The effect of cognitive load on linguistic fluency, lexical production, and emotional polarity: An exploratory study on self-directed French

Mathilde Hutin, Frédéric Tomas

Hutin, Mathilde & Tomas, Frédéric. 2025. The effect of cognitive load on linguistic fluency, lexical production, and emotional polarity: An exploratory study on self-directed French. GCLA 2025; 13: 1–28. De Gruyter Mouton. 10.1515/gcla-2025-0007

The effect of cognitive load on linguistic fluency, lexical production, and emotional polarity: An exploratory study on self-directed French

Mathilde Hutin[1,2] & Frédéric Tomas[3]

[1] Université de Lorraine, CNRS, ATILF, F-54000 Nancy, France
[2] Université catholique de Louvain, F.R.S.-FNRS, Institut Langage & Communication, Louvain-la-Neuve, Belgium
[3] Tilburg center for Cognition and Communication, Department of Communication and Cognition, Tilburg University, Netherlands

mathilde.hutin@cnrs.fr; f.j.y.tomas@tilburguniversity.edu

Abstract. Cognitive load refers to the mental resources available in working memory, which can vary due to, e.g., the inherent complexity of a task. Studies have investigated the effect of a higher cognitive demand on speech, either on fluency, lexical use or linguistic display of emotions. Fewer have investigated all these parameters jointly, and most focus on English. Here we manipulate working memory load with an image-depiction task with or without access to the image and investigate whether cognitive load impacts speech production in French, i.e., correlates with both fluency metrics (number of word tokens and lemmas, lemma-token ratio, counts and rates of filled pauses, interruptions and repetitions) and/or with lexical metrics (use of words from a given lexical field or with a given connotation). Our results show that, compared to their peers in the low cognitive load condition (with access to the image), speakers under high cognitive load (who describe the image from memory) indeed tend to use more disfluencies but show little differences in the use of specific vocabulary or in emotional words. Although the absence of lexical differences may be due to inherent language- or task-specific differences, the confirmation that an increase in cognitive load implies an increase in disfluency indicates that at least this one parameter is a cross-linguistic indicator of speech production under high cognitive load.

Keywords: Fluency, lexical access, emotions, cognitive load, Belgian French

1. Introduction: Fluency and vocabulary usage as cognitive pragmatic phenomena

Language is often said to be a window towards the way people think. The way one speaks can reveal their state of mind, and therefore influence the way the listener infers pragmatic elements from the message. One of the ways researchers have tried to understand this overlap between linguistics and cognition is through the study of disfluencies, and in particular the role of filled pauses such as *um* and *uh* (e.g., Christenfeld, 1994; Laserna et al., 2014, Schneider 2014), while other studies focus on lexical measurements, such as word meaning and connotation (Khawaja et al., 2014).

Filled pauses have been argued to communicate a metacognitive state of uncertainty about one's discourse, and answers to general questions preceded by *uh* or *um* are perceived as less likely to be correct (Brennan & Williams, 1995). Whether this signal is sent out willingly by the speaker is still debated, as cases have been made for both conscious and unconscious reasons for the existence of filled pauses (Christenfeld, 1994; Fox Tree, 2001, 2007; Hutin et al. 2024). Independently of one's willingness to use filled pauses, filled pauses as well as other disfluencies, such as self-corrections, false starts or self-interruptions, have been associated with an increase in processing complex thoughts such as navigating an icon on a screen while answering questions orally (Müller et al., 2001). The quantity of speech uttered by participants has also been shown to correlate with an increase in task complexity (Khawaja et al., 2014). However, other indices of (dis)fluency such as the number of word-types or lemmas, as well as lemma-token ratios, have been less investigated.

Regarding lexical indices *per se*, several global syntactic and semantic characteristics have been observed to correlate with the decrease of available resources for speech (Berthold, 1998). Studies over at least the last two decades have investigated an array of characteristics possibly correlating with a simultaneous higher cognitive demand across various types of tasks. It has been shown that the use of first-person plurals (Sexton & Helmreich, 2000) and that of plural pronouns in general (Khawaja et al., 2014), the lexical fields of human cognitive processes and of perception (Khawaja et al., 2014), agreement and disagreement vocabulary as well as inclusive words such as “along”, “with”, “everyone”, etc. (Khawaja et al., 2014), morpho-phonological complexity, notably length of words (Lennon & Burdick, 2004; Khawaja et al., 2014) and syntactic complexity, such as length of sentences (Lennon & Burdick, 2004; Khawaja et al. 2014), indeed all correlate with an increase in complex thought processing.

Regarding the emotional polarity of speakers, an increased use of negative emotion words and expletives, a decrease in positive emotion words, and a decrease in emotion words altogether (Khawaja et al., 2014) have been shown to correlate with an increase in processing complex thoughts.

In the established terminology of cognitive science, this increase in processing complex thoughts is often referred to as an increase in *cognitive load*. Cognitive load refers to the mental resources a person can invest in a specific task (Chandler & Sweller, 1991; Paas et al., 2003; Sweller, 1988). This cognitive load can vary due to several factors, such as the inherent complexity of the task (for instance, needing to remember more information than one's working memory can handle or needing to complete the task with a time constraint) or the way the task is presented or carried out (for example, having to type on a keyboard with keys arranged differently from the familiar layout; Paas et al., 2010).

In the present study, cognitive load in speech is explored through the working memory model developed by Baddeley (2012), a model allowing to express how cognitive load can affect the production of linguistic elements. This model proposes four main components of working memory. The central executive oversees attention control and coordinates the other subsystems, i.e., the phonological loop, the visuospatial sketchpad, and the episodic buffer. The phonological loop stores and rehearses verbal and acoustic information. The visuospatial sketchpad handles visual imagery and spatial information. Finally, the most recently added component, the episodic buffer (Baddeley 2000), integrates representations from these systems with long-term memory into unitary episodic representations. The episodic buffer is controlled by the central executive, has multimodal storage capacity, and uses attentional mechanisms to actively maintain integrated memory episodes. According to Baddeley's model, the phonological loop is responsible for temporary storage of acoustic and speech-based information (Baddeley, 2000). The phonological loop thus facilitates describing tasks and recalling information in working memory (Baddeley, 2000). In other words, in the case of descriptions of stimuli that are accessible immediately to the senses, such as the depiction of an image that the speaker can see, the phonological loop is activated and borrows very few items from semantic long-term memory besides the name of the forms (i.e., triangle, square) and relative geographic localization (e.g., “above this”, “on the left of that”...). On the contrary, in case someone must describe something from memory, their phonological loop becomes the mediator between their verbal description and long-term memory. Recalling information from long-term memory can thus be considered as cognitively more demanding as it involves (i) the recall in long-term memory of information that is not readily available, and (ii) long-term memory can involve inaccuracies that would be avoided in the description from perception (Tan & Jiang, 2020). For these reasons, one can expect that the phonological loop is given less resources and could therefore produce less fluent information about what the speaker perceives, requiring a higher level of pragmatic investment and, therefore, more words. Given this theoretical framework, cognitive load in this study has been manipulated by giving the speakers access to the image they should depict (low cognitive load) or not (high cognitive load).

The aim of the present study is twofold. On the one hand, we investigate whether cognitive load indeed correlates with fluency metrics, such as the number of word-tokens and lemmas, as well as the lemma-token ratio, per recording, counts and rates of interruptions and

repetitions as well as the counts and rates of filled pauses. On the other hand, we explore the effect of cognitive load on lexical metrics, such as the use of words from a given lexical field or with a given connotation. The novelty of the study resides in both the fact that we explore fluency, lexical and emotional indices of cognitive load in one and the same study, and the fact that we investigate a language that was less investigated in this regard, i.e., Belgian French. In the following, we first present our data and methodology in Section 2, then our results in Section 3: Subsection 3.1 focuses on the fluency metrics, while Subsections 3.2 and 3.3 focus on the lexical metrics. Section 4 concludes and discusses the results.

2. Data and Methodology

This study exploits a subset of the TAngram COrpus, also known as TACO (Broisson, 2019; Broisson & Degand, 2022), which was developed in 2018 at Louvain University (Belgium) and is now freely available on demand. It consists of descriptions of a tangram puzzle first to a partner (allocentric condition) and then to oneself (egocentric condition), either under low or high cognitive load. Since the high cognitive load condition was not achieved the same way in the allocentric and egocentric conditions, the two allo- and egocentric conditions are not comparable and we use here only the data from the second task, i.e., the egocentric condition, henceforth EGO.

We chose the TACO corpus because it is, to the best of our knowledge, the only one allowing a controlled, in-depth comparison of low vs. high cognitive load in French self-directed speech. Knutsen et al. (2020) explore the speech of only 5 participants retelling movie excerpts for which cognitive load varied on multiple dimensions (silent vs. talking movie, action or not, and duration of the except). Another French tangram corpus by Danino et al. (2020) comprises 18 pairs of speakers in either a low or a high cognitive load condition, the latter being achieved through the imposition of a 10-minute time constraint. In addition to comprising dialogues instead of self-directed speech and achieving cognitive load without manipulating memory load, this data was created for psychological studies and would have required extra processing for linguistic analyses.

2.1. Experimental material

A tangram is a dissection puzzle consisting of seven flat polygons which are supposed to be put together to replicate a pattern based solely on its silhouette.

In TACO, the experimental items consist of a set of seven printed colored geometrical shapes commonly used in tangram puzzles – a small pink triangle, a small yellow triangle, a middle-sized yellow triangle, a big light-green triangle, a big blue triangle, a small purple square and a red parallelogram. The experimental items also comprise two puzzle silhouettes, one of a cat and one of a butterfly, given to participants alternatively. The two test items and their solutions are given in Figures 1 and 2.

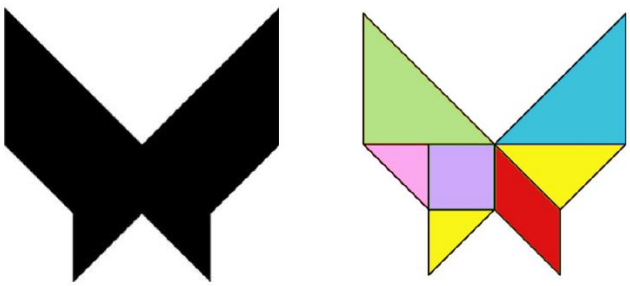

Figure 1. Butterfly tangram puzzle (silhouette on the left, colored solution on the right).

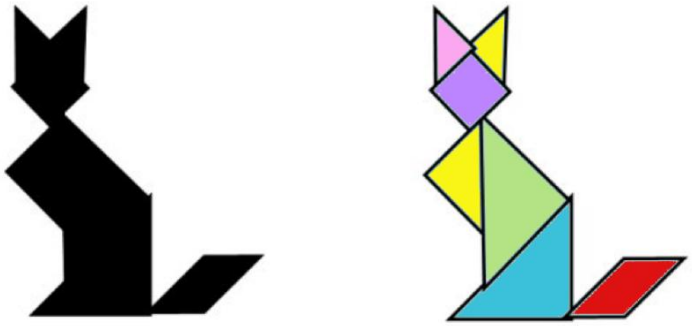

Figure 2. Cat tangram puzzle (silhouette on the left, colored solution on the right).

Participants were also provided with a written copy of instructions given by the experimenter to keep with them at all times, a bell which they could ring to call the experimenter into the room if a (substantial) problem arose, as well as a timer.

2.2. Experimental design

After completing the first task in pairs, the participants were asked to complete the second task, i.e., the egocentric condition, which yielded the presently used corpus, EGO. They were instructed to retell the instructions to the puzzle they had just solved (or attempted to solve) during the allocentric condition. They had to use only verbal cues, in less than ten minutes, in order for their future self to be able to reassemble the puzzle three months later.

To be in adequation with the definition of cognitive load assumed in the introduction (Baddeley, 2000), cognitive load was manipulated, following Roßnagel (2000), by presenting participants in the low cognitive load condition with the assembled puzzle during the task, while asking participants in the high cognitive load condition to recall the puzzle from memory. In the latter case, participants were given an additional five minutes between the instructions from the experimenter and the beginning of the recording, in order to remember how to assemble the puzzle.

The data was collected in a quiet room using the internal microphone of a Nikon D5000 camera, as well as that of an iPhone X as backup.

2.3. Participants

The participants were 12 men and 12 women (n=24)[1] recruited among the acquaintances of the experimenter. They are native speakers of Belgian French from the Hainaut province in Belgium. However, the data of one male speaker was not taken into account in the present study, as he was distracted during the experiment.

The remaining 23 participants' age ranges from 19 to 55 years old (μ=29.3, σ=12.9, Md=24), although younger speakers between the ages of 19 and 26 represent the majority of the sample (n=16, 69.56%).

Besides the age and gender information, the data is anonymized. The participants each received an ID following the template [the participant's initials + gender (coded as F for female, M for male) + age of the speaker], e.g., ASF24 is a 24-year-old woman whose initials are AS.

[1] As we are using an already-existing corpus, we are limited by the content that it offers, which fits the resource constraints justification from Lakens (2022). Results will be interpreted carefully, in line with the principles for cautious statistical inferences (Barnes & Lewin, 2005).

2.4. Data

The data in EGO thus contains 23 audio files for a total duration of approximately 1 hour. The gender, age and duration of speech are given for each participant in each condition in Table 1.

Cognitive load	Speaker ID	Gender	Age	Duration of speech
Low	ANM55	M	55	03:36
	CAF39	F	39	01:21
	CHF26	F	26	04:23
	CLF21	F	21	02:05
	FAM38	M	38	01:06
	HEF25	F	25	03:18
	LOM19	M	19	01:54
	MAF38	F	38	00:52
	NAF52	F	52	03:33
	PAF25	F	25	03:06
	VEF55	F	55	02:19
	Subtotals			**27:33**
	Means		**35.70**	**02:30**
High	ASF24	F	24	07:47
	BRM20	M	20	01:10
	CAF26	F	26	02:32
	CHF52	F	52	03:32
	COM20	M	20	01:41
	JLM19	M	19	01:45
	JSM19	M	19	01:09
	LUM19	M	19	01:07
	MSM19	M	19	01:30
	MTM19	M	19	01:17
	ROM21	M	21	01:17
	SAF24	F	24	05:54
	Subtotals			**30:41**
	Means		**23.50**	**02:33**
TOTAL				**58:14**
MEANS			**29.35**	**02:25**

Table 1. Speaker ID, gender, age and duration of the recording for each participant in the low (top) and high (bottom) cognitive load condition.

The low cognitive load condition comprises 8 female and 3 male participants (n=11), aged between 19 and 55 (μ=35.73, σ=13.60), who spoke between 52 seconds and 4 minutes 23 seconds, for a total of 27 minutes 33 seconds, i.e., a mean of 2 minutes 30 seconds per speaker (σ=1:09).

The high cognitive load condition comprises 4 female and 8 male participants (n=12), aged between 19 and 52 (μ=23.5, σ=9.30), who spoke between 1 minute 7 seconds and 7 minutes 47 seconds, for a total of 30 minutes 41 seconds, i.e., a mean of 2 minutes 33 seconds per speaker (σ=2:09).

The fact that speakers in both conditions speak for comparable durations may be due to the fact that they were all limited to a 10-minute timeframe, but it may also indicate that the two conditions were well-balanced enough to yield similar behavior regarding duration of speech, thus enhancing the quality of the comparison for the other factors (fluency and lexical use). Note however, as a precursor of forthcoming analyses, that the speakers in the high cognitive load condition display more inter-individual variation than in the low cognitive load condition.

2.5. Data processing and analysis

Each audio file was transcribed orthographically in standard French (retaining standard contracted forms only) in Praat (6.1) (Boersma & Weenink, 2019) by Broisson (2019).

Any instance of a person's name is transcribed either as "NAME" or the appropriate participant ID. Filled pauses are transcribed as "euh", "euhm", "hum", "m/", "mm/", "h/", "pf/", "hm", "mh", "tch/", "ts/", or "FP" if their articulation was ambiguous. Truncations are marked by a forward slash "/". Laughter is marked with "(LAUGH)" and any unclear text is marked with "UNCLEAR". Given that several inconsistencies were identified, such as "euhm" transcribed as "euh mh", we homogenized the transcriptions manually.

For the purpose of the first sub-study on fluency, we first extract several pieces of information using Python (3.10.12) (Van Rossum & Drake 2009):

- number of word-tokens: the count of words uttered by the participant
- number of lemmas: the count of different word lemmas used by the participant
- frequency of lemmas: (the number of lemmas / the number of tokens) * 100
- number of filled pauses: the count of occurrences of "euh", "euhm", "hum", "m/", "mm/", "h/", "pf/", "hm", "tch/", "ts/", or "FP"
- frequency of filled pauses: (the number of filled pauses / the number of tokens) * 100
- number of interruptions: the count of interrupted or corrected words (marked with "/")
- frequency of interruptions: (the number of interruptions / the number of tokens) * 100
- number of repetitions: the count of word-tokens identical to the preceding word-token
- frequency of repetitions: (the number of repetitions / the number of tokens) * 100
- overall number of disfluencies: the total count of filled pauses, interruptions and repetitions
- frequency of disfluencies: (the number of disfluencies / the number of tokens) * 100

The transcription text files as well as the Python script used to extract the information are available at the following repository: https://osf.io/amjuy/.

For the second sub-study on lexical information, we analyze the transcribed data by relying on the LIWC-22 software (Piolat et al., 2011). The dictionary used for this purpose was the 2007 French dictionary developed by Piolat et al. (2011). We extract lexical information on both word meanings, i.e., lexical fields, and word connotations that have been shown to correlate with cognitive load in past research.

The word meanings that we test thus comprise:

- the use of first-person plurals (Sexton & Helmreich, 2000) and plural pronouns (Khawaja et al., 2014),
- the lexical field of human cognitive processes (Khawaja et al., 2014),
- the lexical field of perception (Khawaja et al., 2014), in particular that of sight,
- and agreement and disagreement vocabulary as well as inclusive words such as "along", "with", "everyone", etc. (Khawaja et al., 2014).

Given that the data comes from oral utterances, in which sentences are not an applicable concept, and given that the data is constituted of only egocentric, i.e., self-directed, monologual speech, that prevents from using another kind of unit such as speech turn, we do not investigate the effect of cognitive load on length of sentences as markers of syntactic

complexity (Lennon & Burdick, 2004; Khawaja et al., 2014). However, given the nature of the task, we also add the following lexical metrics:

- the lexical field of "relativity", i.e., space-, number- and movement-related terms,
- the lexical field of tentativeness,
- the lexical field of hesitancy or uncertainty,
- swear words.

In addition to swear words, we also test the use of negative emotion words, following Khawaja et al. (2014).

For the sake of clarity, the exact hypotheses and predictions for each variable will be exposed in the appropriate section before the corresponding results.

The final dataset is analyzed using R (4.2.1) (R Core Team, 2021) and jamovi (The jamovi project, 2023). On jamovi, we conduct pairwise comparisons. If no violations of homogeneity of variance or normality are observed, independent-samples t-tests are used. In the case of violations of the homogeneity of variance assumption only, we report Welch's t-test. In all other cases, we report the Mann-Whitney U statistic.

3. Results

3.1. Fluency metrics

The counts for (dis)fluency metrics are given in Table 2.

Speaker ID	Condition	Gender	Age	Count of word tokens	Count of lemmas	Count of filled pauses	Count of interruptions	Count of repetitions	Total count of disfluencies
CAF39	low	female	39	214	59	0	1	0	1
CLF21	low	female	21	313	91	7	1	0	8
MAF38	low	female	38	170	58	6	1	0	7
LOM29	low	male	29	171	39	4	4	0	8
FAM38	low	male	38	154	51	2	1	6	9
HEF25	low	female	25	540	116	9	7	0	16
ANM55	low	male	55	646	120	12	2	0	14
VEF55	low	female	55	413	96	14	4	0	18
PAF25	low	female	25	503	113	24	4	3	31
CHF26	low	female	26	651	108	29	2	2	33
NAF52	low	female	52	596	119	24	8	6	38
Subtotal				**4,371**	**970**	**131**	**35**	**17**	**183**
COM19	high	male	19	186	60	7	0	0	7
JSM20	high	male	20	135	53	1	1	1	3
LUM19	high	male	19	175	52	2	5	0	7
MTM19	high	male	19	208	67	6	3	1	10
JLM19	high	male	19	281	67	10	4	0	14
CAF26	high	female	26	394	88	14	2	0	16
BRM19	high	male	19	196	68	12	5	0	17
ROM10	high	male	20	234	78	12	7	0	19

CHF52	high	female	52	433	92	16	4	5	25
MAM19	high	male	19	283	68	17	7	0	24
SAF24	high	female	24	907	122	79	3	5	87
ASF24	high	female	24	1,266	155	106	14	17	137
Subtotal				**4,698**	**970**	**282**	**55**	**29**	**366**
TOTAL				**9,069**	**1,940**	**413**	**90**	**46**	**549**

Table 2. Counts of word-tokens, lemmas, filled pauses, interruptions, repetitions and total number of disfluencies for each participant in the low (top) and high (bottom) cognitive load conditions.

3.1.1. Number of word-tokens

Regarding the number of words under high vs. low cognitive load, past literature provides contradictory predictions. On the one hand, some studies (Kleinman & Serfaty, 1989) found a decrease in speech quantity under high workload, while on the other, many studies have shown that under complex and high mental load tasks, team members communicate more with each other to provide more information or explanation in order to handle the increased task load (Foushee & Helmreich, 1988; Oser et al., 1991; Katz et al., 1998; Janssen et al., 2010; Khawaja et al., 2014).

These latter results are in line with Baddeley (2000)'s model, where depicting a picture from memory means that the phonological loop is given less sensory resources and could therefore produce less fluent information about what the speaker perceives, requiring a higher level of pragmatic investment and, therefore, more words. However, these studies focus on dialogical data where cognitive load is manipulated in various ways: Is self-directed speech similarly influenced by image- vs. memory-based depiction tasks? Do we also observe more words in the high rather than in the low cognitive load condition?

In EGO, speakers produce the lowest (n=135) and highest (n=1,266) numbers of words in the high cognitive load, for a total of 4,698 words, i.e., a mean of 391.50 words per speaker and a standard deviation of 344.80. Speakers produce much more homogenous word counts in the low cognitive load condition, with productions ranging from 154 to 651 words, for a total of 4,371 words, i.e., a mean of 397.36 words per speaker but a standard deviation of only 200.06.

Since word count is not normally distributed, the non-parametric Mann-Whitney U test is used to assess whether cognitive load affects the number of words uttered by participants. Results show little difference in median number of words between the high cognitive load group (Md=215) and the low cognitive load group (Md=355), U=55.50, p=.538, r=.16. This indicates that the high vs. low cognitive load does not significantly impact the amount of tokens that will be produced, contrary to findings in previous studies. However, our counts seem to indicate that higher cognitive load will favor diversity in the participants' speech quantity, as if speakers were more unequal in their way to handle cognitive load than in their way to deal with a low cognitive load task.

3.1.2. Number of lemmas and lemma-token ratio

Regarding the effect of cognitive load on counts and rates of lemmas, to the best of our knowledge, no specific studies have been conducted. Roßnagel (1995) however shows that, when speaking under higher cognitive load, i.e., while explaining how to assemble a machine without access to the instructions, speakers tend to only name technical terms while they tend to both name and describe technical terms when not under pressure, i.e., when given the assembly instructions, which in our case would imply more diversified vocabulary, i.e., more lemmas in the low-cognitive load condition. Moreover, the working memory model (Baddeley, 2012) seems to indicate that cognitive load will impact negatively lexical density (or vocabulary

richness), i.e., among others, diversity of lemmas, since it will monopolize the cognitive resources necessary to search and select lexemes from the speakers' mental lexicon. Khawaja et al. (2014) confirm this hypothesis, and we therefore expect, in line with their results as well as Roßnagel (1995)'s, to find less vocabulary diversity in the high than in the low cognitive load condition.

However, we could also expect results to go against what was found by Khawaja et al. (2014), while still being in line with Baddeley (2012). The reason behind this is to be found in the different cognitive-load inducing tasks: In Khawaja et al. (2014), the increase in cognitive load concerns the size and number of fires to manage in an emergency management scenario, with the idea that the more fires, the harder the collaboration, while in our study, the increase in cognitive load is found in the absence of a stimulus to describe, requiring higher pragmatic levels of description to the self, and therefore, potentially more words.

In EGO, speakers in the low cognitive load condition produce overall less different words, ranging from 39 to 120, while speakers in the high cognitive load condition produce between 42 and 155 different words: Speakers exposed to a high cognitive load had a lower mean per speaker (μ=80.83 vs. 88.18 in the low cognitive load condition) and a similar standard deviation (σ=30.35 vs. 30.62). An independent samples t-test is conducted to determine whether cognitive load affects the number of lemmas. Results show little difference in number of lemmas between the high cognitive load group (μ=80.83, σ=30.35) and the low cognitive load group (μ=88.18, σ=30.62), $t(21)$=-0.58, p=.570, Cohen's d=-0.24. This indicates that there is no significant difference between the two conditions with regards to the number of different lemmas used by speakers.

Moreover, the ratio of different lemmas as a function of the amount of word-tokens is overall similar in the high cognitive load condition (12.24% to 39.26%) and in the low cognitive load condition (16.59% to 34.12%), but with a higher mean (μ=26.55%) and a higher standard deviation (σ=8.44%) than in the low cognitive load condition (μ=24.45%, σ=5.77%). An independent samples t-test is also conducted to determine whether cognitive load affects the lemma-token ratio: Results show no difference in lemma-token ratios between the high cognitive load group (μ=26.55, σ=8.44) and the low cognitive load group (μ=24.45, σ=5.77), t(22)=0.78, p=.49, 95% CI [-4.41, 8.98], Cohen's d=0.29, thus indicating that speakers in the high cognitive load condition seem to have a slight tendency (not a statistically significant one, though) to use more diversified vocabulary than the ones in the low cognitive load condition, which would be in line with the fact that the phonological loop, while facilitating describing tasks and recalling information in working memory, also impedes access to long-term memory, and therefore to the mental lexicon.

3.1.3. Counts and rates of filled pauses

Regarding filled pauses, we would expect more of them in high cognitive load speech, as they have been shown to be used unconsciously for discursive self-management (Hutin et al., 2024), may be used to indicate uncertainty (Brennan & Williams, 1995) and have indeed been shown to correlate with higher task complexity (Christenfeld, 1994; Müller et al., 2001).

In EGO, speakers submitted to high cognitive load produced 1 to 106 filled pauses, while speakers in the low cognitive load condition produced between 0 and 29. The mean number of filled pauses is divided in two between the two conditions (HCL=23.5, LCL=11.91) and the standard deviation in three (HCL=33.12, LCL=9.79). The ratio of the number of filled pauses as a function of the amount of word-tokens goes from 0.00% to 4.77% in the low cognitive load condition and from 0.74% to 8.71% in the high cognitive load condition, with again a higher mean and a higher standard deviation in the high (μ=4.47%, σ=2.50%) than in the low cognitive load condition (μ=2.69%, σ=1.47%). Moreover, filled pauses are the most represented type of oralized disfluency (we do not take silent pauses into account in the present study), as they make up for 69.86% of the disfluencies for speakers in the low cognitive load condition and 73.04% in the high cognitive load condition.

Regarding the number of filled pauses, neither the normality assumption (Shapiro-Wilk W=0.71, p<.001) nor the homogeneity of variance assumption (Levene's F(1,

21)=4.58, *p*=.044) is met. We thus rely on the Mann-Whitney U test statistics to determine whether the cognitive load condition affects this variable. Results show no significant difference in median number of tokens between the high (Md=12.00) and low cognitive load group (Md=9.00), *U*=57.00, *p*=.600, *r*=-0.14. However, since the assumptions of normality (Shapiro-Wilk W=0.97, p=.614) and homogeneity of variances (Levene's F(1, 21)=2.72, p=.114) are met for the rate of filled pauses, an independent samples t-test (Student's t) is conducted to determine whether cognitive load affects this rate. Results show close-to-significant differences in the rate of filled pauses between the high cognitive load group (μ=4.47, σ=2.50) and the low cognitive load group (μ=2.69, σ=1.47), t(21)=2.06, p=.052, Cohen's d=0.86.

These results indicate that speakers under high cognitive load indeed seem to resort more often to filled pauses. However, whether these are used to indicate uncertainty or betray the buffering of the working memory system, remains an open question.

3.1.4. Counts and rates of self-interruptions

Regarding interruptions and self-corrections, past literature provides contradictory results. Deese (1980) and Oviatt (1995) compare previously planned vs. unplanned speech; Bromme & Wehner (1987), speech describing a restaurant menu vs. a puzzle; Jou & Harris (1992), content playback without vs. with additional mental arithmetics, and Rummer (1996) speech without vs. with an additional visual judgement task: None of these studies reports significant results regarding the effect of cognitive load or associated time constraints on the production of self-corrections. However, Müller et al. (2001) manipulate time pressure and complexity of a navigation task on a screen and find an effect of cognitive load on a collection of disfluencies including self-corrections and interruptions but not filled pauses. Moreover, Marx (1984) finds contradictory results depending on the task: With cognitive load implemented with a time pressure, she finds that time pressure has an effect on the production of self-corrections in the speech of participants aiming at persuading their interlocutors, while it does not in the speech of participants monologuing aloud about a problem. Given these results, we would expect no effect of cognitive load on interruptions in our egocentric speech data.

In EGO, the count of interrupted words ranges from 1 to 8 (μ=3.18, σ=2.48) in the low cognitive load condition and from 0 to 14 in the high cognitive load condition (μ=4.58, σ=3.66). Compared to the number of word-tokens, speakers in the low cognitive load condition interrupted between 0.31% and 2.34% of their words (μ=0.85%, σ=0.62%), against 0.00% to 2.99% for speakers in the high cognitive load condition (μ=1.45%, σ=1.03%). It therefore seems that speakers in the low cognitive load condition interrupt their words less often and more consistently compared with each other.

Since the normality assumption for the number of interruptions is violated (Shapiro-Wilk W=0.90, p=.023), a Mann-Whitney U test is conducted to determine whether cognitive load indeed affects the number of interruptions. Results show no significant differences in the median number of interruptions between the high (Md=4.00) and the low cognitive load group (Md=2.00), U=50.00, p=.334, r=-0.24. The effect of cognitive load is also assessed on the rate of interruptions. Since the assumptions of normality (Shapiro-Wilk W=0.94, p=.157) and homogeneity of variances (Levene's F(1, 21)=3.95, p=.060) are met, an independent samples t-test (Student's t) is conducted. However, results again show no significant differences in the rate of interruptions between the high cognitive load group (μ=1.45, σ=1.03) and the low cognitive load group (μ=0.85, σ=0.62), t(21)=1.65, p=.114, Cohen's d=0.69.

Our results thus seem to confirm past research on the non-effect of cognitive load on self-interruptions.

3.1.5. Counts and rates of repeated words

Regarding the production of repetitions, past research again provides ambiguous results. Bromme & Wehner (1987) and Jou & Harris (1992) find non-significant differences between the high- and low-cognitive load conditions, while Müller et al. (2001) find an effect of cognitive load on self-corrections and interruptions combined. Moreover, Marx (1984) also finds

significant results in the dialogual task but not in the monologual one. We therefore do not expect an effect of cognitive load on the number and rate of repetitions in our data.

In EGO, the number of repeated words ranges from 0 to 6 (μ=1.54, σ=2.42) in the low and from 0 to 17 (μ=2.41, σ=4.96) in the high cognitive load condition. The rates of repeated words as a function of the number of word-tokens range from 0.00% to 3.89% (μ=0.53%, σ=1.16%) in the low and from 0.00% to 1.34% (μ=0.35%, σ=0.49%) in the high cognitive load condition. The means and standard deviations indicate that speakers in the low cognitive load condition repeat their words slightly more often and less consistently. This unexpected tendency is discussed in the conclusion.

The number of repetitions being not normally distributed (Shapiro-Wilk W=0.64, p<.001), we rely on the Mann-Whitney U statistics to determine whether they are affected by cognitive load. Results show no significant difference in median number of tokens between the high cognitive load group (Md=0.00) and low cognitive load group (Md=0.00), U=64.50, p=.944, r=-0.02. A Mann-Whitney U test is also used to determine whether the rate of repetitions differs between the conditions, as the normality assumption is violated for this variable (Shapiro-Wilk W=0.63, p<.001). We find no significant difference between the high (Md=0.00) and low cognitive load groups (Md=0.00), U=62.50, p=.834, r=-0.05.

Our results thus again support past research on the non-effect of cognitive load on the production of repetitions.

3.1.6. Counts and rates of filled pauses, self-interruptions and repetitions combined

However, the lack of significance of our results on the counts and rates of filled pauses, interruptions and repetitions may be due to the small amount of data, since EGO is only 1 hour long and comprises only 11 and 12 participants in each condition respectively. To enhance the counts of observations, we now provide the analysis for oralized disfluencies of all types combined into one category.

Overall, in EGO, all types of disfluencies (filled pauses, interruptions, and repetitions) make up for 0.47% to 6.38% (μ=4.07%, σ=1.86%) of the speech of speakers not submitted to cognitive load, while they make up for 2.74% to 10.82% (μ=6.27%, σ=2.74%) of the speech of speakers who were describing the puzzle from memory. The high cognitive load thus seems to favor the use of disfluencies, as well as a larger heterogeneity among speakers regarding the diversity of words used and the amounts of filled pauses, interruptions and repetitions. However, the effect of cognitive load on the number of disfluencies is assessed with the Mann-Whitney U test and we find, again, no significant difference between the high cognitive load group (Md=16.50) and the low cognitive load group (Md=14.00), U=58.00, p=.64, r=-.12.

However, the effect of cognitive load on the *rate* of disfluencies is assessed with an independent samples t-test. Since the assumptions of normality (Shapiro-Wilk W=0.97, p=.690) and homogeneity of variances (Levene's F(1, 21)=4.03, p=.058) are met, the Student's t-test results are reported. A significant difference is observed, t(21)=2.24, p=.036, Cohen’s d=0.93. As could be expected, a higher disfluency rate is to be observed in the high cognitive load condition (μ=6.27, σ=2.74) compared to the low cognitive load condition (μ=4.07, σ=1.86). This shows that, although particular metrics for each type of disfluency do not yield significant results, the observation of the rate of all types of hesitation-related oralized disfluencies (we exclude silent pauses and laughter) show that speakers under high cognitive load indeed tend to hesitate more than their peers in the low cognitive load condition.

3.2. Lexical metrics: Metrics of word meaning

In this subsection, we focus on the effect that cognitive load may have on the use of the vocabulary, focusing on lexical fields related to geometry, since we expect such words to be of particular interest in this tangram task, as well as perception-related vocabulary, in particular that of sight, which may also yield interesting results given the nature of the task. We then investigate cognition-related terms, as well as the vocabulary of hesitation and tentativeness.

Finally, we analyze the use of inclusion words and more specifically of singular vs. plural and first-person vs. third-person pronouns.

3.2.1. Geometry-related terms

Regarding geometry-related words, we investigate in particular space-, number-, and movement-related terms, and the combination of all three lexical fields under the term "relativity". Given the specificity of the variable, no such analysis has been provided by past studies.

With the normality violated, a Mann-Whitney U test determines the effect of cognitive load on space-related terms. Results show that the high cognitive load group (μ=10.76, σ=2.38) mentions fewer space-related terms than the low cognitive load group (μ=14.12, σ=4.19). The difference is statistically significant, U=29.0, p=.023, r=0.56.

With all assumptions met, an independent samples t-test determined the effect of cognitive load on number-related terms. Results show that the high cognitive load group (μ=3.61, σ=1.34) scores higher on number-related lexical items than the low cognitive load group (μ=2.40, σ=1.37), a statistically significant difference, t(21)=2.14, p=.044, Cohen's d=0.89.

We then determine whether cognitive load affects movement-related words. With all assumptions met, an independent samples t-test determines whether cognitive load affects movement-related words. Results show that the high cognitive load group (μ=6.10, σ=3.21) scores slightly higher on movement words than the low cognitive load group (μ=5.64, σ=3.63), but this difference is not statistically significant, t(21)=0.33, p=.746, Cohen's d=0.14.

Finally, considering the absence of violations of assumptions, an independent samples t-test assesses whether cognitive load affects the production of relativity-related words. Results show that the high cognitive load group (μ=22.22, σ=4.03) scores lower on relativity-related words than the low cognitive load group (μ=25.17, σ=5.07), but this difference is not statistically significant, t(21)=-1.55, p=.135, Cohen's d=-0.65.

Thus, regarding geometry-related words, compared to their peers under low cognitive load, speakers under high cognitive load seem to disfavor space-related words but favor number-related, and to some extent movement-related words. However, when all geometry words are pooled together, the difference loses its significance.

3.2.2. Cognition- and perception-related terms

Under high cognitive load, participants are more likely to be involved in active cognitive processes like thinking, evaluating, and analyzing (Baddeley, 2003; Sweller et al., 2011). Khawaja et al. (2014) thus hypothesize that words designating human cognitive processes (e.g., "think", "consider", "know"...) would increase with increased load, thus mirroring the speakers' self-awareness of increased mental effort. Similarly, the increased use of perceptual words (e.g., "see", "feel", "reckon") would mirror the participants' higher effort to concentrate on the task and to understand the environment that surrounds it in order to improve the overall performance. Khawaja et al. (2014) confirm both hypotheses, however cognitive load in their study relies on increasing visual material while in ours it relies on removing it: We thus hypothesize that the difference in the usage of perceptual words would display the reverse tendency to that observed by Khawaja et al. (2014). Moreover, we would go one step further and suggest that the use of tentative and hesitancy words is bound to increase as well under higher cognitive load, thus mirroring the participants' acknowledging the task's difficulty.

With the assumption of homogeneity of variance violated for cognition-related terms (Levene's p=.002), Welch's t-test is used to determine whether cognitive load affects this variable. Results show little difference in mean cognition-related words between the high (μ=13.50, σ=1.61) and the low cognitive load group (μ=13.41, σ=4.02), Welch's t(12.9)=0.07, p=.947. The effect size is negligible (Cohen's d=0.03). This contradicts Khawaja et al. (2014)'s results.

Regarding the vocabulary of perception, all assumptions are met, meaning that an independent samples t-test is conducted to evaluate the effect of cognitive load on perception-related words. Results show that the high cognitive load group (μ=5.55, σ=1.73) scores lower on perception-related terms than the low cognitive load group (μ=6.59, σ=1.76), but this difference is not statistically significant, t(21)=-1.44, p=.165, Cohen's d=-0.60.

To determine whether cognitive load affects sight-related words, an independent samples t-test is conducted as no assumption violations are observed. Results show that the high cognitive load group (μ=3.12, σ=1.53) scores lower on sight-related words than the low cognitive load group (μ=4.99, σ=1.51), the difference being statistically significant, t(21)=-2.97, p=.007, Cohen's d=-1.24. Speakers under high cognitive load thus seem to disfavor the use of sight-related words: Were this conclusion extended to the use of perception words in general, our results would go against Khawaja et al. (2014)'s, who find the opposite preference, thus aligning with our hypothesis.

Finally, regarding the vocabulary of uncertainty, an independent samples t-test assesses differences between cognitive load conditions for lexical markers of hesitancy since all assumptions are met. Results show that the high cognitive load group (μ=4.69, σ=2.47) scores higher than the low cognitive load group (μ=3.02, σ=1.60), but this difference is not statistically significant, t(21)=1.90, p=.071, Cohen's d=0.79.

Similarly, with all assumptions met, an independent samples t-test assesses the effect of cognitive load on lexical items depicting tentativeness. Results show that the high cognitive load group (μ=1.59, σ=0.77) scores higher on tentative words than the low cognitive load group (μ=1.02, σ=0.58), but this difference is not statistically significant, t(21)=2.00, p=.058, Cohen's d=0.84. For this variable, then, the general observations are in adequacy with our expectations, but the results are not statistically significant, which may also be due to the small amount of data. More research on this matter would be needed.

3.2.3. Inclusion-related terms

Finally, we investigate the participants' relation to themselves and to others by analyzing the vocabulary of inclusion as well as their use of pronouns. Khawaja et al. (2014) hypothesize that an increase in cognitive load will imply an increase in the use of inclusive words, following the observation that people, working in teams and performing complex tasks together prefer to share their load as the task difficulty increases (Kirschner et al., 2009). Similarly, they also expect that increasing the task complexity will cause the decrease of singular pronouns and the increase of plural ones. Both hypotheses are confirmed in Khawaja et al. (2014)'s study on multilogual speech, but what about monologual, self-directed speech?

For EGO, all assumptions are met, meaning that an independent samples t-test can be used to assess the effect of cognitive load on inclusion-related terms (such as "together", "with"...). Results show small differences in mean inclusion terms between the high cognitive load group (μ=9.40, σ=2.18) and the low cognitive load group (μ=8.84, σ=2.05). This difference is not statistically significant, t(21)=0.63, p=.533, Cohen's d=0.26.

Regarding pronouns, with all assumptions met, an independent samples t-test determines whether cognitive load affects terms related to the first-person singular. Results show little difference between the high cognitive load group (μ=4.05, σ=3.18) and the low cognitive load group (μ=3.77, σ=3.47), t(21)=0.20, p=.845, Cohen's d=0.08. Similarly, with the normality assumption violated for first-person plural terms, the Mann-Whitney U test is used. Results show very little use in either the high (Md=0.00) or the low cognitive load group (Md=0.00), with no significant difference, U=65.0, p=.950, r=0.02.

These results again contradict Khawaja et al. (2014)'s results, but they are not surprising since the participants were alone in the room at the time of the experiment and speaking to a future self: Even though it is not impossible to use first-person plural pronouns in this case, which in French is similar to the underspecified pronoun "on" (Engl. "one"), the display of the experiment was not prone to elicit such differences.

Regarding third-person pronouns, with all assumptions met, an independent samples t-test determines whether cognitive load affects third-person singular items. Results show that

the high cognitive load group (μ=7.67, σ=1.97) uses fewer third-person singular terms compared to the low cognitive load group (μ=9.45, σ=3.20), but this difference is not statistically significant, t(21)=-1.62, p=.121, Cohen's d=-0.68. With the normality assumption violated for third-person plural terms, the Mann-Whitney U test is used. Results show little difference in median third-person plural terms between the high (Md=0.00) and low cognitive load group (Md=0.00), U=65.0, p=.970, r=-0.02.

Our results thus do not confirm past hypotheses regarding inclusion and togetherness. However, given the type of speech used in this corpus, i.e., egocentric, self-directed speech, this is not surprising. These results should be confirmed with other dialogual data, for instance with the allocentric data from TACO.

3.3. Lexical metrics: Metrics of word connotation

In this subsection, we focus on the emotional polarity of the speech uttered by the participants, in particular by focusing on the use of negative emotion words, and more specifically on the use of swear words. Negative emotion words, and in particular swear words taken as indices of negative feelings, are expected to increase as the task complexity increases, since high task load is known to cause higher anger and frustration (Farmer & Brownson, 2003). This hypothesis has been confirmed in English-speaking interactions (Khawaja et al., 2014).

With assumptions violated, the non-parametric Mann-Whitney U test is used to assess the effect of cognitive load on words connoting negative emotions. Results show that the high cognitive load group (Md=0.45) indeed scores higher on negative emotion words compared to the low cognitive load group (Md=0.00), but this difference is not statistically significant, U=47.50, p=.139, r=.04.

Similarly, regarding swear words, both the assumptions of normality and homogeneity of variance are violated. Therefore, the non-parametric Mann-Whitney U test is used to explore the effect of cognitive load on swear words. Results show, again, a non-significant difference, as the high cognitive load group (Md=0.27) uses slightly fewer words identified by the algorithm as swear words compared to the low cognitive load group (Md=0.55), U=50.50, p=.208, r=.30.

In conclusion, although speakers of French under higher cognitive load seem to display more negative emotions, as do English speakers, the results are not significant enough to draw any clear-cut conclusion.

4. Conclusions and Discussion

In this paper, we investigate the effect of cognitive load on speech in Belgian French, in particular on fluency, vocabulary and emotional polarity. The results are summarized in Table 3.

Metrics	LCL	HCL	Delta	Significance
Disfluencies				
Number of word-tokens per speaker	μ=397.36 σ=200.06	μ=391.5 σ=344.80	μ=5.86 σ=-144.74	Not significant
Number of lemmas per speaker	μ=88.18 σ=30.63	μ=80.83 σ=30.35	μ=7.35 σ=-0.28	Not significant
Lemma-token ratio	μ=24.45% σ=5.77%	μ=26.55% σ=8.44%	μ=-2.10% σ=-2.67%	Not significant
Number of filled pauses per speaker	μ=11.91 σ=9.79	μ=23.50 σ=33.12	μ=-11.59 σ=-23.33	Not significant
Rates of filled pauses	μ=2.69% σ=1.47%	μ=4.47% σ=2.50%	μ=-1.78% σ=-1.03%	Not significant
Number of self-interruptions by speaker	μ=3.18 σ=2.48	μ=4.58 σ=3.66	μ=-1.40 σ=-1.18	Not significant

Rates of self-interruptions	μ=0.85% σ=0.62%	μ=1.45% σ=1.03%	μ=0.60% σ=-0.41%	Not significant
Number of repeated words per speaker	μ=1.54 σ=2.42	μ=2.41 σ=4.96	μ=-0.87 σ=-2.54	Not significant
Rates of repeated words	μ=0.53% σ=1.16%	μ=0.35% σ=0.49%	μ=0.18 σ=0.67	Not significant
Number of disfluencies	μ=16,64 σ=12.18	μ=30.50 σ=40.10	μ=-13.86 σ=-27.91	Not significant
Rates of disfluencies (filled pauses, self-interruptions and repeated words)	**μ=4.07% σ=1.86%**	**μ=6.27% σ=2.74%**	**μ=-2.20 σ=-0.88**	**Significant**
Lexical metrics				
Rates of space-related words	**μ=14.12 σ=4.19**	**μ=10.76 σ=2.38**	**μ=3.36 σ=1,81**	**Significant**
Rates of number-related words	**μ=2.40 σ=1.37**	**μ=3.61 σ=1.34**	**μ=-1.21 σ=0.03**	**Significant**
Rates of movement-related words	μ=5.64 σ=3.63	μ=6.10 σ=3.21	μ=-0.46 σ=0.42	Not significant
Rates of relativity-related words (space, number and movement)	μ=25.17 σ=5.07	μ=22.22 σ=4.03	μ=2.95 σ=1.04	Not significant
Rates of cognition-related words	μ=13.41 σ=4.02	μ=13.50 σ=1.61	μ=-0.09 σ=2.41	Not significant
Rates of perception-related words	μ=6.59 σ=1.76	μ=5.55 σ=1.73	μ=1.04 σ=0.03	Not significant
Rates of sight-related words	**μ=4.99 σ=1.51**	**μ=3.12 σ=1.53**	**μ=1.87 σ=-0.02**	**Significant**
Rates of uncertainty words	μ=3.02 σ=1.60	μ=4.69 σ=2.47	μ=-1.67 σ=-0.87	Not significant
Rates of tentativeness-related words	μ=1.02 σ=0.58	μ=1.59 σ=0.77	μ=-0.57 σ=-0.19	Not significant
Rates of inclusion-related words	μ=8.84 σ=2.05	μ=9.40 σ=2.18	μ=-0.56 σ=-0.13	Not significant
Rates of first-person singular pronouns	μ=3.77 σ=3.47	μ=4.05 σ=3.18	μ=-0.28 σ=0.29	Not significant
Rates of third-person pronouns	μ=9.45 σ=3.20	μ=7.67 σ=1.97	μ=1.78 σ=1.23	Not significant
Rates of words connoting negative emotions	Md=0.00	Md=0.45	Md=-0.45	Not significant
Rates of swear words	Md=0.55	Md=0.27	Md=0.28	Not significant

Table 3. Mean (μ) and standard deviation (σ) for each metrics in the low cognitive load (LCL) and high cognitive load (HCL) conditions, the delta between the baseline (LCL) and the experimental condition (HCL), and statistical significance of the results. Significant differences between the two conditions are in bold.

Regarding fluency, we show that the counts of word-tokens and lemmas do not significantly differ between speakers under high cognitive load and speakers under low cognitive load. However, the lemma-token ratio shows some small (non-significant) differences, which display the unexpected tendency, i.e., that speakers in the high cognitive load condition tend to use more diversified vocabulary. Regarding disfluencies, counts and rates of filled pauses, word interruptions and word repetitions are investigated separately. Observations confirm the expected tendencies, with more disfluencies in high-cognitive load speech, and also show that the higher cognitive load favors inter-individual variability in the number of words and the use of disfluency devices. However, results of statistical analyses show that none of these disfluency types are significantly affected by the higher cognitive load. Since the lack of significance may be due to the small amount of data used in this study, we also investigate all disfluencies merged into one macro-category. For disfluencies in general, there is no difference between the two groups of speakers regarding their raw counts, however

there is a significant difference in the rates of disfluencies proportionally to the amount of speech uttered, with more disfluencies in the speech of participants in the high cognitive load condition, which confirms past literature.

Regarding lexical metrics, we investigate geometry-related terms, that were expected to show some lexical difference given the nature of the task, and we find a significant difference in the space- and number-related words, with less space-related vocabulary but more number-related vocabulary in the high cognitive load condition. We also investigate cognition- and perception-related terms but find no difference between the two speaker groups, contrary to past research. However, when focusing on a specific type of perception words, i.e., the vocabulary of sight, we do find that the high cognitive load speakers use less of them, which could be expected from the nature of the task. We also investigate the vocabulary of hesitancy and tentativeness and find the expected tendencies, with more of such words in the high cognitive load condition, but not in a statistically significant way. Finally, we analyze the vocabulary of inclusion and the use of pronouns (singular vs. plural, first vs. third person), but find no differences between the groups on either parameter.

Finally, we investigate the emotional polarity of speakers in each group and find that speakers under high cognitive load indeed use more negative emotion words, but not in a significant way. When focusing on swear words as a proxy for expression of anger and frustration, results are not significant either.

To sum up, we confirm the result of past literature on other languages by showing that speakers of Belgian French indeed have a higher tendency to disfluency when submitted to a higher cognitive load. Our results thus indicate that at least this one parameter could be a cross-linguistic cue of higher cognitive demand during speech production. This should of course be confirmed on a larger sample of languages from more diversified language families.

However, significant lexical differences are to be found in the more specific vocabulary related to space, numbers and sight, while we find no differences regarding other general vocabulary, such as cognition-, perception-, tentativeness- or inclusion-related words used by the speakers when they are conscious of being submitted to a (difficult) task, nor regarding the emotional polarity of words. These results differ from Khawaja et al. (2014)'s results on English, which is probably due to the different tasks used to increase cognitive load (more vs. less visual material) and the different settings (multilogual vs. monologual / self-directed speech). It may also be that English and French differ altogether in their uses of such meta-linguistic vocabulary that comment on ongoing actions, but further studies would be needed to confirm this hypothesis.

The present study could benefit from additional measurements. In particular, the duration of filled pauses, as well as that of silent pauses, could refine our understanding of the impact of cognitive load on fluency. Along similar lines, a more precise investigation of repetitions may be in order. In the present work, were counted as repetitions only repetitions of one word immediately after itself. Yet when looking at the data, it appears that some repetitions may involve larger chunks of words, sometimes with correction. A manual identification of such sequences would be useful for future investigations. Finally, it should be noted that the differences observed between the two groups may be due to other variables than the experimental condition. In particular, genders were not balanced among the two groups, with much more female speakers in the low cognitive load group and much more male speakers in the high cognitive load group. Speakers from the low cognitive load group are also much older than those from the high cognitive load group (δ=13.30). Given that older age may impair performance or ease in such technology-oriented settings, it is possible that some differences would have been more obvious or more significant between the two groups if the speakers from the low cognitive load condition had been younger. A more balanced dataset is thus needed to confirm the present observations.

Nonetheless, our results fill the gap in the literature regarding the effect of cognitive load on the processing of the French language. We show that some tendencies found in other studies on other languages such as English and German can be generalized, and could thus be cross-linguistic tendencies, i.e., language universals. However, these results should be confirmed with a larger amount of data, since, as could be seen, some of the results are not

significant due to the small amount of data available. Moreover, some differences may need to be investigated on other types of speech, such as the use of pronouns which was not properly observable given the particular monologual, self-directed speech elicited. In future studies, we intend to augment EGO to re-run these analyses and confirm or infirm the preliminary results found here, and also to conduct these analyses on the other half of TACO, where dialogical speech is elicited, in combination with Knusen et al. (2020)'s dialogual tangram corpus.

5. Acknowledgements

The authors wish to thank the two anonymous reviewers who not only provided valuable insights, but also a more accurate, cleaner and more elegant version of the original Python script used in this study. This work was also partly supported by an F.R.S.-FNRS grant to the project *PPaDisM: Phonetic Patterns in Discourse Markers* (PI: Mathilde Hutin).

6. References

Baddeley, Alan. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science*, 4, 417–423. https://doi.org/10.1016/s1364-6613(00)01538-2

Baddeley, Alan. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, 36, 189–208. https://doi.org/10.1016/s0021-9924(03)00019-4

Baddeley, Alan. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*. Vol. 63:1-29 (Volume publication date January 2012). First published online as a Review in Advance on September 27, 2011. https://doi.org/10.1146/annurev-psych-120710-100422.

Barnes, Sally, & Lewin, Cathy. (2005). An introduction to inferential statistics: Testing for differences and relationships. *Research methods in the social sciences*, 226-235.

Berthold, Andre. (1998) *Repräsentation und Verarbeitung sprachlicher Indikatoren für kognitive Ressourcenbeschränkungen* [Representation and processing of linguistic indicators of cognitive resource limitations]. Master's thesis, Department of Computer Science, Saarland University, Germany, 1998

Boersma, Paul & Weenink, David (2019). *Praat: doing phonetics by computer* [Computer program]. Version 6.1, retrieved 13 July 2019 from http://www.praat.org/

Brennan, S. E. & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of memory and language*, vol. 34, no. 3: 383–398. https://doi.org/10.1006/jmla.1995.1017

Broisson, Zoë. (2019). *How egocentric is language production? Evidence from discourse marker use under cognitive load*. Master's thesis. UCLouvain. https://hdl.handle.net/2078.2/14297

Broisson, Zoë & Degand, Liesbeth. (2022). 6 - How egocentric is discourse marker use? Investigating the impact of speaker orientation and cognitive load on discourse marker production. *Discourse Markers in Interaction: From Production to Comprehension*, edited by Maria-Josep Cuenca and Liesbeth Degand, Berlin, Boston: De Gruyter Mouton, 121-158. https://doi.org/10.1515/9783110790351-006

Bromme, Rainer & Wehner, Theo. (1987). Zum Zusammenhang von Sprechgeschwindigkeit und Sprechfehlern mit der Aufgabenschwierigkeit beim lauten Denken. *Zeitschrift für experimentelle und angewandte Psychologie*, 34, 1-16

Chandler, Paul & Sweller, John. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8, 293–332. https://doi.org/10.1207/s1532690xci0804_2

Christenfeld, Nicholas. (1994). Options and ums. *Journal of Language and Social Psychology*, 13(2), 192–199. https://doi.org/10.1177/0261927X94132005

Danino, Charlotte, Knutsen, Dominique & Col, Gilles. 2020. Naviguer dans le dialogue et faire voir ce que l'on dit : approches linguistique et psycholinguistique de "voilà". In Marta Saiz-Sánchez; Amalia Rodríguez Somolinos; Sonia Gómez-Jordana Ferary (eds) : *Marques*

d'oralité et représentation de l'oral en français, 20, Presses universitaires Savoie Mont Blanc, 10 2377410146. ⟨hal-02949361v2⟩

Deese, James. (1980). Pauses, prosody, and the demands of production in language. In H. W. Dechert & M. Raupach (Hrsg.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler*, 69-84. The Hague: Mouton. https://doi.org/10.1515/9783110816570.69

Farmer, E., & Brownson, A. (2003). *Review of workload measurement, analysis and interpretation methods* (No. CARE-Integra-TRS-130-02-WP2). Brussels, Belgium: European Organisation for the Safety of Air Navigation.

Foushee, H. C., & Helmreich, R. L. (1988). Group interaction and flight crew performance. In E. Wiener & D. Nagel (Eds.), *Human factors in aviation*, 189–227. New York, NY: Academic Press. https://psycnet.apa.org/record/1988-98354-006

Fox Tree, Jean E. (2001). Listeners' uses of *um* and *uh* in speech comprehension. *Memory & Cognition* 29(2), 320–326. https://doi.org/10.3758/bf03194926

Fox Tree, Jean E. (2007). Functional spontaneous speech phenomena. *Perspectives on Fluency and Fluency Disorders*, 17(2), 17-19. https://doi.org/10.1044/ffd17.2.17

Hutin, Mathilde, Hu, Junfei & Degand, Liesbeth. (2024). Uh, um and mh: Are filled pauses prone to conversational converge? *Proceedings of Interspeech 2024*, 3575-3579. https://doi.org/10.21437/Interspeech.2024-1168

The jamovi project (2023). *jamovi* (Version 2.3) [Computer Software]. Retrieved from https://www.jamovi.org

Janssen, Jeroen, Kirschner, Femke, Erkens, Gijsbert, Kirschner, Paul A., & Paas, Fred. (2010). Making the black box of collaborative learning transparent: Combining process-oriented and cognitive load approaches. *Educational Psychology Review*, 22, 139–154. https://doi.org/10.1007/s10648-010-9131-x

Jou, Jerwen & Harris, Richard J. (1992). The effect of divided attention on speech production. *Bulletin of the Psychonomic Society*, 30, 301-304. https://doi.org/10.3758/BF03330471

Katz, C., Fraser, E. B., & Wagner, T. L. (1998). Rotary-wing crew communication patterns across workload levels. *RTO HFM Symposium on "Current Aeromedical Issues in Rotary Wing Operations,"* 14.1–14.3.

Khawaja, M. Asif, Chen, Fang & Marcus, Nadine. (2014) Measuring Cognitive Load Using Linguistic Features: Implications for Usability Evaluation and Adaptive Interaction Design, *International Journal of Human-Computer Interaction*, 30:5, 343-368, DOI: http://dx.doi.org/10.1080/10447318.2013.860579

Kirschner, Femke, Paas, Fred & Kirschner, Paul A. (2009). Cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review*, 21, 31-42. https://doi.org/10.1007/s10648-008-9095-2

Kleinman, D. L., & Serfaty, Daniel. (1989). Team performance assessment in distributed decision making. *Proceedings of the Interactive Networked Simulation for Training Conference*, 22-27.

Knutsen, Dominique, Col, Gilles and Rouet, Jean-François. 2020. L'apport de la méthode expérimentale à l'étude de certains aspects de voilà". *Polysémie, usages et fonctions de « voilà »*, edited by Gilles Col, Charlotte Danino and Stéphane Bikialo, Berlin, Boston: De Gruyter, 2020, 259-298. https://doi.org/10.1515/9783110622454-008

Lakens, Daniël. (2022). Sample size justification. *Collabra: psychology*, 8(1), 33267. https://doi.org/10.1525/collabra.33267

Laserna, Charlyn M., Seih, Yi-Tai, & Pennebaker, James W. (2014). Um… Who Like Says You Know: Filler Word Use as a Function of Age, Gender, and Personality. *Journal of Language and Social Psychology*, 33(3), 328-338. https://doi.org/10.1177/0261927X14526993

Lennon, C., & Burdick, H. (2004). *The Lexile Framework as an Approach for Reading Measurement and Success* (White paper from The Lexile Framework for Reading). Retrieved from http://www.Lexile.com

Marx, Edeltrud. (1984). *Über die Wirkung von Zeitdruck auf Sprachproduktionsprozesse*. Dissertation, Universität Münster

Müller, Christian, Großmann-Hutter, Barbara, Jameson, Anthony, Rummer, Ralf, & Wittig, Frank. (2001). Recognizing time pressure and cognitive load on the basis of speech: An experimental study. *UM2001, user modeling: Proceedings of the Eighth International Conference*, 24–33. Berlin, Germany: Springer. https://doi.org/10.1007/3-540-44566-8_3

Oser, Randall, Prince, Carolyn, Morgan, Ben & Simpson, Steven. (1991). *An analysis of aircrew communication patterns and content* (No. 90-009). Orlando, FL: Naval Training Systems Centre, Human Factors Division. Retrieved from https://apps.dtic.mil/sti/pdfs/ADA246618.pdf

Oviatt, Sharon. (1995). Predicting spoken disfluencies during human-computer interaction. *Computer Speech and Language*, 9, 19-35. https://doi.org/10.1006/csla.1995.0002

Paas, Fred, Tuovinen, Juhani E., Tabbers, Huib, & van Gerven, Pascal. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38, 63–71. https://doi.org/10.1207/S15326985EP3801_8

Paas, Fred, van Gog, Tamara & Sweller, John. (2010). Cognitive Load Theory: New Conceptualizations, Specifications, and Integrated Research Perspectives. *Educational Psychology Review* 22, 115–121. https://doi.org/10.1007/s10648-010-9133-8

Piolat, Annie, Booth, R.J., Chung, C.K., Davids, M., Pennebaker, James W. (2011). La version française du dictionnaire pour le LIWC : modalités de construction et exemples d'utilisation, *Psychologie Française*, Vol. 56 (3), 145-159. https://doi.org/10.1016/j.psfr.2011.07.002.

R Core Team (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/

Roßnagel, Christian. (1995). Kognitive Belastung und Hörerorientierung beim monologischen Instruieren. *Zeitschrift für experimentelle und angewandte Psychologie*, 42, 94-110.

Roßnagel, Christian. (2000). Cognitive load and perspective-taking: Applying the automatic controlled distinction to verbal communication. *European J. of Social Psychology*, 30 (3), 429-445. https://doi.org/10.1002/(SICI)1099-0992(200005/06)30:3<429::AID-EJSP3>3.0.CO;2-V

Rummer, Ralf. (1996). *Kognitive Beanspruchung beim Sprechen*. Weinheim: Beltz.

Schneider, Ulrike. (2014). *Frequency, Chunks and Hesitations. A Usage-based Analysis of Chunking in English*. Freiburg: Universität Freiburg.

Sexton, J. Bryan, & Helmreich, Robert L. (2000). Analyzing cockpit communication: The links between language, performance, error, and workload. *Journal of the Human Performance in Extreme Environments*, 5, 63–68. https://doi.org/10.7771/2327-2937.1007

Sweller, John. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12, 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, John, Ayres, Paul, & Kalyuga, Slava. (2011). *Cognitive load theory*. New York, NY: Springer-Verlag. https://doi.org/10.1007/978-1-4419-8126-4

Tan, Deborah H., & Jiang, Yuhong V. (2020). Tell me what you saw: The usefulness of verbal descriptions for others. *Quarterly Journal of Experimental Psychology*, 73(8), 1227-1241. https://doi.org/10.1177/1747021820915356

Van Rossum, Guido, & Drake, Fred L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace. ISBN**:** 978-1-4414-1269-0