulaicity (which goes hand in hand with recognisability) is clearly seen in a study of readers' eye movements by Underwood et al. (2004), comparing duration and frequency of fixation on the final word in formulaic vs. non-formulaic contexts. Similarly, in a self-paced reading task, Conklin and Schmitt (2008) showed that formulaic sequences are read more quickly than non-formulaic sequences. Tremblay and Baayen (2010) and Tremblay et al. (2011) found that scores in recall tasks were higher for formulaic language sequences, suggesting that these are not processed word by word and thus demand relatively little working memory capacity.

One area of formulaicity research that shows an interesting potential relationship to interpreting is the use of L2 corpora in studies of language learning: while learners' relative lack of confidence tends to make them rely on formulaic sequences more than proficient speakers (Hyland 2008), the latter use a greater variety of these sequences (Chen & Baker 2010; Paquot & Granger 2012). Once again, the relatively low processing cost of formulaic language is apparent here.

Research into the effects of formulaicity on interpreters' performance is scarce. As a general principle, the automation of cognitive operations is believed to decrease interpreters' cognitive load (Gile 1995). Processing of formulaic chunks of language is assumed to be a holistic process, involving no parsing of separate items (Eyckmans 2007; Van Rietvelde et al. 2010) and thus conducive to a degree of automation. Eyckmans (2007) reports that formulaic source text phrasing is positively correlated with fluency in a sight translation task carried out by students of interpreting. Consistent with this, Van Rietvelde et al. (2010) demonstrate that set phrases in the source text used for a simultaneous interpreting task prompt anticipation by student interpreters: the authors' explanation is that formulaic sequences unburden working memory in simultaneous interpreting and leave more available capacity for other cognitive tasks.

A crucial aspect of our analysis is that we will study the four informational load indicators identified above (delivery rate, lexical density, proportion of numbers and formulaicity) in both the source text and its simultaneous interpretation (following Plevoets & Defrancq 2016). The reason is that the interpreter's cognitive

load is expected to increase not only when informational load is higher in the source text, but also when the interpretation is faster and/or lexically/numerically denser. By the same token, we expect lower cognitive load for the interpreter when s/he is both processing and producing more formulaic language. Our analysis is threefold, comparing the input and output load of interpreters with the output load in monolingual speech.

3. Data

The data for this study are taken from two corpora of spoken Dutch: one of interpretations, and one of non-interpreted monolingual speech. The corpus of interpretations is the EPICG: an ongoing project at Ghent University's Department of Translation, Interpreting and Communication, this comprises plenary speeches and their interpretations recorded in the European Parliament between 2006 and 2008. French, English, Dutch and German are source and target languages in the EPICG, together with Spanish as a source language only. The corpus, which is compiled by transcribing the audio-visual recordings in the European Parliament website according to the protocol of the VALIBEL corpus (Bachy et al. 2007), currently totals about 270,000 tokens. For the current study, only the sub-corpus of the French source speeches and their Dutch interpretations will be used: it has been annotated for part-of-speech tags, lemmas and chunks with the *LeTs Pre-process Toolkit* (Van de Kauter et al. 2013). Additionally, it was manually sentence-aligned by means of SDL Trados WinAlign (2014).

The corpus of non-interpreted Dutch is the sub-corpus of parliamentary speeches in the *Corpus Gesproken Nederlands* (CGN), or Spoken Dutch Corpus (Oostdijk 2000). As the parliamentary speeches sub-corpus is part (g) of the CGN, it will be abbreviated below as CGNg. The CGN was a joint project of several Netherlandic and Flemish universities, running from 1998 to 2003. The Netherlandic and Flemish parts of CGNg comprise about 220,000 and 140,000 tokens respectively. All speeches have been tagged for parts of speech and lemmas.

Table 2 gives an overview (text, sentence/utterance and token counts) of the EPICG and CGNg corpora. The number of texts in the EPICG corpus is the same for the French source texts and the Dutch target texts. Comparing the two corpora, the counts for all parameters are far higher for the CGNg, the Netherlandic texts being much longer than the Flemish texts and those in the EPICG corpus: there are only 85 Netherlandic Dutch texts, but they total over 10,000 utterances. Since the distinction between Netherlandic Dutch and Flemish Dutch is relevant only in a sociolinguistic perspective, it will not be taken into account here.

Table 2. Summary overview of the corpora

		Nr. of texts	Nr. of sentences/utterances	Nr. of tokens
EPICG	FRA (source):	107	1,455	39,239
	DUT (target):	107	1,431	34,968
CGNg	TOTAL:	240	19,046	349,058
	(Flanders:	155	8,293	133,006)
	(The Netherlands:	85	10,753	213,052)

4. Method

The first step was to count every instance of *uh(m)* in both the EPICG and the CGNg corpora (*euh(m)* was the conventional transcription in the EPICG). This was done by means of a Python script. Originally, *uh* and *uhm* were counted separately, but this distinction was not maintained because the *-m* variant was infrequent.

The informational load indicators (delivery rate, lexical density, numbers) were operationalised in the same way as in Plevoets and Defrancq (2016), with lexical density and numbers scored at utterance level. Delivery rate was calculated at text level, as the total number of words in a speech divided by its duration in minutes. The rationale for using this method is that, since the time measurements in the EPICG do not correspond to utterance boundaries, computation of delivery rates per utterance is currently not yet possible.

Lexical density was counted as content words divided by the sum of content and function words, the distinction between the two being based on the part-of-speech tags in both corpora. Content words are all nouns, adjectives, adverbs derived from adjectives and non-auxiliary verbs; function words are all articles, pronouns, prepositions, conjunctions, so-called pronominal adverbs (e.g., *daarin*/'therein', *hiernaar*/'hereto', *waarvan*/'whereof' etc. – see Haeseryn et al. 1997: 490–503) and auxiliary verbs.

The frequency of numbers was calculated as a proportion of total words.

In assessing formulaicity, various ways have been proposed to operationalise an N-gram. Some studies do this by statistical association measures such as Mutual Information (see Paquot & Granger 2012 for an overview in L2 research). However, the most widely used criterion is the frequency threshold proposed by Biber et al. (1999): a particular sequence of words constitutes an N-gram if its combined occurrence is at least 10 per million words, and Biber et al. (2004) even use the much stricter cut-off of 40 per million words. It is customary in most research to restrict the analysis to 3-grams and/or 4-grams. Our study will also do so, to obtain a first rough selection of potential 3- or 4-grams. The reason is

the vastly uneven token sizes of our two corpora: neither the French source texts nor the Dutch target texts in the EPICG reach a count of 40,000 words, while the CGNg totals about 350,000 words. We therefore set a frequency threshold of 4 for the EPICG and 35 for the CGNg, which in both cases amounts to a relatively conservative cut-off of 100 per million words. The resulting 3- and 4-grams were checked manually so as to prune away spurious hits. The final list of N-grams for the French source texts contains examples such as the following:

- *de la commission* (of the commission)
- *en tout cas* (in any case)
- *il nous faut* (we need to)
- *droits de l'homme* (human rights)
- *en ce qui concerne* (concerning)
- *c'est ce que* (it is what).

Some examples of the final N-grams in the Dutch texts (both interpretations and monolingual speeches) are:

- *aan de orde* (at stake/hand)
- *de conferentie van* (the conference of)
- *zo snel mogelijk* (as soon as possible)
- *in het kader van* (in the context of)
- *op het vlak van* (regarding)
- *met het oog op* (with a view on).

The full lists can be obtained from the authors.

The quantitative measure for formulaicity is defined as the number of 3- and 4-grams in an utterance, divided by the total number of 3- or 4-word sequences that are in principle possible in the utterance. For instance, an utterance of five words contains three 3-word sequences and two 4-word sequences, so its divisor is 5; an utterance of six words contains four 3-word sequences and three 4-word sequences, so its divisor is seven; in the same vein, the divisor of any other utterance with token size n can be computed as $2n - 5$.

Delivery rate, lexical density, frequency of numbers and formulaicity were evaluated as predictors for the frequency of $uh(m)$, using two analyses. The first analysis was a classic comparison of the EPICG input and output texts: all four parameters were measured in both, so that the sentence alignment of the corpus would make it possible to examine them as eight predictors for occurrence of $uh(m)$ in the target texts. The frequency of $uh(m)$ in the source text (calculated as a mean value in relation to the total words per source utterance) is included as a ninth predictor, because it might shed light on Goldman-Eisler's (1967) view that disfluencies of this type in the input provide opportunities for the interpreter

to ‘catch up’ with the speaker by focusing solely on production – a hypothesis for which Gerver (1975) found no convincing evidence.

The second analysis was a comparison of the Dutch interpretations and the monolingual Dutch speeches: here, the four informational load indicators were measured in relation to occurrences of $uh(m)$ within each of the two samples (i.e. interpreted Dutch vs. non-interpreted Dutch). The distinction is coded as a categorical variable, labelled as *NED_in* and *NED_or* respectively.

The two analyses evidently have a different focus: the first separates the interpreter’s processing of input and output, whereas the second differentiates processing of output in interpreting and monolingual production.

Both analyses were run with a Generalised Additive Mixed-effects Model (commonly abbreviated as GAMM), with a twofold rationale. Firstly, the possibility of the informational load indicators showing a non-linear distribution has already been mentioned (see Introduction) – hence the preference for ‘smoothing splines’ (see Wood 2006), meaning curves which capture the general trend of the data without a precise mathematical formula. Smoothing splines are capable not only of testing whether a predictor is statistically significant, but also of detecting the actual shape of the effect curve. The effects in a GAMM are typically inspected in the form of ‘effects graphs’ (Fox 2003), paying particular attention to the orientation and shape of curves and the shape of the confidence bands.

Secondly, individual observations are measured at utterance level. Table 2 shows a total of 19,046 and 1,431 Dutch utterances in the CGNg and EPICG respectively. As the utterances are ‘nested’ in texts (240 and 107 Dutch texts in the CGNg and EPICG respectively), observations may be more similar within any given text than in different texts, giving rise to spurious variation at the text level. To accommodate for this, we treat the texts as a ‘random factor’: in this way, all estimated effects of interest (i.e. the ‘fixed effects’) actually occur at utterance level (see Baayen 2008: 241–302 and Hox 2010 for accounts of mixed-effects modelling). The sole exception will be for delivery rate, for which the EPICG allows only whole-of-speech computation. Technically, this makes delivery rate a ‘second-level’ predictor (i.e. modelled at text level), while lexical density, numbers and formulaicity are ‘first-level’ (i.e. utterance-level) predictors. However, this means that potential variation of delivery rate within texts will not be taken into account. This is an unfortunate drawback of our data management, which we hope to amend in future versions of the corpus.

The frequency of $uh(m)$ cannot be analysed against a Gaussian distribution, as in linear models (i.e. ANOVA and/or linear regression), and has to be modelled as a Poisson variable (see Agresti 2013: 122–130). However, as the frequency of $uh(m)$ is likely to grow in relation to utterance length, the total number of words in an utterance has to be included as a control variable (or ‘offset’); the

corresponding analysis, which models relative frequency against a total count, is called a 'rate model' (see Agresti 2013: 385–391, for a discussion of rate models; Faraway 2006: 61–63; R Core Team 2016, on the programming code for the R software). The general fitting of the GAMM is done by means of the R package 'mgcv' (Wood 2017).

Finally, Table 3 provides some descriptive statistics. These are based on absolute frequencies for total word counts (per utterance) and occurrences of $uh(m)$; and on relative frequencies for the four informational load indicators.

To obtain mean word counts, the number of tokens in Table 2 are divided by the number of utterances: $346,058 / 19,046 = 18.170$ per utterance in monolingual Dutch speeches; $34,968 / 1,431 = 24.351$ in interpreted Dutch; $39,239 / 1,455 = 26.968$ in the French source texts. Mean occurrences of $uh(m)$ are calculated as $10,519 / 19,046 = 0.552$ for monolingual Dutch production; $1,738 / 1,431 = 1.210$ for interpreted Dutch; and $260 / 1,455 = 0.179$ for the French source texts.

Delivery rate, measured at text level, is calculated per minute. The mean delivery rate shown in Table 3 for the original Dutch texts (165.674) is based on the mean values of 190.404 for the first monolingual Dutch text, 169.2824 for the second, 155.4812 for the third, and so on for the 240 texts in the *NED_or* corpus. The same calculation is made for the 107 texts in the *NED_in* corpus and the 107 texts in the *FRA* corpus.

Statistics for lexical density, numbers and formulaicity are calculated at utterance level. Thus, mean lexical density for monolingual Dutch production (0.390) is based on mean values of 0.667 in the first utterance of the first *NED_or* text, 0.334 in the second utterance, 0.400 in the third utterance, and so on for all 19,046 utterances in the entire *NED_or* corpus. The same calculation is made for the 1,431 utterances and the 1,455 utterances in the *NED_in* and *FRA* corpora respectively.

Lexical density, numbers and formulaicity, all indicated as relative frequencies, are normalised differently: lexical density as total content words, divided by total content words plus function words (which may be less than the total number of words per utterance); frequency of numbers as the total count divided by the total number of words; and formulaicity as the total number of 3- or 4-grams, divided by twice the total number of words minus 5.

Table 3. Descriptive statistics for total words, occurrences of *uh(m)* and the four informational load indicators

Number of words	<i>NED_or</i>	<i>NED_in</i>	<i>FRA</i>	Total
Absolute frequency	346,058	34,968	39,239	420,265
Mean	18.170	24.351	26.968	19.158
Standard deviation	15.650	17.796	19.284	16.268
Skewness	1.948	1.360	1.496	1.859
Kurtosis	6.541	2.328	3.245	5.707
Occurrences of <i>uh(m)</i>	<i>NED_or</i>	<i>NED_in</i>	<i>FRA</i>	Total
Absolute frequency	10,519	1,738	260	12,517
Mean	0.552	1.210	0.179	0.571
Standard deviation	1.197	1.752	0.696	1.231
Skewness	3.545	2.798	7.409	3.607
Kurtosis	18.947	15.331	83.997	20.626
Delivery rate	<i>NED_or</i>	<i>NED_in</i>	<i>FRA</i>	Total
Mean	165.674	155.894	176.229	165.734
Standard deviation	22.803	24.453	29.024	23.666
Skewness	−0.326	0.482	−0.043	−0.183
Kurtosis	−0.106	1.282	1.711	0.321
Lexical density	<i>NED_or</i>	<i>NED_in</i>	<i>FRA</i>	Total
Mean	0.390	0.429	0.443	0.396
Standard deviation	0.175	0.106	0.118	0.169
Skewness	0.709	0.747	1.538	0.676
Kurtosis	3.124	3.481	7.606	3.356
Proportion of numbers	<i>NED_or</i>	<i>NED_in</i>	<i>FRA</i>	Total
Mean	0.019	0.016	0.016	0.019
Standard deviation	0.082	0.040	0.040	0.078
Skewness	8.799	3.409	4.055	9.037
Kurtosis	93.318	13.821	25.483	100.828
Formulaicity	<i>NED_or</i>	<i>NED_in</i>	<i>FRA</i>	Total
Mean	0.028	0.040	0.050	0.030
Standard deviation	0.124	0.141	0.127	0.126
Skewness	6.962	6.227	6.028	6.823
Kurtosis	50.197	38.989	39.620	48.347

5. Results

5.1 Analysis 1: Comparison of input and output in interpreting

The four informational load indicators and the frequency of $uh(m)$ were compared in the source and target texts. The model is fairly good, with deviance close to its residual degrees of freedom (Tang et al. 2012:180–181): 1,301.031 and 1,308.168 respectively. Table 4 gives the effects of the fixed factors – i.e. the intercept and the nine numerical predictors. Since the cognitive load indicators are estimated by means of smoothing splines, their effects cannot be expressed as single coefficients. Table 4 therefore indicates degrees of freedom and significances for each smoothing spline, while the predictors are shown individually in Figures 1–5. We will discuss each figure in detail, focussing on the orientation and shape of the curves and the shape of the confidence bands.

Table 4 shows a significant relationship of $uh(m)$ with only three of the nine predictors: lexical density (source text) and formulaicity (source and target text). However, the six non-significant predictors (source text delivery rate, numbers and the frequency of $uh(m)$; target text delivery, lexical density and numbers) are also plotted for illustrative purposes. A refinement of the model, taking into account only the significant predictors, did not improve the fit in comparison to the comprehensive model with nine predictors.

Table 4. Significances of the predictors for interpreters' source and target texts (Analysis 1)

	Estimated coefficient	Standard Error	Z value	p-value
Intercept	−3.24121	0.07034	−46.08	<0.0001
	Effective degrees of freedom	Reference degrees of freedom	Chi-square	p-value
Source text (ST) delivery rate	2.10	2.28	1.531	0.5295
ST lexical density	1.00	1	8.022	0.0046
ST % of numbers	1.00	1	2.099	0.1474
ST formulaicity	1.00	1	7.398	0.0065
ST % of $uh(m)$	1.00	1	0.070	0.7912
Target text (TT) delivery rate	1.00	1	0.108	0.7421
TT lexical density	1.12	1.22	3.414	0.0844
TT % of numbers	1.00	1	2.441	0.1182
TT formulaicity	1.00	1	30.187	<0.0001

Figure 1 shows the effect of source and target text delivery rate: neither effect is statistically significant. Since a horizontal line can be drawn through the 95% confidence region created by the (dashed) confidence bands, the estimated effect (solid line) is not significantly different from no effect. This lack of a significant difference does not seem to result from paucity of data: though delivery rate is a second-level predictor (i.e. at text level), the 107 texts in the EPICG (see Table 2) seem to provide sufficient basis for identification of any potentially significant effects. We therefore conclude that there is no evidence here of a relationship between source or target text delivery rates and the frequency of $uh(m)$ in the interpretations.

Figure 2 shows how the frequency of $uh(m)$ in the target texts is related to source and target text lexical density. Given an average target text utterance length of 24.351 (see Table 2), the expected frequency per utterance of $uh(m)$ in the target text is about 0.609 ($= 24.351 \times 0.025$) if the source text has a lexical density of about 0.150; and 1.826 ($= 24.351 \times 0.075$) for a source text lexical density of 1. Target text lexical density is near significant only at the 10% level: the predicted target text occurrence rate of $uh(m)$ is about 0.030 per utterances with a lexical density of 0, rising to about 0.050 when lexical density is 1. Lexical density is 0 in short utterances such as *Ja* (Yes), *Die zijn dat ook* (They are also like that) or *Ik ben het daarmee eens* (I agree with that), made up solely of functional words. In absolute terms, the numbers mean that the frequency of $uh(m)$ will be about 0.731 ($= 24.351 \times 0.030$) at a lexical density of 0, increasing to about 1.218 ($= 24.351 \times 0.050$) $uh(m)$ at a lexical density of 1. Both results are consistent with the principle that increasing informational content goes hand in hand with higher cognitive load for speakers and listeners, possibly leading to processing difficulties in the specific case of interpreters (see Dillinger 1994; Tommola & Helevä 1998). One sign of these difficulties can be a growing rate of disfluencies.

Figure 3 illustrates the effects of numbers on the target text frequency of $uh(m)$, with no significant relationship: a straight horizontal line can be drawn through the 95% confidence regions. In addition, the scale on the horizontal axes of Figure 3 indicates that the maximum occurrence of numbers is about 0.35. Since the mean frequency of numbers reported above for source and target texts (Table 2) was even lower (0.016, or 1.6% in both), the lack of a significant relationship between numbers and $uh(m)$ may simply reflect a scarcity of data. Be that as it may, Figure 3 shows opposite trends for the source texts (ascending curve) and target texts (descending curve). In our earlier study (Plevoets & Defrancq 2016), we hypothesised that *omission* might be a confounding variable: while source texts containing more numbers are obviously harder for an interpreter, s/he might choose simply not to convey some of these items, which could afford a tentative explanation for the opposite trends in the source and target texts.

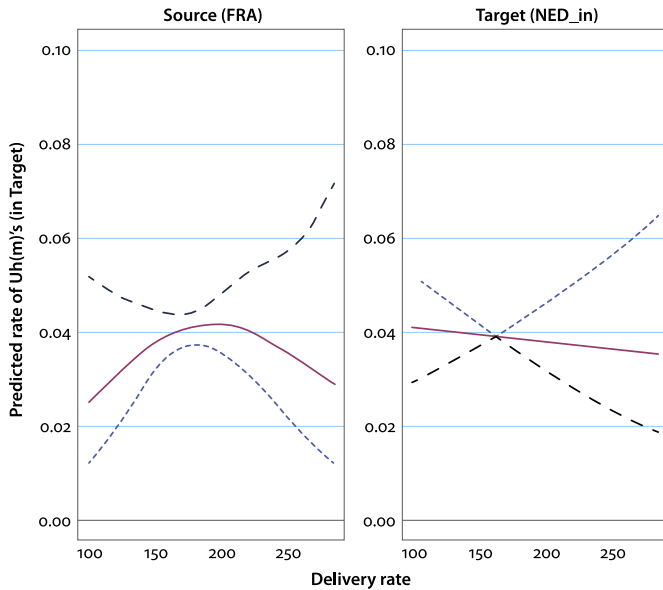


Figure 1. The effect of source and target text delivery rate on target text frequency of $uh(m)$

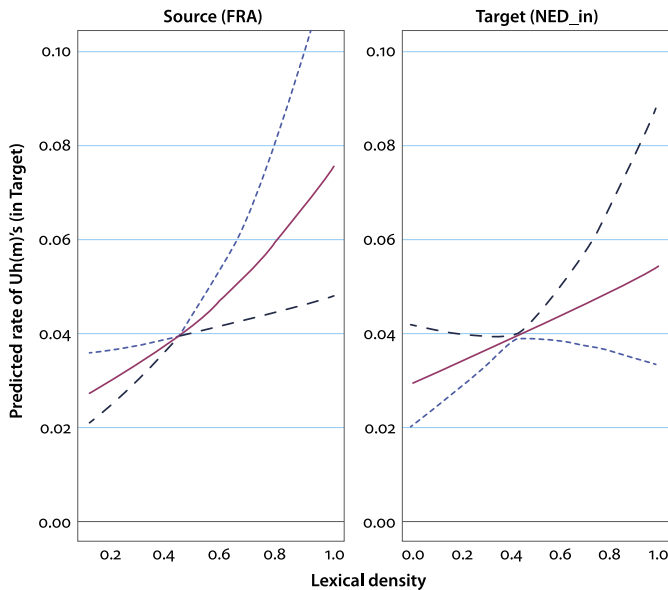


Figure 2. The effect of source and target text lexical density on target text frequency of $uh(m)$

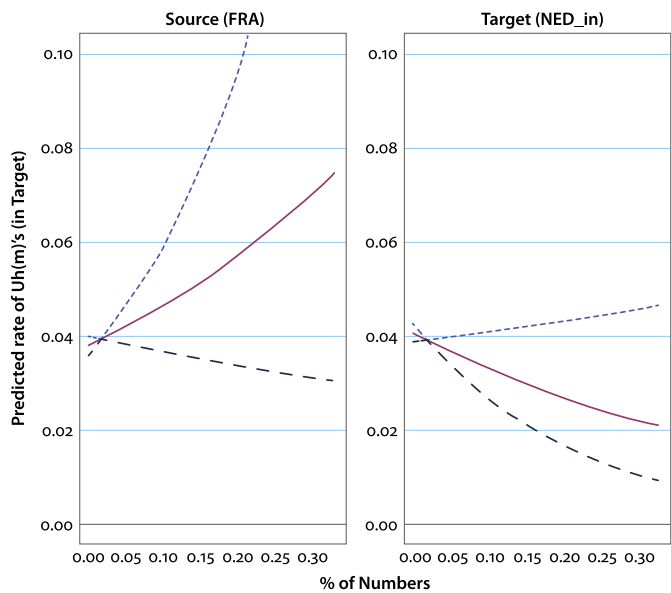


Figure 3. The effect of source and target text numbers on target text frequency of $uh(m)$

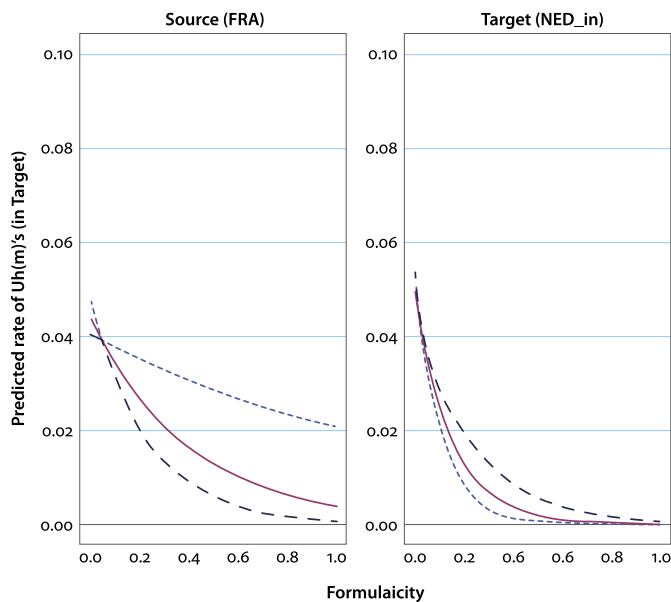


Figure 4. The effect of source and target text formulaicity on target text frequency of $uh(m)$

Figure 4 illustrates the effects of source and target text formulaicity on the target text frequency of $uh(m)$. In both cases, there is a statistically significant decrease in occurrence of filled pauses (this being more marked in the target texts). Predicted approximate source text frequency of $uh(m)$ decreases from 0.042 with a formulaicity of 0 to 0.003 with a formulaicity of 1. The corresponding values in target texts are about 0.050 and 0 respectively. This reinforces the observation (see Methodology, above) that more formulaic texts are cognitively easier to process and to produce, leading to a lower rate of disfluencies.

Finally, Figure 5 shows the relationship between the frequency of $uh(m)$ in the source and target texts. This is not significant, with a consistent level of about 4%. With respect to the discussion between Goldman-Eisler (1967) and Gerver (1975) about the tactical advantage the interpreter can obtain by making production coincide as far as possible with source text pauses, our results tend to support the latter's view that there is no evidence for this.

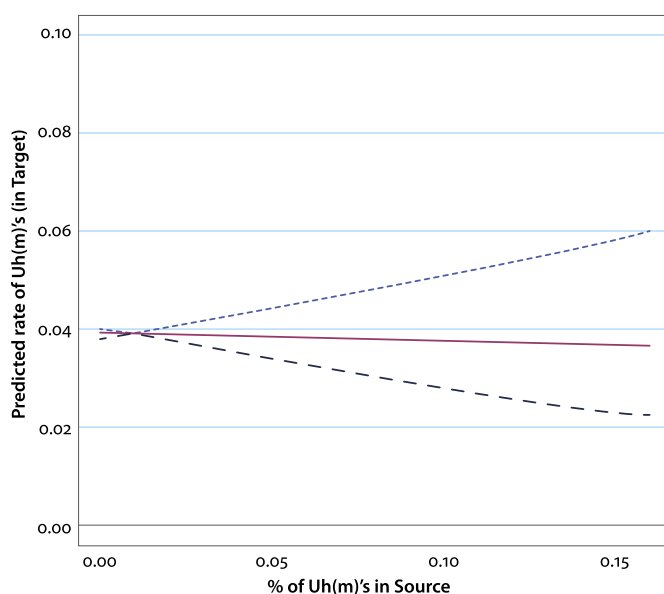


Figure 5. Relationship between frequency of $uh(m)$ in the source and target texts

5.2 Analysis 2: Comparison of output in interpreting and monolingual speech

The relationship of the four informational load indicators with the frequency of $uh(m)$ was also studied in interpretations and monolingual speech. The model was again acceptable, deviance and residual degrees of freedom being 18,108.74 and

20,078.88 respectively. Table 5 shows the estimated fixed effects, which are then visualised in Figures 6–10.

Table 5. Significance of the predictors in Dutch: interpretation vs. monolingual speech (Analysis 2)

	Estimated coefficient	Standard Error	Z value	p-value
Intercept	−3.9818	0.0821	−48.473	<0.0001
lang:NED_in	0.6991	0.1227	5.698	<0.0001
	Effective degrees of freedom	Reference degrees of freedom	Chi-square	p-value
Delivery rate × lang:NED_or	4.119	4.276	153.051	<0.0001
Delivery rate × lang:NED_in	1.001	1.001	0.274	0.6011
Lexical density × lang:NED_or	7.168	8.110	122.624	<0.0001
Lexical density × lang:NED_in	2.009	2.587	6.246	0.0738
% of Numbers × lang:NED_or	3.097	3.837	9.731	0.0408
% of Numbers × lang:NED_in	1	1	0.157	0.6916
Formulaicity × lang:NED_or	4.516	5.485	34.239	<0.0001
Formulaicity × lang:NED_in	1	1	37.975	<0.0001

Figure 6 shows the main effect of non-interpreted Dutch versus interpreted Dutch: the significantly higher frequency of *uh(m)* in the latter is consistent with the familiar idea that interpreting will tend to involve higher cognitive load than spontaneous speech (Goldman-Eisler 1967; Setton 1999).

The relationship between delivery rate and the frequency of *uh(m)* is illustrated in Figure 7. While it is not significant for interpreting, there is a significant increase for monolingual speech (although the effect becomes uncertain after about 190 words per minute). The explanation for this result in monolingual speech is almost certainly that speaking faster can involve higher cognitive load, resulting in more disfluencies. The non-significant result for interpreting is consistent with the finding in Analysis 1, suggesting that interpreters’ cognitive load does not increase with their own delivery rate. The different results for interpreting and monolingual speech could also be related to differences in training: interpreters are trained speakers and therefore plausibly have little or no difficulty managing

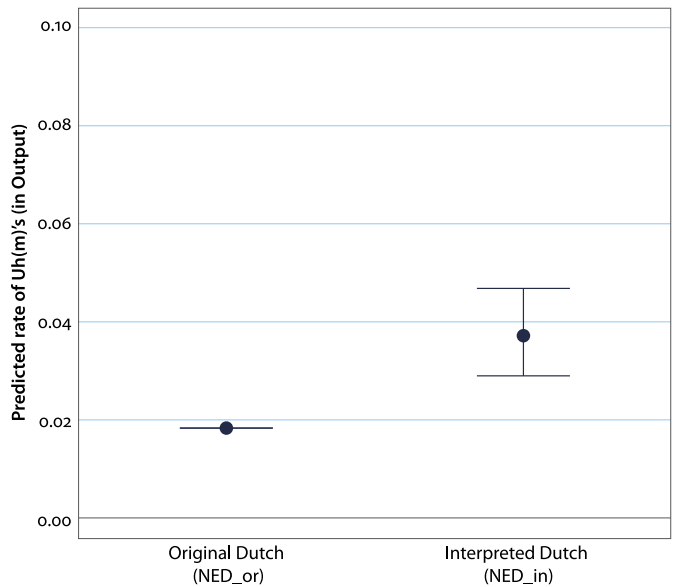


Figure 6. Main difference in frequency of $uh(m)$ between monolingual Dutch and interpreted Dutch

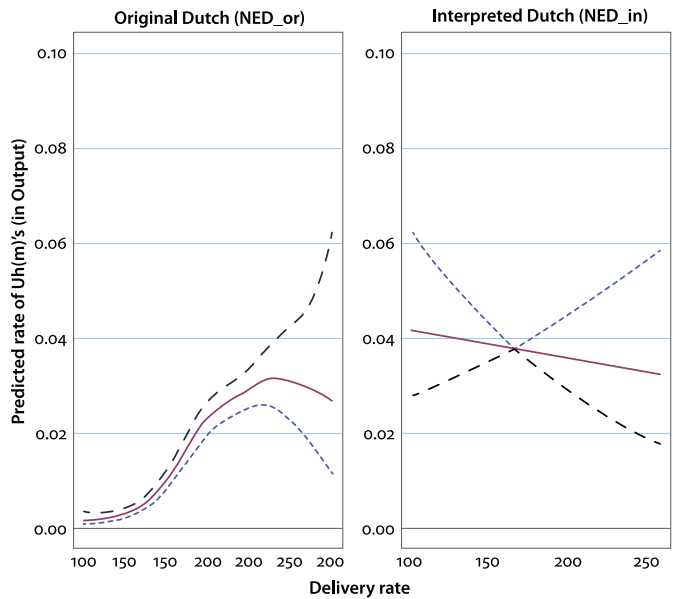


Figure 7. Effect of delivery rate on frequency of $uh(m)$ in monolingual Dutch and interpreted Dutch

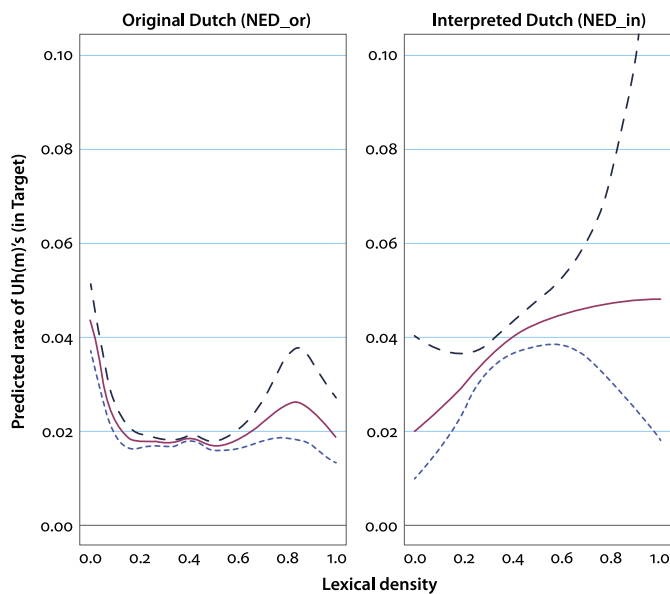


Figure 8. Effect of lexical density on frequency of *uh(m)* in monolingual Dutch and interpreted Dutch

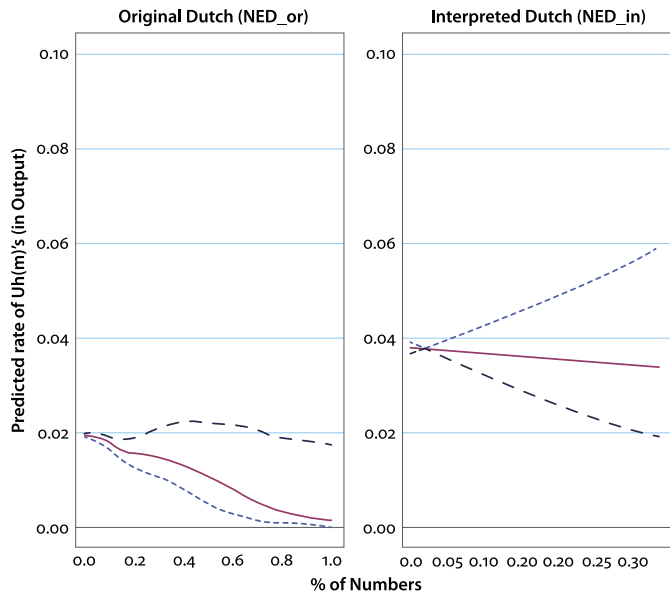


Figure 9. Effect of numbers on frequency of *uh(m)* in monolingual Dutch and interpreted Dutch

increases in their own delivery rate, while parliamentarians are not necessarily trained in public speaking.

Figure 8 shows the relationship between lexical density and the frequency of $uh(m)$. A fairly clear, near-significant linear trend emerges for interpreting; the absence of a clear pattern in monolingual speech may be explained by the highly scripted nature of the speeches concerned, as a result of which high cognitive load will not necessarily be associated with an increase in disfluencies.

In Figure 9, which examines the effects of numbers on the frequency of $uh(m)$, there is a statistically significant negative relationship in monolingual speech. This trend may, again, reflect the scripted and highly rehearsed nature of the parliamentary speeches.

Finally, Figure 10 shows the relationship between formulaicity and the frequency of $uh(m)$: there is a significant negative trend for both monolingual and interpreted speech. The undulating pattern in the monolingual speech curve may reflect parliamentary speakers' difficulties with some formulaic sequences. These results reinforce the findings for formulaicity in Analysis 1: formulaic sequences tend to make a text easier, in terms of both production and comprehension.

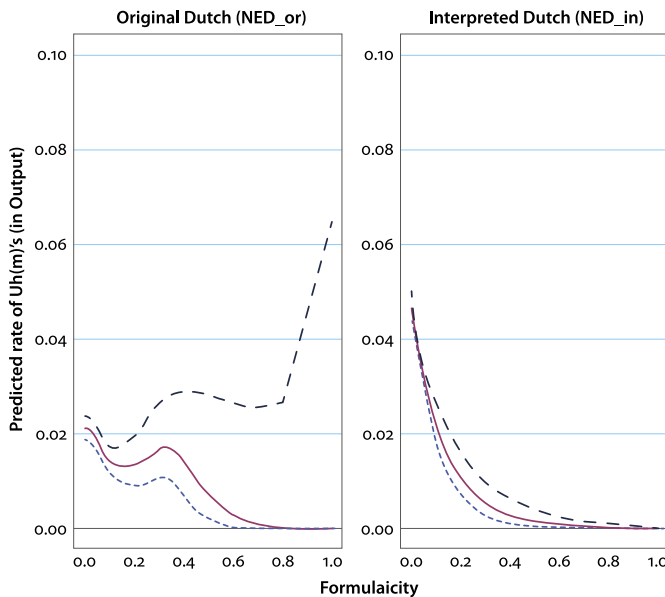


Figure 10. Effect of formulaicity on frequency of $uh(m)$ in monolingual Dutch and interpreted Dutch

6. Discussion

The results lend support to some basic findings from the literature on interpreting. Lexical density (particularly in the source text) was seen to increase the interpreter's cognitive load. On the other hand, we found no significant relationship between numbers and cognitive load, possibly because the data were insufficient. However, as we found in an earlier study (Plevoets & Defrancq 2016), we recognize that the interpreter may avoid cognitive overload by omitting numbers. Another variable affecting cognitive load is the formulaicity of the message, which has a predictably attenuating effect. An interesting finding is that this applies to both the input and the output: interpreters seem to have less processing difficulty not only with formulaic source texts, but also when their own production contains more formulaic sequences.

Various aspects of production were further analysed in the comparison of monolingual and interpreted Dutch, reinforcing the idea that interpreting involves a significantly higher cognitive load than monolingual production. That said, the relationship between delivery rate and disfluencies is significant only in monolingual speech and not in interpreting – which may arguably reflect interpreters' intensive training in speaking skills. One rather surprising finding is that neither lexical density nor the frequency of numbers shows a significant relationship with disfluencies in monolingual production, the probable explanation being the prepared and scripted nature of parliamentary speeches in this setting. Interpreting, by contrast, has more in common with impromptu speech (Shlesinger 1989; Taylor 1989): this means that interpreters will tend to experience more cognitive difficulty as the informational content of their output increases. On the other hand, formulaicity shows a negative relationship with the interpreter's (and, to a markedly lesser extent, the monolingual speaker's) cognitive load.

These results enable us to answer our three research questions (stated in Section 1). Regarding comparison of the utterance-level analysis here and the text-level breakdown in Plevoets and Defrancq (2016), the informational load indicators in the two studies do not fully coincide. The main difference is probably the positive relationship between source text delivery rate and the frequency of $uh(m)$ in the interpretations, which was significant only in the earlier study: this may be because the effect of formulaicity proves so strong in the present study that delivery rate is no longer significant. The second and third research questions (target text disfluencies in relation to source text informational load, and in comparison to monolingual speech disfluencies respectively) were addressed in Analyses 1 and 2. Lexical density appears to be the main source text predictor for the interpreter's cognitive load for listening and comprehension, while the cognitive load for target text production shows a marked negative relationship with its formu-

laicity. This indicates that input load is chiefly related to content, whereas output load is related to form. Interpreters seem to experience most problems with the processing of the source text's informational content, but benefit to a large extent from formulaic sequences in the input (predictability) and in their target text production (automatisation).

Our results are therefore consistent with theoretical models of interpreting which stress the balance between cognitive resources and demands (Gile 1997, 2009; Seeber 2011). Interpreters' cognitive load is higher when the lexical density of the source text increases, but decreases when the source text and/or target text contain more formulaic sequences. Further research is needed to determine whether interpreters deliberately use formulaic sequences as a compensatory mechanism to offset increasing cognitive load.

7. Conclusion

This paper investigated cognitive load in interpreting by using a Generalised Additive Mixed-effects Model to analyse the frequency of *uh(m)*-type disfluencies as a function of four informational load indicators: delivery rate, lexical density, numbers and formulaicity (measured as the proportion of 3- and 4-grams). The study was based on two corpora: the EPICG (containing 107 Dutch target texts and French source texts) and the non-interpreted CGNg. The analysis was twofold: a classic study of input and output in the EPICG; and a comparison of interpretations (EPICG) with monolingual speech production (CGNg), following on from Mona Baker's work in translation studies. Our results indicate a cognitive equilibrium: the cognitive load of interpreters increases with source text lexical density, but decreases when source texts are more formulaic. Interpreters also experience lower cognitive load when they can use more formulaic sequences in the target language.

These results suggest further scope for study of interpreters' strategies for cognitive load management. For example, we have already mentioned the interest of examining whether interpreters use formulaic sequences strategically to reduce cognitive load. Another point is that our data suggest a possible role of omission in the reduction of cognitive load, particularly in the case of numbers. Our hypothesis is that some source text numbers, which were not interpreted, nevertheless generated disfluencies in parts of the Dutch interpretations where no numbers occurred. Clarification of these questions will help towards a fuller account of interpreters' cognitive load and of how they manage it.

References

- Agresti, A. (2013). *Categorical data analysis*. Hoboken, NJ: John Wiley & Sons.
- Alessandrini, M. S. (1990). Translating numbers in consecutive interpretation: An experimental study. *The Interpreters' Newsletter* 3, 77–80.
- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combination. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications*. Oxford: Oxford University Press, 101–122.
- Arnold, J. E., Fagnano, M. & Tanenhaus, M. K. (2003). Disfluencies signal thee, um, new information. *Journal of Psycholinguistic Research* 3, 25–36.
<https://doi.org/10.1023/A:1021980931292>
- Arnold, J. E., Wasow, T., Losongco, A. & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76, 28–55. <https://doi.org/10.1353/lan.2000.0045>
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511801686>
- Bachy, S., Dister, A., Francard, M., Geron, G., Giroul, V., Hambye, P., Simon, A. -C. & Wilmet, R. (2007). *Conventions de transcription régissant les corpus de la banque de données VALIBEL*. [Transcription conventions of the corpora included in the VALIBEL Database]
www.uclouvain.be/cps/ucl/doc/valibel/documents/conventions_valibel_2004.PDF (accessed 1 October 2015).
- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair*. Amsterdam: John Benjamins, 233–250. <https://doi.org/10.1075/z.64.15bak>
- Barik, H. C. (1975). Simultaneous interpretation: Qualitative and linguistic data. *Language and Speech* 18, 272–297.
- Bendazzoli, C. (2017). Corpus-based interpreting studies: Past, present and future developments of a (wired) cottage industry. In C. Bendazzoli, M. Russo & B. Defrancq (Eds.), *Making way in corpus-based interpreting studies*. Singapore: Springer.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *The Longman grammar of spoken and written English*. London: Longman.
- Biber, D., Conrad, S. & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25, 371–405.
<https://doi.org/10.1093/applin/25.3.371>
- Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F. & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech* 44, 123–147. <https://doi.org/10.1177/00238309010440020101>
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.
<https://doi.org/10.1037/10037-000>
- Cecot, M. (2001). Pauses in simultaneous interpretation: A contrastive analysis of professional interpreters' performances. *The Interpreters' Newsletter* 11, 63–85.
- Chen, Y. -H. & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology* 14, 30–49.
- Chernov, G. V. (2004). *Inference and anticipation in simultaneous interpreting: A probability-prediction model*. Amsterdam: John Benjamins. <https://doi.org/10.1075/btl.57>

- Chmiel, A. & Mazur, I. (2013). Eye tracking sight translation performed by trainee interpreters. In C. Way, S. Vandepitte, R. Meylaerts & M. Bartłomiejczyk (Eds.), *Tracks and treks in translation studies*. Amsterdam: John Benjamins, 189–205.
<https://doi.org/10.1075/btl.108.10chm>
- Clark, H. H. & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition* 84, 73–111. [https://doi.org/10.1016/S0010-0277\(02\)00017-3](https://doi.org/10.1016/S0010-0277(02)00017-3)
- Conklin, K. & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and non-native speakers?. *Applied Linguistics* 29, 72–89.
<https://doi.org/10.1093/applin/amm022>
- Conklin, K. & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics* 32, 45–61. <https://doi.org/10.1017/S0267190512000074>
- Dillinger, M. (1994). Comprehension during interpreting: What do interpreters know that bilinguals don't? In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation*. Amsterdam: John Benjamins, 155–189.
<https://doi.org/10.1075/btl.3.14dil>
- Defrancq, B. (2015). Corpus-based research into the presumed effects of short EVS. *Interpreting* 17 (1), 26–45.
- Eyckmans, J. (2007). Taking SLA research to interpreter-training: Does knowledge of phrases foster fluency? In F. Boers, J. Darquennes & R. Temmerman (Eds.), *Multilingualism and applied comparative linguistics, Volume 1: Pedagogical perspectives*. Cambridge: Cambridge Scholar Publishing, 89–105.
- Faraway, J. J. (2006). *Extending the linear model with R*. Boca Raton: Chapman & Hall/CRC.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology* 32 (3), 221–233.
<https://doi.org/10.1037/h0057532>
- Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software* 8 (15), 1–27. <https://doi.org/10.18637/jss.v008.i15>
- Gerver, D. (1969). The effects of source language presentation rate on the performance of simultaneous conference interpreters. In E. Foulke (Ed.), *Proceedings of the 2nd Louisville Conference on Rate and/or Frequency Controlled Speech*. University of Louisville: Centre for Rate-Controlled Recordings, 162–184.
- Gerver, D. (1975). A psychological approach to simultaneous interpretation. *Meta* 20 (2), 119–128. <https://doi.org/10.7202/002885ar>
- Gerver, D. (1976). Empirical studies of simultaneous interpretation: A review and a model. In R. W. Brislin (Ed.), *Translation: Applications and Research*. New York: Gardner Press, 165–207.
- Gibson, T. R. (1993). *Towards a discourse theory of abstracts and abstracting*. Nottingham: University of Nottingham.
- Gile, D. (1995). *Regards sur la recherche en interprétation de conférence*. Lille: Presses Universitaires de Lille.
- Gile, D. (1997). Conference interpreting as a cognitive management problem. In J. H. Danks, G. M. Shreve, S. B. Fountain & M. McBeath (Eds.), *Cognitive processes in translation and interpreting*. Thousand Oaks/London/New Delhi: Sage Publications, 196–214.
- Gile, D. (1999). Testing the Effort Models' tightrope hypothesis in simultaneous interpreting – A contribution. *Hermes* 22, 51–79.
- Gile, D. (2008). Local cognitive load in simultaneous interpreting and its implications for empirical research. *Forum* 6, 59–77.

- Gile, D. (2009). *Basic concepts and models for interpreter and translator training. Revised edition*. Amsterdam: John Benjamins. <https://doi.org/10.1075/btl.8>
- Goldman-Eisler, F. (1967). Sequential temporal patterns and cognitive processes in speech. *Language and Speech* 10 (3), 122–132.
- Haeseryn, W., Romijn, K., Geerts, G., De Rooij, J. & Van den Toorn, M. C. (1997). *Algemene Nederlandse Spraakkunst*. [General Dutch Grammar] Groningen/Deurne: Martinus Nijhoff/Wolters Plantyn.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications. Second edition*. New York: Routledge.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27, 4–21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G. & Keenan, J. M. (1975). Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior* 14, 158–169. [https://doi.org/10.1016/S0022-5371\(75\)80065-X](https://doi.org/10.1016/S0022-5371(75)80065-X)
- Kurz, I. (2008). The impact of non-native English on students' interpreting performance. In G. Hansen, A. Chesterman & H. Gerzymisch-Arbogast (Eds.), *Efforts and models in interpreting and translation research: A tribute to Daniel Gile*. Amsterdam: John Benjamins, 179–192.
- Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition* 14, 41–104. [https://doi.org/10.1016/0010-0277\(83\)90026-4](https://doi.org/10.1016/0010-0277(83)90026-4)
- Mazza, C. (2001). Numbers in simultaneous interpretation. *The Interpreters' Newsletter* 11, 87–104.
- Mead, P. (2000). Control of pauses by trainee interpreters in their A and B languages. *The Interpreters' Newsletter* 10, 89–102.
- Moser, B. (1978). Simultaneous interpretation: A hypothetical model and its practical application. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication. Proceedings of the NATO symposium, Venice, Italy, September 26-October 1, 1977*. New York/London: Plenum Press, 353–368.
- Oostdijk, N. (2000). The Spoken Dutch Corpus: Overview and first evaluation. In M. Gravididou, G. Carayannis, S. Markantonatou, S. Piperidis & G. Stainhaouer (Eds.), *Proceedings of the Second International Conference on Language Resources and Evaluation*. Paris: ELRA, 887–894.
- Paquot, M. & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Linguistics* 32, 130–149. <https://doi.org/10.1017/S0267190512000098>
- Paradis, M. (1994). Toward a neurolinguistic theory of simultaneous translation: The framework. *International Journal of Psycholinguistics* 9 (3), 319–335.
- Pinochi, D. (2009). Simultaneous interpretation of numbers: Comparing German and English to Italian. An experimental study. *The Interpreters' Newsletter* 14, 33–57.
- Pio, S. (2003). The relation between ST delivery rate and quality in simultaneous interpretation. *The Interpreters' Newsletter* 12, 69–100.
- Plevoets, K. & Defrancq, B. (2016). The effect of informational load on disfluencies in interpreting: A corpus-based regression analysis. *Translation and Interpreting Studies* 11 (2), 202–224.
- R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org> (accessed 1 January 2017).

- Riccardi, A. (1998). Interpreting strategies and creativity. In A. Beylard-Ozeroff, J. Kralova & B. Moser-Mercer (Eds.), *Translators' strategies and creativity*. Amsterdam: John Benjamins, 171–180. <https://doi.org/10.1075/btl.27.24ric>
- Seeber, K. (2011). Cognitive load in simultaneous interpreting: Existing theories – new models. *Interpreting* 13 (2), 176–204. <https://doi.org/10.1075/intp.13.2.02see>
- Seeber, K. & Kerzel, D. (2012). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism* 16 (2), 228–242. <https://doi.org/10.1177/1367006911402982>
- Setton, R. (1999). *Simultaneous interpretation: A cognitive-pragmatic analysis*. Amsterdam: John Benjamins. <https://doi.org/10.1075/btl.28>
- Shlesinger, M. (1989). *Simultaneous interpretation as a factor in effecting shifts in the position of texts on the oral-literate continuum*. MA thesis, Tel Aviv University.
- Shlesinger, M. (1998). Corpus-based interpreting studies as an offshoot of corpus-based translation studies. *Meta* 43 (4), 486–493. <https://doi.org/10.7202/004136ar>
- Straniero Sergio, F. & Falbo, C. (Eds.) (2012). *Breaking ground in corpus-based interpreting studies*. Bern: Peter Lang. <https://doi.org/10.3726/978-3-0351-0377-9>
- Stubbs, M. (2007). An example of frequent English phraseology: Distribution, structures and functions. In R. Facchinetti (Ed.), *Corpus linguistics 25 years on*. Amsterdam: Rodopi, 89–105. https://doi.org/10.1163/9789401204347_007
- Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30, 485–496. [https://doi.org/10.1016/S0378-2166\(98\)00014-9](https://doi.org/10.1016/S0378-2166(98)00014-9)
- Tang, W., He, H. & Xin, M. T. (2012). *Applied categorical and count data analysis*. Boca Raton: Chapman & Hall/CRC.
- Taylor, C. (1989). Primary and secondary orality in teaching interpreting technique. In J. M. Dodds (Ed.), *Aspects of English: Miscellaneous papers for English teachers and specialists*. Udine: Campanotto Editore, 93–102.
- Tissi, B. (2000). Silent pauses and disfluencies in simultaneous interpretation: A descriptive analysis. *The Interpreters' Newsletter* 10, 103–127.
- Tommola, J. & Helevä, M. (1998). Language direction and source text complexity: Effects on trainee performance in simultaneous interpreting. In L. Bowker, M. Cronin, D. Kenny & J. Pearson (Eds.), *Unity in diversity? Current trends in translation studies*. Manchester: St. Jerome Publishing, 177–186.
- Tremblay, A. & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication*. London/New York: Continuum, 151–173.
- Tremblay, A., Derwing, B., Libben, G. & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning* 61, 569–613. <https://doi.org/10.1111/j.1467-9922.2010.00622.x>
- Underwood, G., Schmitt, N. & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences*. Amsterdam: John Benjamins, 153–172. <https://doi.org/10.1075/llt.9.09und>
- Van de Kauter, M., Coorman, G., Lefever, E., Desmet, B., Macken, L. & Hoste, V. (2013). LeTs Preprocess: The multilingual LT3 linguistic preprocessing toolkit. *Computational Linguistics in the Netherlands Journal* 3, 103–120.

- Van Rietvelde, S., Eyckmans, J. & Bauwens, D. (2010). As time goes by: Phraseological competence and linguistic anticipation in the interpreting performance. *Artesis VT Working Papers in Translation Studies*. Antwerp: Artesis.
- Voor, J.B. & Miller, J.M. (1965). The effect of practice on the comprehension of worded speech. *Speech Monographs* 32, 452–455. <https://doi.org/10.1080/03637756509375469>
- Watanabe, M., Hirose, K., Den, Y. & Minematsu, N. (2008). Filled pauses as cues to the complexity of up-coming phrases for native and non-native listeners. *Speech Communication* 50, 81–94. <https://doi.org/10.1016/j.specom.2007.06.002>
- Welford, A. T. (1952). The 'psychological refractory period' and the timing of high speed performance – a review and a theory. *British Journal of Psychology* 43, 2–19.
- SDL Trados WinAlign (2014). *SDL Trados WinAlign Tutorial*.
<http://www.translationzone.com/resources/downloads/winalign-tutorial.html> (accessed 1 October 2015).
- Wood, S. N. (2017). *Generalized additive models: An introduction with R*. Boca Raton: Chapman & Hall/CRC.

Authors' addresses

Koen Plevoets
Department of Translation
Ghent University
Groot-Brittanniëlaan 45
B-9000 Ghent
Belgium
koen.plevoets@ugent.be

Bart Defrancq
Interpreting and Communication
Ghent University
Groot-Brittanniëlaan 45
B-9000 Ghent
Belgium
koen.plevoets@lstat.kuleuven.be

Biographical notes

Koen Plevoets is a coordinator of FLAMES (Flanders Training Network for Methodology and Statistics) at KU Leuven. He holds Master in Artificial Intelligence and Master of Statistics degrees, and he defended his Ph.D. in 2008 on the socio-stylistic variation of morphological features in colloquial Belgian Dutch. From 2010 until 2016, he worked on several projects in corpus-based translation and interpreting studies. His research interests are interpreters' cognitive load and disfluencies. He is the author of the R packages *corregp* and *svs*.

Bart Defrancq is an Associate Professor at Ghent University, where he was awarded his Ph.D. in Linguistics in 2002. He teaches interpretation and legal translation. His research is corpus-based and focuses on topics ranging from contrastive linguistics to translation and interpreting studies, areas in which he has published widely. He is a member of the EQTIS research team at Ghent University and the editor of several special issues of *Languages in Contrast*, *Belgian Journal of Linguistics*, and *Across Languages and Cultures*.