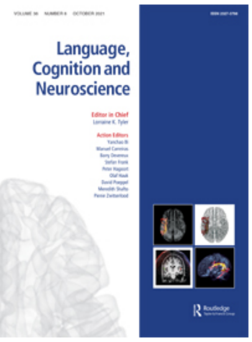

Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Eye-movements can help disentangle mechanisms underlying disfluency

Aurélie Pistono & Robert J. Hartsuiker

To cite this article: Aurélie Pistono & Robert J. Hartsuiker (2021) Eye-movements can help disentangle mechanisms underlying disfluency, Language, Cognition and Neuroscience, 36:8, 1038-1055, DOI: 10.1080/23273798.2021.1905166

To link to this article: https://doi.org/10.1080/23273798.2021.1905166

© European Union 2021. Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 26 Mar 2021.

View supplementary material

Submit your article to this journal

Article views: 1391

View related articles

View Crossmark data

Citing articles: 4 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=plcp21

LANGUAGE, COGNITION AND NEUROSCIENCE
2021, VOL. 36, NO. 8, 1038–1055
https://doi.org/10.1080/23273798.2021.1905166

REGULAR ARTICLE

OPEN ACCESS

Eye-movements can help disentangle mechanisms underlying disfluency

Aurélie Pistono and Robert J. Hartsuiker

Department of Experimental Psychology, Ghent University, Gent, Belgium

ABSTRACT

To reveal the underlying cause of disfluency, several authors related the pattern of disfluencies to difficulties at specific levels of production, using a Network Task. Given that disfluencies are multifactorial, we combined this paradigm with eye-tracking to disentangle disfluency related to word preparation difficulties from others (e.g. stalling strategies). We manipulated lexical and grammatical selection difficulty. In Experiment 1, lines connecting the pictures varied in length, which led participants to use a strategy and inspect other areas than the upcoming picture when pictures were preceded by long lines. Experiment 2 only used short lines. In both experiments, lexical selection difficulty promoted self-corrections, pauses and longer fixation latency prior to naming. Multivariate Pattern Analyses demonstrated that disfluency and eye-movement data patterns can predict lexical selection difficulty. Eye-tracking could provide complementary information about network tasks, by disentangling disfluencies related to picture naming from disfluencies related to self-monitoring or stalling strategies.

ARTICLE HISTORY
Received 21 April 2020
Accepted 10 March 2021

KEYWORDS
Disfluency; eye tracking; connected speech; lexical access; network task

Introduction

Natural speech production is full of disfluencies, which are defined as *phenomena that interrupt the flow of speech and do not add propositional content to an utterance* (Fox Tree, 1995, p. 709). This term includes various phenomena such as filled or silent pauses, repeated words, and self-corrections. Despite the high frequency of disfluencies, the question remains as to why speakers are disfluent. It is often argued that disfluencies occur when the speaker faces difficulty in language production (e.g. in creating a message or finding a word, Maclay & Osgood, 1959), but it remains unclear whether this is true for all levels of language production defined in current psycholinguistic models (e.g. Dell, 1986; Levelt, 1989).

There is however a consensus that there are multiple factors underlying disfluencies. According to one approach (Clark & Fox Tree, 2002), filled pauses such as “uh” or “um” are considered as words, used by speakers to announce to their listeners that they are initiating what they expect to be a delay before speaking. These approaches argue for specific functions of some disfluencies that are not always related to word preparation difficulties – for instance stalling for time or creating a particular effect in a discourse. Within the language production system, several processing levels may be involved in the production of disfluencies. Some studies used a Network Task to investigate this issue (Hartsuiker & Notebaert, 2010; Oomen & Postma, 2001; Oomen & Postma, 2002; Schnadt & Corley, 2006). In this task (Figure 1), participants describe a route taken by a point marker through a network of pictures so that a listener could fill in a blank network by listening to the description. This paradigm allows for the manipulation of the items so as to create difficulties at specific production stages (e.g. conceptual generation) while holding other stages constant. It has been shown that impeding the conceptual generation of a message using blurry images increased the rate of disfluencies (Schnadt & Corley, 2006). The network task also taps into self-monitoring processes (Nozari & Novick, 2017) as it elicits a substantial number of errors and self-repairs (Oomen & Postma, 2001). Hartsuiker and Notebaert (2010) caused difficulty in the initial stage of lexical access by manipulating name agreement (i.e. the number of different names speakers use to refer to an object) and showed that pictures with low agreement names induced more pauses and self-corrections than pictures with high agreement names. These authors also considered grammatical gender, which is marked on determiners in Dutch. In languages with grammatical

CONTACT Aurélie Pistono aurelie.pistono@ugent.be Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, Gent 9000, Belgium

Supplemental data for this article can be accessed at https://doi.org/10.1080/23273798.2021.1905166.

© European Union 2021. Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

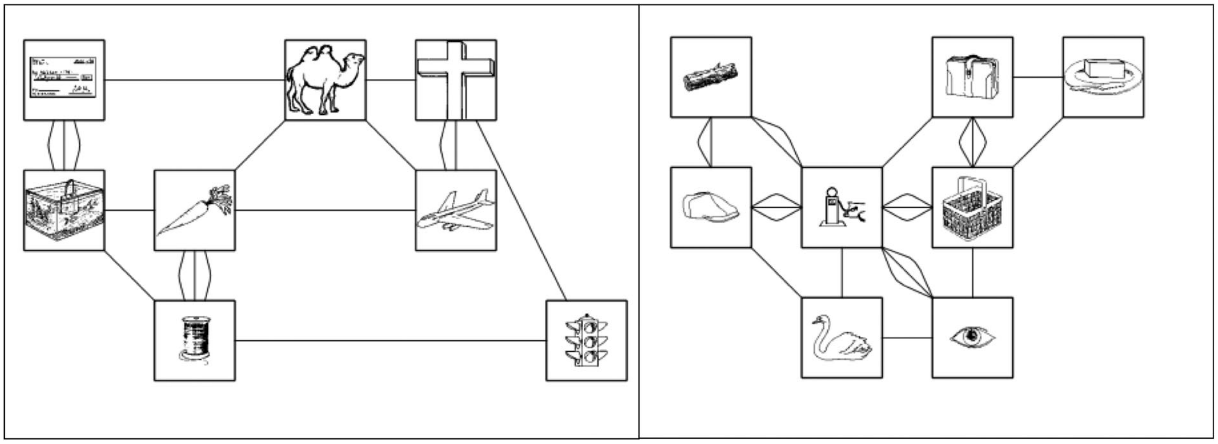

Figure 1. Example of a network for Experiment 1 (left) and Experiment 2 (right).

gender, determiner selection occurs after noun selection as the determiner must agree with the gender of the noun (e.g. Dhooge et al., 2016). Given that in Dutch the common-gender determiner "de" occurs more frequently than the neuter-gender determiner "het", it is possible that choosing "het" is more difficult and therefore results in more disfluencies. In accordance with this hypothesis, Hartsuiker and Notebaert (2010) showed that picture names with the less-frequent (i.e. neuter) Dutch gender induced more disfluencies than picture names with the more-frequent (i.e. common) Dutch gender.

In other words, various mechanisms can lead to disfluency: on the one hand, several levels of the language system are responsible for disfluency production, on the other hand, not all disfluencies occur with the language system (e.g. some reflect a stalling strategy). It is therefore important to disentangle their underlying functions, and uncover whether different types of difficulties are associated with different disfluency phenomena. To do so, the current study combines network tasks with eye-tracking, which allows identifying whether, when, and for how long speakers look at items they have to mention. It will address two broad questions: which pattern of disfluency and eye-movements occur with lexical selection difficulty and grammatical selection difficulty? Which disfluencies and eye-movements are associated with other mechanisms than difficulties in speech encoding, such as stalling strategies?

Indeed, eye movements to visual objects are closely tied to the production of speech about these objects (see Ferreira & Rehrig, 2019 for a review of the current literature on eye-movements and language production). Speakers usually fixate upon an object just before mentioning it, and the dynamic between eyes and voice (eye-voice span, Levin & Buckler-Addis, 1979) varies with the time required to plan the oral production. In particular, eye-voice span before speech onset (onset-EVS, reflecting the latency between the start of the first gaze at the picture and the onset of its name) increases when pictures are relatively difficult to name, such as low name agreement pictures (Griffin, 2001), pictures with low-frequent names, or visually degraded images (Meyer et al., 1998). EVS is also influenced by other mechanisms than linguistic properties during naturalistic scene descriptions (i.e. perceptual, conceptual and structural guidance, Coco & Keller, 2015). EVS following name onset is also tied to language production. Using single sentence description, Griffin (2004) showed that eye-movements follow the order of picture description: from 100 to 300 ms before saying an object's name, speakers shift their gaze to the next object to be named. This offset-EVS (i.e. latency between the end of the last gaze at the picture and the offset of its name) has been argued to coincide with the end of phonological encoding.

Because of this close coupling between speaking and seeing, eye-movements may be indicative of the mechanism underlying the disfluency. Nevertheless, to date only few studies have analysed disfluencies in relation with gaze. Focusing on a corpus of self-corrected speech errors, Griffin (2004) showed that the EVS is different for correct productions and errors. Specifically, speakers are usually still gazing at the object while initiating an erroneous name. This finding means that self-corrections do not necessarily result from rushed word preparation (i.e. if so, speakers should spend less time gazing at pictures before uttering errors). On the contrary, it indicates that when self-corrected speech errors occur, knowing which object a speaker gazed at is informative about the detection and monitoring of this error. Another study showed that disfluencies and gazes are sometimes related to a more general incremental message planning (Brown-Schmidt & Tanenhaus, 2006). In this study, pairs of participants took turns telling each other to click on a target picture among various items. Some trials included a contrast picture that differed from the target in size. The authors showed that the timing of speakers' fixations on the contrast picture predicted the type of phrase and the production of disfluency (e.g. the contrast can be mentioned with a pre-nominal adjective: *the small triangle* vs. post-noun repair: *the triangle, uh small one*). In other words, the authors showed that disfluency was a way to gain time to add additional information to a planned utterance, when the speaker saw new information that had to be integrated into the current utterance. According to this study, some disfluencies are rather strategic than unintentional, which is consistent with the predictions made by Clark and Fox Tree (2002) about strategic filled pauses.

We therefore propose that eye-movements can partly disentangle the underlying functions of disfluencies. First, eye-tracking will allow us to examine whether longer connected-speech production (network tasks) induces a similar eye-speech synchronisation as picture naming or single sentence production (i.e. studies mentioned above). We will test whether visual attention increases with lexical selection difficulty and grammatical selection difficulty. Second, we will also investigate whether participants sometimes make use of anticipatory and late fixations (i.e. do not follow the pace of the marker) when describing a network, and whether these behaviours are associated with disfluency. In particular, disfluency occurring while the speaker is gazing at an upcoming picture probably reflects a "stalling strategy" (i.e. similarly to Brown-Schmidt & Tanenhaus, 2006). On the contrary, additional time spent on an picture after speech onset (in terms of number of fixations, offset-EVS or late fixations), will probably be related to self-monitoring processes (Griffin, 2004). Finally, we will discriminate, regardless of the lexical and grammatical difficulties of the pictures, disfluencies occurring during onset-EVS from other disfluencies. Indeed, while previous studies using network tasks analysed disfluencies related to picture naming, we will be able to tackle disfluencies that are actually occurring when the speaker is gazing at an item and about to produce its name. These disfluencies will be more likely to reflect word preparation difficulty. To test these assumptions, we will replicate the network task study of Hartsuiker and Notebaert (2010) on name agreement with two changes. First, the previous study included more pictures with common-gender than with neuter-gender names, but we selected equal numbers of each gender to test grammatical selection as well. As in the previous study, we held word frequency, age of acquisition, and word length in syllables constant. Second, as mentioned above, we combined this paradigm with eye-tracking, to address three questions: (i) Which pattern of disfluency occurs depending on lexical selection difficulty and grammatical selection difficulty? We aim at replicating the finding that lexical selection difficulty induces pauses and self-corrections and testing whether grammatical selection difficulty (i.e. neuter gender) induces disfluency. (ii) What is the pattern of eye-movements during lexical or grammatical selection? We predict that lexical selection difficulty will induce longer onset-EVS, similarly to single picture naming, and more fixations on the picture. Indeed, the number of fixations on a picture during a naming task predicts anomia for the same item in patients with lexical difficulties (Reilly et al., 2020). It is therefore possible that lexical access difficulties induce a similar pattern of eye-movements in healthy participants. We have no predictions regarding grammatical selection. (iii) What is the relationship between disfluencies and eye-movements? We predict that, regardless of the manipulations (name agreement and gender), some disfluencies will predict longer onset-EVS, while others will be associated with anticipatory and late fixations. In other words, not all disfluencies will be related to picture naming difficulties.

Finally, we aimed at examining whether, by contrast, lexical selection difficulty or grammatical selection difficulty could be predicted based on the pattern of eye-movements and disfluency associated with it, using multivariate pattern analyses (MVPA, Haynes & Rees, 2006). Indeed, traditionally, each variable is treated as a dependent variable to determine whether that variable varies according to the experimental conditions. By contrast, MVPA extracts the information contained in the pattern of information available, to test whether experimental conditions can be distinguished from one another on the basis of the patterns observed. MVPA has mostly been used in neuroimaging, to infer stimulus specific representation (e.g. Senoussi et al., 2016) or cognitive state (e.g. Craddock et al., 2009) based on the pattern of cortical activity. More recently, it has been demonstrated that the viewing task a person is engaged in (scene memorisation task, reading task, scene search task, or pseudo-reading task) could be classified from the pattern of eye movements (Henderson et al., 2013). This means that eye movements code a specific pattern of information about a viewing task that a classifier can learn and then use to predict which task viewers were performing. In the present study, we applied multivariate pattern classification to examine whether the type of linguistic difficulty could be inferred based on the pattern of disfluency and eye-movements. More precisely, we tested whether name agreement and gender could be predicted based on the pattern of eye-movements or the pattern of disfluency (and which features were the most consistent at a group level). This type of analysis was done to reinforce previous analyses and provide further information about the stability of the variables under study across participants:

Experiment 1

Data, scripts and written transcripts to run the experiments are made available on OSF: https://osf.io/9yhcb/.

Material and methods

Participants

Twenty bachelor students, all native speakers of Dutch, participated in the experiment in exchange for course credit. The samples have been calculated using guidelines for mixed models in designs with repeated measures, for which 1600 observations per condition are required (Brysbaert & Stevens, 2018). One participant was excluded after analyses because more than 80% of the trials were excluded. In the current study, a trial refers to a picture. A trial was excluded if the participant: used the wrong name (i.e. not the dominant name) while naming it; did not produce anything; used the plural or an indefinite determiner (plural and indefinite determiners are not gender-marked in Dutch) or a diminutive (which always has neuter gender in Dutch); omitted the determiner. In total, 864 trials (on a total of 3200, 27%), were excluded in Experiment 1. The final sample consisted of 19 participants (16 Females and 3 Males); mean age was 19 ± 1 years old.

Material

We constructed 20 networks using a programme written in Psychopy (Peirce, 2007). Each network consisted of eight interconnected black- and-white line pictures. The pictures were either connected by one, two, or three straight lines or curves. Lines were either horizontal, vertical, or diagonal. Curves could be horizontal or vertical. The route through the network was indicated by a moving red dot that traversed the network in 42 s (Similarly to Hartsuiker & Notebaert, 2010). For Experiment 1, networks were created randomly and could have either short or long lines.

To construct the networks, 160 line drawings were selected from the set of pictures that Severens et al. (2005) normed for Dutch. Eighty pictures had high name agreement and eighty had low name agreement. Name agreement was based on the H-statistic (Snodgrass & Vanderwart, 1980). The lower H, the fewer names are used; when all participants use the same name, H is 0. Mean H for low name agreement pictures was 1.8 (±0.44) and mean H for high name agreement pictures was 0.4 (±0.38). Eighty pictures had a common gender name and eighty had a neuter gender name. In each network there were two pictures with low name agreement-common gender; two pictures with low name agreement-neuter gender; two pictures with high name agreement-common gender; two pictures with high name agreement-neuter gender. The type and number of lines connecting the pictures, as well as the order and location of appearance of the 160 pictures were randomised across participants. The pictures were matched across sets for the log frequency, age of acquisition, number of syllables, and number of phonemes of the dominant names (Table 1).

Apparatus

The experiment was implemented using Psychopy (Peirce, 2007), to display networks and record both eye-tracking and speech production. Eye movements from participants' dominant eye were monitored with an EyeLink 1000+ system, with a sampling rate of 500 Hz. The monitor display resolution was 1921 × 1081 pixels, at a viewing distance of 88 cm. Pictures' resolution was 150 × 150 pixels, subtending 2.8° of visual angle. Head movements were minimised with chin/head rests.

Procedure

The participants were tested individually in a quiet room. First, in a familiarisation phase, the participants saw the 160 pictures subsequently, along with their dominant names. Participants were instructed to study each picture and its name, and to press the space bar to see

Table 1. Mean (±SD) log frequency per million, age of acquisition (AoA), number of syllables, name agreement (H-statistic), in isolation for the low (LNA) and high (HNA) name agreement pictures, common and neuter gender pictures (from Severens et al., 2005).

	LNA	HNA	*p* value	Common	Neuter	*p* value
log frequency	1.3 ± 0.7	1.4 ± 0.48	0.14	1.3 ± 0.73	1.4 ± 0.44	0.20
AoA	5.9 ± 1	5.7 ± 0.95	0.33	5.9 ± 0.79	5.8 ± 1.2	0.45
Number of syllables	1.7 ± 0.59	1.5 ± 0.59	0.34	1.6 ± 0.72	1.6 ± 0.76	0.46
H-statistic	1.8 ± 0.44	0.4 ± 0.38	<.0001	1.1 ± 0.84	1.1 ± 0.84	0.9

the next picture. This was done to ensure that difficulties with low-agreement stimuli were truly difficulties in choosing among lexical items, not difficulties in visual object recognition. Then the participants took their place in front of a computer screen which displayed an example network. Instructions were given to provide an accurate description of the network using complete sentences and to synchronise the description with the dot that moved through the network. Instructions emphasised that a complete description mentioned the route of the dot, including the shape and direction of the lines or curves, and including the objects. The instructions explicitly mentioned that it was not necessary to use the name from the familiarisation phase for each object, to ensure that difficulties with low-agreement stimuli were difficulties in choosing among lexical items, and not difficulties in visual object recognition (similarly to Hartsuiker & Notebaert, 2010). Participants were told that their descriptions would be played to listeners who had to fill in an empty network, which only showed the position of the objects. Subsequently, three practice networks were run. The first network was described by the experimenter (to illustrate the task) and the next two networks were described by the participant. During this training phase, participants were already using the chin/head rest, to adapt the apparatus for the experiment if needed. Feedback about any failure to comply with the instructions was provided when necessary. During the eye-tracking experiment, each network was preceded by a fixation cross in the upper centre of the computer screen and started with a two seconds period for visual inspection after which the dot appeared and started its path (see procedure Figure 2). Each experiment was split into two runs of ten networks to perform a recalibration halfway through the experiment. Within each run, the procedure was automatised, so that participants did not have to press any key.

Scoring and data analysis

All productions were transcribed and scored by a native Dutch speaker and checked by another native Dutch speaker. Disfluencies were noted for utterances related to paths, but we will only report analyses on utterances related to objects, because we manipulated properties of objects. However, disfluencies following object names were not included (e.g. disfluencies occurring once the object has already been named: *to the canoe. From the cal- canoe*, which represented less than 1% of all disfluencies). Disfluencies were grouped into broad categories, to ensure a sufficient amount of data within each category: repetitions (of a sound, syllable, word, or phrase), filled pauses, silent pauses, prolongations, and self-corrections (substitutions, additions, or deletions). Self-corrections were treated as one category because less than 1% of self-corrections were not substitutions. Examples are provided in Table 2. One of the transcribers first independently transcribed and scored all networks. In a subsequent phase, a second transcriber listened to all the productions and checked the transcriptions. They disagreed on 18% of

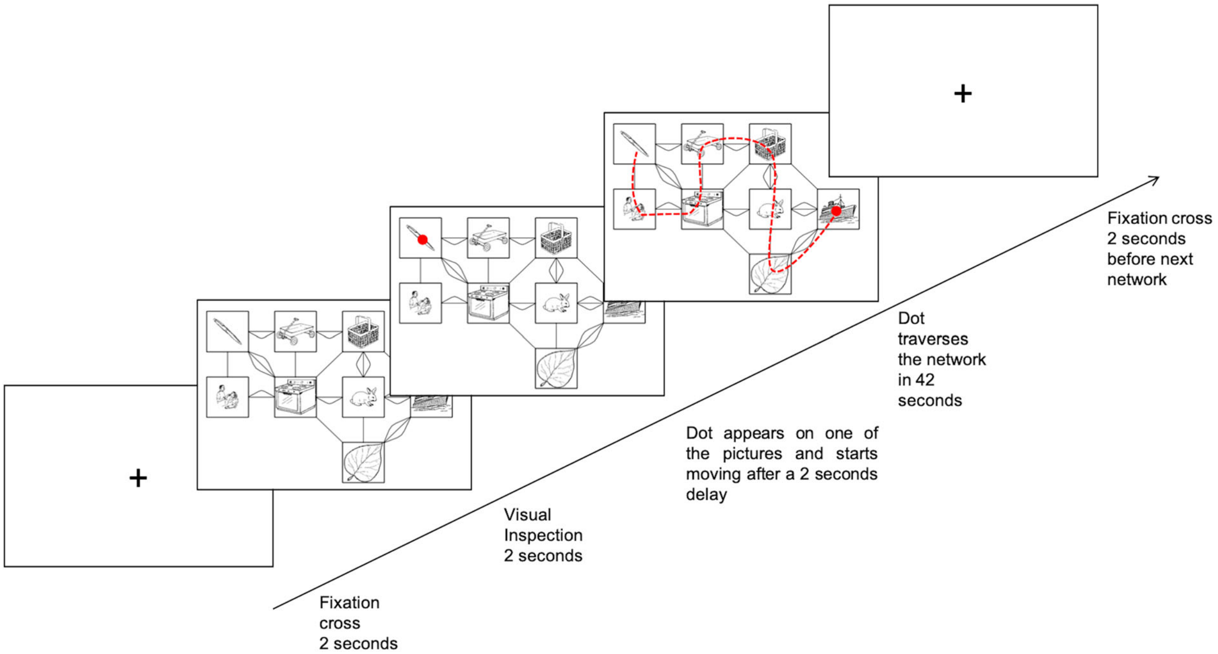

Figure 2. Procedure of each experiment. The arrow represents the time course of the experiment.

Table 2. Definitions and examples of disfluencies in each category.

Category		Definition	Example
Self-correction	Substitution	When the speaker stops and resumes with a substitution for a word.	naar het naar de [/] brief *to the [neuter-gender] to the [common gender] letter*
	Addition	when the speaker stops and resumes with the addition of new material.	naar de sla kropsla [//] *to the lettuce cabbage-lettuce*
	Deletion	When a speaker stops without completing an utterance and resumes with a new utterance.	naar de [///] dan gaat hij naar boven via *to the [///] then it goes up via*
	Other	When the speaker stops and resumes with a grammatical or lexical error.	de kano de boot [////] *de canoe de boat [////]* (when canoe was the right target)
Repetition		Repetitions of sounds, syllables, words or (part) phrases.	naar naar [r] het meisje *to to the girl*
Pause	Silent pause	When the speaker delays the speech stream by being silent.	naar (.) het meisje *to (.) the girl*
	Filled pause	When the speaker delays the speech stream by inserting a filler (e.g. uh, um)	um (h) naar het meisje *um (h) to the girl*
	Prolongation	When the speaker delays the speech stream by prolonging a speech sound.	naar (p) het meisje *to (p) the girl*

trials for Experiment 1 and 20% of the trials for Experiment 2. Disagreements were solved by a third person.

For the analysis of eye-movements, fixation positions were categorised by object Areas of Interest (AoI) corresponding to each picture. Five variables were then considered to test the effect of name agreement, grammatical gender, and disfluency on eye-movements, as described below (Table 3).

Results

Descriptive

There was at least one disfluency on 21% of trials (i.e. pictures): 2.6% included at least one self-correction, 7.3% a silent pause, 4.1% a filled pause, and 4% a prolongation. Repetitions were not analysed because there were only 11 observations in this category. Regarding eye-movements, 4.7% of pictures included at least one anticipatory fixation and 12.9% at least one late fixation. The average tracking ratio over the entire task was 95.6% (min: 88.8%; max: 99.4%).

Table 3. Description of eye-movements variables.

Variable	Definition/Description	Formula
Onset-EVS	Latency between the start of the first gaze at the picture and the onset of its name.	[first gaze onset time – speech onset time]
Offset-EVS	Latency between the end of the last gaze at the picture and the offset of its name.	[speech offset time – last gaze onset time]
Number of fixations	Number of fixations occurring from the first to the last gaze on the picture.	[First + N gazes]
Number of anticipatory fixations	Number of fixations that occur on the picture before the dot traversed it.	Naming a picture while gazing at an upcoming one
Number of late fixations	Number of fixations that occur on the picture after the dot traversed it.	Naming a picture while gazing at a previous one

Disfluency

All disfluencies: The effects of Name Agreement, Gender, and their interaction were tested for disfluency (all phenomena together) using linear mixed effects models by means of the lme4 package (Bates et al., 2015) in R (version 3.6.1). For the random part of the model, the maximal random effects structure (Barr et al., 2013) was included. We then chose a backward-selection heuristic. We used the likelihood ratio test criterion by reducing the model complexity until a further reduction would imply a significant loss in the goodness-of-fit (Matuschek et al., 2017). For the measure of disfluency, this resulted in a random intercept for network order and image order, a random slope for agreement and gender over subjects, and a random slope for agreement, gender, agreement*gender over items. The fixed part consisted of the agreement*gender interaction. There was a significant effect of name agreement (χ^2 (1) = 9.9, $p < .01$, Figure 3) indicating more disfluencies before low name agreement; a significant effect of gender (χ^2 (1) = 10.87, $p < .001$, Figure 3) indicating more disfluencies before common gender, and an agreement*gender interaction (χ^2 (1) = 6.63, $p = .01$, Figure 3), indicating more disfluency for common gender when name agreement is low.

Per disfluency category: Generalised linear mixed effects models tested for effects of name agreement, gender, and their interaction for each disfluency type (binomial) separately, using the same methods as described above (Matuschek et al., 2017). Self-corrections and filled pauses were tested with a random intercept for item and subject. There were more *self-corrections* before low name agreement pictures (χ^2 (1) = 4.91, $p < .05$). There was a trend suggesting more *filled pauses* before low name agreement pictures (χ^2 (1) = 3.58, $p < .06$). There were also trends regarding

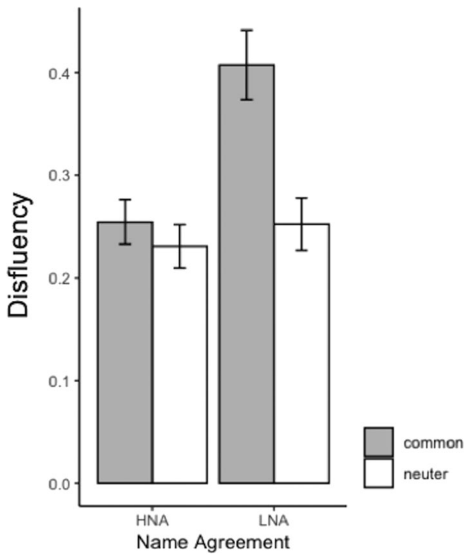

Figure 3. Proportion of total disfluency depending on name agreement and gender. Error bars show the standard error of the mean. HNA: High Name Agreement; LNA: Low Name Agreement.

gender (χ^2 (1) = 3.82, $p < .06$) and name agreement*gender (χ^2 (1) = 3.54, $p < .06$), suggesting more filled pauses before common gender pictures, and common gender pictures with low name agreement. *Silent pauses* were tested with a random intercept for network order, a random slope for agreement and gender over items, and a random slope for agreement over subjects. This model also showed a significant effect of name agreement (χ^2 (1) = 6.25, $p = .01$), with more silent pauses before low name agreement. *Prolongations* were tested with a random intercept for item, subject, and image order. There was a significant effect of gender, with more prolongations before common gender (χ^2 (1) = 10.64, $p < .01$). In sum, low name agreement induced self-corrections and silent pauses, whereas common gender induced an increase of prolongations.

Eye-movements

The same method as the one used for disfluency was chosen (Matuschek et al., 2017). After examination of skewness and kurtosis, onset-EVS and offset-EVS were log-transformed to normalise their distribution. Onset-EVS and number of fixations were tested with a random intercept for network order, image order, and items, and a random slope for agreement over subjects. There were significant effects of name agreement (χ^2 (1) = 10.71, $p < .01$) and gender (χ^2 (1) = 5.13, $p < .05$) on onset-EVS, indicating longer onset-EVS before low name agreement and common gender pictures (Figure 4). In the current study, Onset-EVS was measured in milliseconds (similarly to Meyer et al., 2012, for example). Appendix 1 provides another normalisation of this measure (Coco & Keller, 2015), which corroborates current effects. There were no effects on the number of fixations (see Appendix 2). Offset-EVS was tested with a random intercept for network order, subjects, and items and showed a significant effect of (low) name agreement (χ^2 (1) = 11.11, $p < .001$) but no effect of gender (χ^2 (1) = 0.95, $p = .3$) or agreement*gender (χ^2 (1) = 0.04, $p = .8$). The presence of anticipatory or late fixations was tested with a random intercept for image order, subjects, and items. There were no significant effects of name agreement, gender, or their interaction (see Appendix 2).

Effect of network configuration

Because this is to the best of our knowledge the first study to combine eye-tracking with a network task, we explored whether network configuration (i.e. the length of the line preceding the picture) had an influence on eye-movements and disfluencies related to that picture. Indeed, 24.1% of lines were long lines in this experiment. Each model previously used for each variable was therefore compared with a model including the length of the preceding line as a fixed effect (i.e. short or long line).

Disfluencies: There was a significant effect of line length on the production of disfluencies (χ^2 (1) = 8.02, $p < .01$): there were more disfluencies with longer lines. Of the trials that were preceded by a long line, 25.1% had at least one disfluency, compared to 19.7% of trials preceded by a short line. More specifically, this effect was significant for silent pauses (χ^2(1) = 3.99, $p < .05$) and prolongations (χ^2 (1) = 7.34, $p < .01$), both of which showed an increase when pictures were preceded by long lines. The effect of line length on filled pauses (χ^2 (1) = 0.77, $p = 0.8$) and self-corrections (χ^2 (1) = 0.37, $p = 0.5$) was not significant.

Eye-movements: There was a significant effect of line length on the onset-EVS (χ^2 (1) = 6.00, $p < .05$); offset-EVS (χ^2 (1) = 34.91, $p < .001$); and on the number of fixations (χ^2 (1) = 39.13, $p < .001$), indicating shorter latencies and fewer fixations on pictures when they were preceded by a long line.

Links between disfluency and eye-movements

In a further set of analyses, we tested whether disfluency predicts particular eye-movements (regardless of the manipulations of name agreement and gender). That is why we took the presence vs. absence of each disfluency phenomenon as a fixed effect and took the various eye-movement measures as dependent variables. For that purpose, we excluded trials where more than one disfluency was produced. As shown in Table 4, the onset-

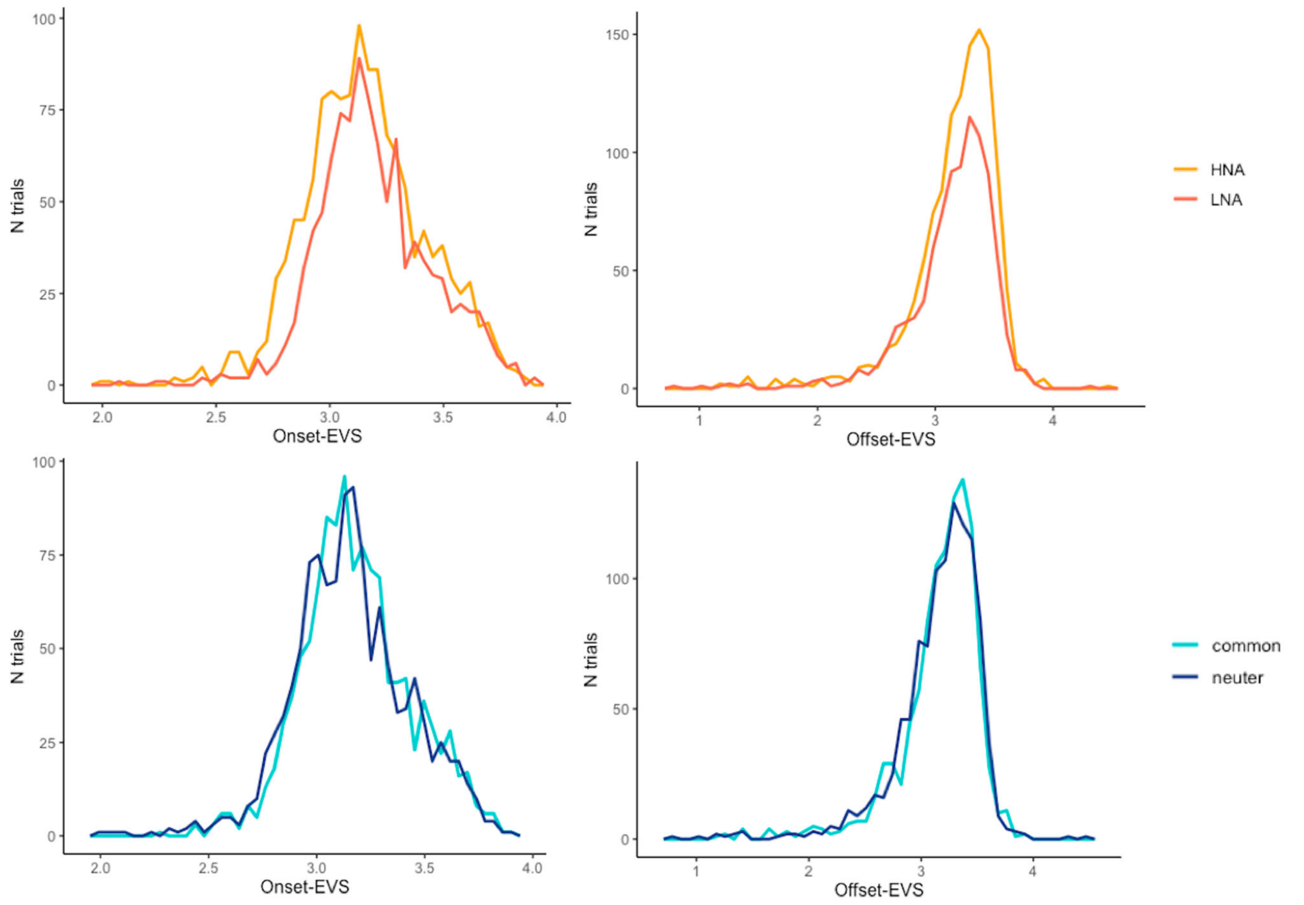

Figure 4. Onset-EVS and Offset-EVS depending on name agreement and gender. HNA: high name agreement; LNA: low name agreement; common: common gender; neuter: neuter gender.

EVS was longer when a self-correction, a silent pause, or a filled pause was produced. Regarding late fixations, participants produced significantly more late fixations towards a picture when this picture induced a self-correction or a filled pause. Offset-EVS, number of fixations and the presence of anticipatory fixations were not related to the production of disfluencies related to the picture.

Multivariate pattern analyses of disfluency and eye-movements

Finally, to investigate whether the item a participant was naming (i.e. its name agreement or gender) could be identified based on the pattern of eye movement or disfluency, we performed multivariate pattern classification, using the Scikit-learn toolbox (Pedregosa et al., 2011). Classifiers were trained for each participant to identify whether s/he was about to mention a low or high name agreement item, a common gender, or neuter gender item. We trained a linear discriminant analysis (LDA) classifier on four disfluency features (i.e. self-corrections, silent pauses, filled pauses, prolongations) and five eye-movement features (i.e. onset-EVS; offset-EVS; number of fixations, anticipatory fixations; late fixations). In all MVPA analyses, features were normalised into Z-scores. The classification was performed in a leave-one-out cross-validation approach to ensure unbiased evaluation of classification performance: In a cross-validation fold, the classifier was trained on data from all but one trial and used on the left-out trial to predict its class membership. This procedure was repeated until each trial's class has been predicted. Accuracy was the proportion of correctly classified trials. Classification accuracies for each analysis were compared to chance level, that is 50% for a two-class problem, using a one-tailed T-test. To determine which features played a significant role at a group level, we then performed one-sample t-tests on each feature's contribution in the classification (Haufe et al., 2014). Furthermore, we compared classification accuracies across analyses (i.e. using eye-movements or disfluency) to estimate which data were the most informative to classify the different conditions.

Name agreement: For name agreement, classification accuracies were significantly above chance level (50%)

Table 4. Summary of the effects of each disfluency on eye-movements.

Variable	Random structure	Results
Onset-EVS	random intercept for item, subject, image order	**self-corrections (χ^2 (1) = 7.99 *p* < .01)** **silent pauses (χ^2 (1) = 5.33, *p* < .05)** **filled pauses (χ^2 (1) = 14.64, *p* < .001)** prolongations (χ^2 (1) = 0.58, *p* = .4)
Offset-EVS	random intercept for item, subject, image order	self-corrections (χ^2 (1) = 3.37, *p* < .06) silent pauses (χ^2 (1) = 2.75, *p* = .1) prolongations (χ^2 (1) = 2.9, *p* = .1) filled pauses (χ^2 (1) = 2.4, *p* = .1)
Number of fixations	random intercept for item, subject, image order	self-corrections (χ^2 (1) = 1.55, *p* = .2) filled pauses (χ^2 (1) = 0.07, *p* = .8)
	random intercept for item, network order, image order random slope for silent pauses/prolongations over subjects	silent pauses (χ^2 (1) = 0.36, *p* = .5) prolongations (χ^2 (1) = 0.9, *p* = .4)
Anticipatory fixations	random intercept for subjects, item, image order	filled pauses (χ^2 (1) = 0.69, *p* = .4) silent pauses (χ^2 (1) = 2.3, *p* = .1)
	random intercept for subjects, image order random slope for prolongations/self-corrections over items	prolongations (χ^2 (1) = 0.12, *p* = .7) self-corrections (χ^2 (1) = 0.45, *p* = .5)
Late fixations	random intercept for subjects, item, image order	silent pauses (χ^2 (1) = 0.07, *p* = .8) prolongations (χ^2 (1) = 1.8, *p* = 2)
	random intercept for subjects, network order, image order random slope for self-corrections over items	**self-corrections (χ^2 (1) = 8.1, *p* < .01)**
	random intercept for network order, image order random slope for hesitations over items and subjects	**filled pauses (χ^2 (1) = 7.4, *p* < .01)**

Note: Significant results are in bold.

when analysing disfluencies (57.73% on average; *t*(18) = 13.32, *p* < .001). Silent pauses were the only feature that was consistent across participants for this classification (*t*(18) = 3.55, *p* < .01). Classification accuracies were also above chance when analysing eye-movements (55.18% on average; *t*(18) = 4.89, *p* < .001). The contribution of onset-EVS (*t*(18) = 3.4, *p* < .01) and offset-EVS (*t*(18) = −3.25, *p* < .01) were consistent across participants for this classification. Classification based on disfluency was significantly better than the one based on eye-movements (*t*(28) = −2.12, *p* < .05). These results showed that name agreement can be decoded based on eye-movement or disfluency features, and that disfluencies were more informative (see Figure 5A).

Gender: For gender, classification accuracies were not above chance level, neither when using disfluencies (51.57% on average; *t*(18) = 1.62, *p* = .08) nor with eye-movements (50.09% on average; *t*(18) = 0.08, *p* = .5), which means that gender cannot be predicted based on the pattern of eye-movements or disfluency (see Figure 5B).

Discussion

This experiment replicated the finding that pictures with low name agreement names induce more disfluencies in general, and more self-corrections and silent pauses in particular (Hartsuiker & Notebaert, 2010). However, contrary to the previous study, pictures with a neuter gender name did not elicit more disfluencies. We showed rather that pictures with a common gender name were associated with more prolongations. This

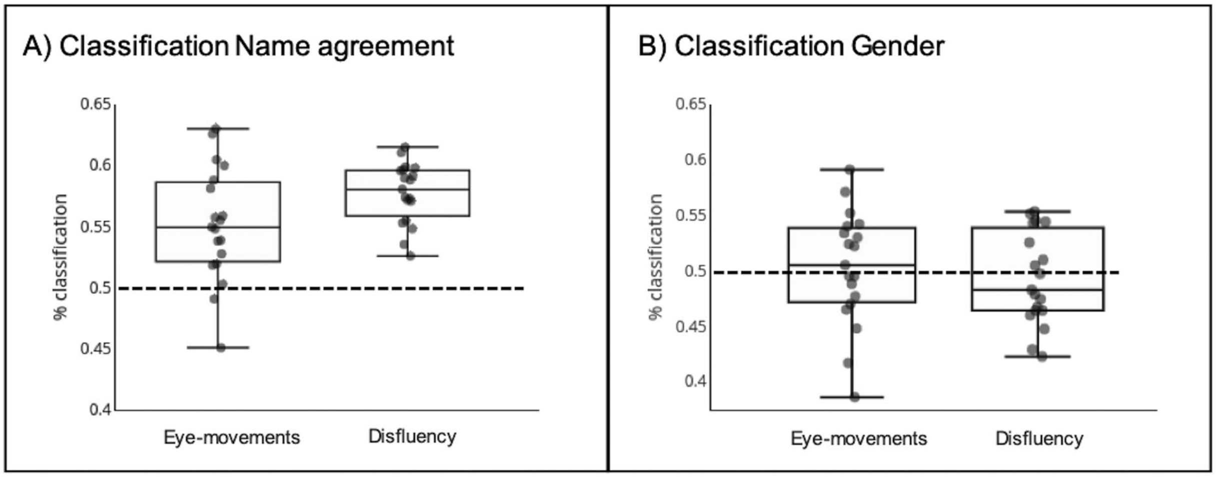

Figure 5. Classification accuracies for each participant for identifying (A) Name agreement, (B) Gender, of the items based on eye-movements or disfluency. The dashed line represents chance level. Each dot represents classification accuracy for a single participant.

unexpected finding might be due to the distribution of each gender in our set of stimuli. Indeed, because we explicitly manipulated this variable, the experiment had the same number of items with common gender and neuter gender. On the other hand, Hartsuiker and Notebaert (2010) presented more pictures with common-gender names than with neuter-gender names, following the distribution of common and neuter gender in the Dutch language. One possibility is that, because of the relatively large proportion of neuter-gender items, participants' attention towards these items increased, which resulted in a reduction of disfluencies associated with this gender. Although this paradoxical effect of gender on disfluency cannot be fully explained, it at least suggests that difficulties at distinct stages of production lead to distinct patterns of disfluencies: difficulties in determiner selection led to a different pattern of disfluencies than difficulties in content word selection.

Apart from disfluency, the current study also examined eye-movements associated with a network task. Few studies analysed eye-movements during linguistic difficulties, and this study is the first to examine longer samples than single utterances. We observed that onset-EVS increased with lexical selection difficulty, thereby replicating and extending the results of studies based on single sentence production (e.g. Griffin, 2001). Thus, it is also possible to observe effects of word preparation difficulty in eye-movements during connected-speech production. However, we did not find a typical offset-EVS in the current study: The participants were usually still gazing at the picture while naming it and even continued gazing for more than 1.5 sec on average after picture naming had been completed. In contrast, previous literature showed that participants do not look at items anymore while articulating their names. Instead, they typically gaze at the next object to be mentioned (Griffin, 2004; Meyer et al., 1998). This difference may be related to the presence of the dot that paced the speech: Participants therefore needed to wait for the dot to catch up if they had rapidly named the picture. We suspect this is why the pattern of offset-EVS had the opposite direction from the onset-EVS (i.e. shorter for low name agreement). The use of the dot also explains why participants did not have as many anticipatory/late fixations as we expected in this type of paradigm (i.e. 4.7% of the trials for anticipatory fixations; 12.9% of the trials for late fixations).

Because this is the first study to combine eye-tracking with a network task, we explored the effect of network configuration on the pattern of eye-movements and disfluency. The length of the line preceding an item had an influence on both types of measure. Participants spent less time gazing at a picture when it was preceded by a long line (when analysing onset-EVS, offset-EVS, and number of fixations), and at the same time they produced more pauses (silent pauses and prolongations). This finding might imply that participants used the extra time they had available when the dot traversed a long line to inspect other areas than the upcoming picture, which led to pauses. To test this assumption, we exploratory considered anticipatory and late fixations produced while the dot was traversing a line ($N = 87$). We counted, for each participant, the proportion of these fixations that was produced during short versus long lines. On average, 39.5% were anticipatory fixations and 45% were late fixations produced while the dot was traversing a long line (versus 13.2% anticipatory fixations and 2.3% late fixations produced while the dot was traversing a short line), which reinforces this hypothesis. It therefore means that disfluency can sometimes be used as a "stalling strategy" and is controlled in part by top down processes (Brown-Schmidt & Tanenhaus, 2006; Clark & Fox Tree, 2002). Although current effects do not involve filled pauses, contrary to Clark and Fox Tree's view, other authors showed that silent pauses could also reflect speaker's strategies at a discourse level rather than speech encoding processes (Pistono et al., 2016; Pistono et al., 2019). This finding also indicates that studies focusing on disfluency using this paradigm need to control the configuration of the network, as it influences the use of pauses.

In sum, we showed that lexical selection difficulty and common-gender names induce longer onset-EVS while the number of fixations as well as the presence of anticipatory or late fixations did not vary with these manipulations. To have a fuller understanding of the production of disfluency and eye-movements, we also examined the effect of each disfluency on eye-movement variables. By doing so, we observed that the production of self-corrections, silent pauses, or filled pauses induced longer onset-EVS. This supports the idea that the time spent gazing at an object prior to naming it reflects linguistic difficulty, but is not indicative of a specific type of difficulty. Indeed, previous studies have shown that this latency increases with several difficulties (e.g. lexical selection or phonological encoding: Griffin, 2001). We also showed that participants produced significantly more late fixations towards a picture when it was previously associated with a self-correction or a filled pause. Given the hypothesis that extra time spent on an item reflects self-monitoring processes, the current results argue for a monitoring account of filled pauses and self-corrections.

Finally, using multivariate pattern analyses we successfully classified low name agreement trials from high name agreement trials, based on the pattern of disfluency or eye-movements associated with it. In particular, classification accuracy was higher when analysing disfluency, which suggests that disfluencies are a more informative and reliable feature to decode name agreement. It is important to mention, however, that only silent pauses contributed consistently to classier performance across participants. This means that, although there was a significant effect of name agreement on self-corrections in linear mixed models, this disfluency was not consistently informative across participants, Items' gender could not be predicted based on multivariate analyses of eye-movements or disfluency. In other words, grammatical selection difficulty could not be decoded based on the information contained in eye-movements or disfluency, which means that results found with linear mixed models were not consistently informative across participants. In sum, the information patterns in eye-movement and disfluency data allow for reliable classification of name agreement but not of gender.

This experiment revealed that lexical and grammatical selection elicit a specific pattern of disfluency and eye-movements. In particular, the presence of lexical selection difficulty can be inferred based on the pattern of eye-movements or disfluency. Moreover, the use of eye-tracking revealed a strong connection between disfluency and gaze before speech onset since disfluencies are predictive of longer gaze before onset. It also revealed the presence of anticipatory and late fixations during the description of a network. These latter phenomena were associated with the production of self-corrections and filled pauses. This suggests that some disfluencies and eye-movements reflect late self-monitoring processes, rather than difficulties in speech encoding. However, given the influence of line length on these results, it is important to control whether they will be consistent when length is held constant. To do so, we conducted a second experiment where only short lines were used.

Experiment 2

Participants

Twenty further bachelor students (5 males and 15 females), all native speakers of Dutch, volunteered to take part in the experiment. Mean age was 19.5 ± 1 years old.

Material and methods

The material and methods were identical to Experiment 1, except that networks were controlled to only have short lines (Figure 1).

Results

Descriptive

We excluded 685 trials (21.4% of pictures) because the wrong target was produced or the gender was omitted. There was at least one disfluency on 16.3% of the included trials. More precisely, 3.7% of the trials included at least one self-correction, 5.6% a silent pause, 4.5% a filled pause, and 5.5% a prolongation. Repetitions were not analysed because there were only 10 observations in this category. Regarding eye-movements, 3.6% of pictures induced at least one anticipatory fixation and 15.1% at least one late fixation. The average tracking ratio was 92.8% (min: 80.9%; max: 99.3%).

Disfluency

All disfluencies: The effects of name agreement, gender, and their interaction were tested with a random intercept for subjects and items. The fixed part consisted of the agreement*gender interaction. There was a significant main effect of name agreement (χ^2 (1) = 10.88, $p < .001$, Figure 6) indicating more disfluencies before low name agreement; a significant effect of gender (χ^2 (1) = 11.48, $p < .001$, Figure 6) indicating more disfluencies before common gender, and an agreement*-gender interaction (χ^2 (1) = 6.51, $p < .05$, Figure 6), indicating that, in the low name agreement condition, participants produced more disfluency before common gender.

Per disfluency category: Generalised linear mixed effects model tested agreement* gender on each

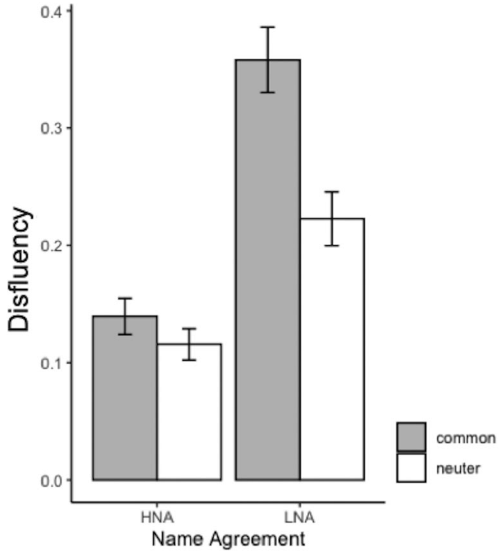

Figure 6. Proportion of total disfluency depending on name agreement and gender. Error bars show the standard error of the mean. HNA: High Name Agreement; LNA: Low Name Agreement.

phenomenon (binomial). Self-corrections and prolongations were tested with a random intercept for network order and image order, a random slope for agreement over subjects, and a random slope for gender over items. There was a significant effect of (low) name agreement on *self-corrections* (χ^2 (1) = 10.19, $p < .01$), whereas the effects of gender (χ^2 (1) = 0.56, $p = .5$) and agreement*gender (χ^2 (1) = 0.19, $p = .7$) were not significant. There were significantly more *prolongations* before common gender nouns (χ^2 (1) = 16.21, $p < .0001$). There was a trend regarding the effect of name agreement (χ^2 (1) = 3.6, $p < .06$), suggesting more prolongations before low name agreement. The agreement*gender interaction was not significant (χ^2 (1) = 0.92, $p = .3$). *Filled pauses* were tested with a random intercept for subjects and items. The model also showed an effect of name agreement (χ^2 (1) = 11.32, $p < .001$), indicating more filled pauses before low name agreement, while other effects were not significant (gender: (χ^2 (1) = 1.54, $p = .2$); agreement*gender: (χ^2 (1) = 1.48, $p = .2$)). *Silent pauses* were tested with a random slope for agreement, gender, and agreement*gender over subjects and items, and a random slope for agreement over network order and image order. This resulted in a significant effect of (low) name agreement (χ^2 (1) = 17.59, $p < .0001$) while the effects of gender (χ^2 (1) = 1.52, $p = .2$) and agreement*gender (χ^2 (1) = 0.02, $p = .9$) were not significant. In sum, items with low name agreement were accompanied by more self-corrections, prolongations (albeit only marginal), filled pauses and silent pauses. Gender only affected prolongations, and gender and agreement never interacted.

Eye-movements

As for Experiment 1, *onset-EVS* was log-transformed. This measure was tested with a random intercept for image order, a random slope for gender over network order, and a random slope for agreement and gender over subjects and items. This resulted in a significant effect of name agreement (χ^2 (1) = 15.53, $p < .0001$, Figure 7). Other effects were not significant (gender: (χ^2 (1) = 1.45, $p = .2$); agreement*gender: (χ^2 (1) = 0.05, $p = .8$)). *Offset-EVS* and *anticipatory fixations* were tested with a random intercept for items, subjects, and image order. The presence of *anticipatory fixations* was not significant affected by any variable (see Appendix 2). *Offset-EVS* was significantly longer

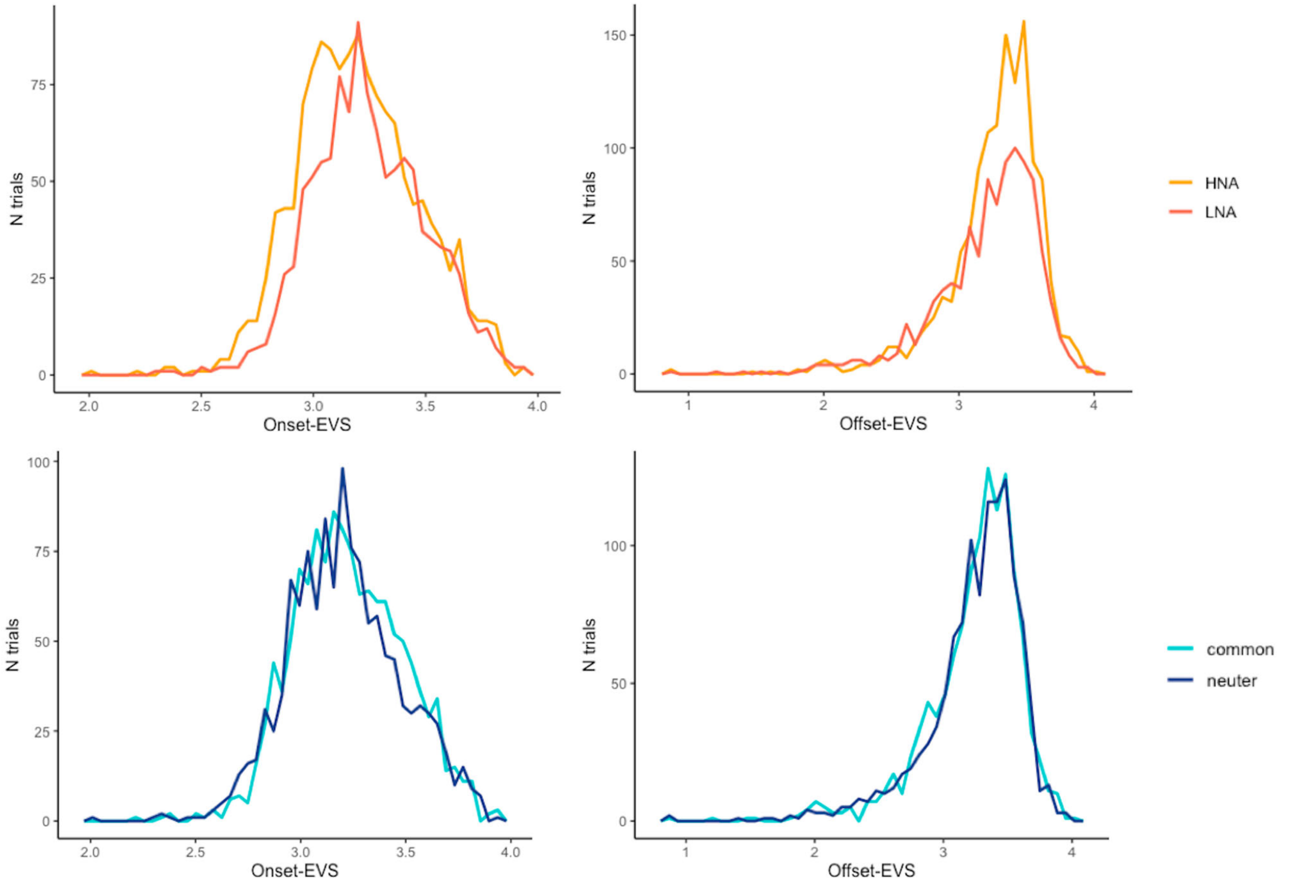

Figure 7. Onset-EVS and Offset-EVS depending on name agreement and Gender. HNA: high name agreement; LNA: low name agreement; common: common gender; neuter: neuter gender.

after high name agreement (χ^2 (1) = 15.59, $p < .001$, Figure 7). Other effects were not significant (gender: (χ^2 (1) = 0, $p = .99$); agreement*gender: (χ^2 (1) = 0.43, $p = .5$)). *The number of fixations* was tested with a random intercept for image order and network order, and a random slope for agreement and gender over subjects and items. The effect of name agreement was significant, indicating more fixations on low name agreement pictures (χ^2 (1) = 5.11, p. < .05). Other effects were not significant (gender: (χ^2 (1) = 2.17, $p = .1$); agreement*gender: (χ^2 (1) = 0.36, $p = .5$)). The presence of *late fixations* was tested with a random intercept for subjects, items, image order, and network order. There were no significant effects of name agreement, gender, or agreement*gender (see Appendix 2 for detailed results).

Links between disfluency and eye-movements

As in Experiment 1, we tested the presence vs. absence of each disfluency phenomenon as a fixed effect and took the various eye-movement measures as dependent variables. Results are presented in Table 5. There was a significant main effect of each disfluency on onset-EVS, indicating longer prior gazes when a disfluency was produced. There was also a significant main effect of all phenomena but prolongations on offset-EVS, indicating shorter offset-EVS when a self-correction, a silent pause, or a filled pause was produced. Disfluency had no effect on the number of fixations or on the presence of anticipatory fixations. There was a significant effect of filled pauses on late fixations: the presence of late fixations towards a picture increased when this picture induced a filled pause. In other words, participants gazed back at a picture more often when they hesitated before naming this picture.

Table 5. Summary of the effects of each disfluency on eye-movements.

Variable	Random structure	Results
Gaze-onset-to-name-onset interval	random intercept for item, subject, image order	**self-corrections (χ^2 (1) = 17.74, $p < .0001$)** **silent pauses (χ^2 (1) = 9.85, $p < .01$)** **prolongations (χ^2 (1) = 4.54, $p < .05$)**
	random intercept for item, subject, network order, image order	**filled pauses (χ^2 (1) = 8.03, $p < .01$)**
Name-offset-to-gaze-offset interval	random intercept for item, subject, image order	**self-corrections (χ^2 (1) = 14.94, $p < .001$)** **silent pauses (χ^2 (1) = 7.31, $p < .01$)** prolongations (χ^2 (1) = 0.3, $p = .6$) **filled pauses (χ^2 (1) = 17.65, $p < .0001$)**
Number of fixations	random intercept for item, subject, network order, image order	filled pauses (χ^2 (1) = 2.83, $p = .1$) silent pauses (χ^2 (1) = 1.04, $p = .3$) prolongations (χ^2 (1) = 1.78, $p = .2$)
	random intercept for item, subject, image order	self-corrections (χ^2 (1) = 0.08, $p = .8$)
Anticipatory fixations	random intercept for subjects, items, image order	self-corrections (χ^2 (1) = 0.61, $p = .4$) filled pauses (χ^2 (1) = 0.62, $p = .4$) silent pauses (χ^2 (1) = 0.01, $p = 1$) prolongations (χ^2 (1) = 1.05, $p = .3$)
Late fixations	random intercept for subjects, items, network order, image order	self-corrections (χ^2 (1) = 0.49, $p = .5$) silent pauses (χ^2 (1) = 2.16, $p = .1$) prolongations (χ^2 (1) = 0.24, $p = .6$)
	random intercept for network order, image order random slope for filled pauses over items and subjects	**filled pauses (χ^2 (1) = 10.58, $p < .0.01$)**

Note: Significant results are in bold.

Multivariate pattern analyses of disfluency and eye-movements

Following the same method as the one developed in Experiment 1, classifiers were trained for each participant to identify whether the participant was about to mention a low or high name agreement item, a common gender, or neuter gender item. We again trained a linear discriminant analysis (LDA) classifier on four disfluency features (i.e. self-corrections, silent pauses, filled pauses, prolongations) and five eye-movement features (i.e. onset-EVS, offset-EVS, number of fixations, anticipatory fixations, late fixations).

Name agreement: For name agreement, classification accuracy using disfluency was significantly above chance (59.57% on average; $t(19) = 7.04$, $p < .001$). The contribution of each feature was consistent across participants for this classification (self-corrections: ($t(19) = 3.6$, $p < .01$); silent pauses: ($t(19) = 6.5$, $p < .0001$); prolongations: ($t(19) = 3.2$, $p < .01$); filled pauses: $t(19) = 3.5$, $p < .01$). Classification accuracies were also above chance when analysing eye-movements (57.89% on average; $t(19) = 7.29$, $p < .001$). Onset-EVS ($t(19) = 4.04$, $p < .0001$) and offset-EVS ($t(19) = -2.11$, $p < .05$) were consistent features for this classification. Classification based on disfluency was not significantly better than the one based on eye-movements ($t(38) = -0.5$, $p = .6$), which means that eye-movement and disfluency are equally informative relative to name agreement (see Figure 8A). Compared to Experiment 1, classification based on disfluency was not significantly higher ($t(25) = -1.25$, $p = .2$). However, classification based on eye-movements was ($t(20) = 2.1$, $p < .05$).

Gender: For gender, classification accuracies were above chance level when analysing disfluencies

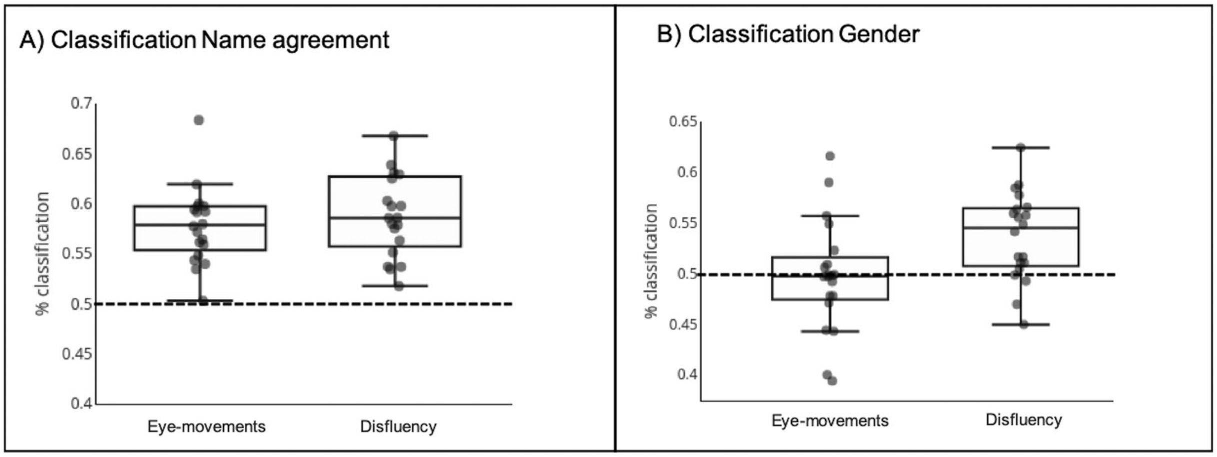

Figure 8. Classification accuracies for each participant for identifying (A) Name agreement, (B) Gender, of the items based on eye-movements or disfluency. The dashed line represents chance level. Each dot represents classification accuracy for a single participant.

(53.72% on average; $t(19) = 3.83$, $p = .001$, Figure 8B) but not when analysing eye-movements (49.41% on average; $t(19) = -.41$, $p = .3$). Only the contribution of prolongations was consistent across participants ($t(19) = -2.5$, $p < .05$). Compared to Experiment 1, classification based on disfluency significantly increased ($t(37) = -3$, $p < .01$) while classification based on eye-movements did not ($t(19) = 0.79$, $p = .4$).

Discussion

Despite the presence of only short lines, this experiment elicited fewer disfluencies than Experiment 1 overall (i.e. 16.3% instead of 21%). Nevertheless, we replicated most results: pictures with low name agreement again led to more self-corrections and silent pauses than pictures with high name agreement, similarly to Experiment 1 and Hartsuiker and Notebaert (2010) (but also to more filled pauses, in contrast to Experiment 1 that only showed a trend). The effect of gender was also similar to the one found in the first experiment: items with a common gender name induced more disfluency (i.e. prolongations), contrary to what was expected on the basis of Hartsuiker and Notebaert's study. In short, low name agreement induced all types of disfluency except for prolongations, and prolongations were the only type of disfluency showing a significant effect of gender. As mentioned above, the findings suggest that difficulties at distinct stages of production lead to distinct patterns of disfluencies.

Concerning eye-movements, the use of a consistent type of configuration (i.e. short lines connecting the pictures only) led to clearer findings. All significant variables showed a similar effect of name agreement, but no effect of gender or agreement*gender. Low name agreement elicited longer onset-EVS, shorter offset-EVS, and more fixations towards the item, in line with our predictions. Contrary to Experiment 1, the effect of gender on eye-movements was not significant, which suggests that the act of determiner selection is not as demanding as the act of lexical selection. Finally, the presence of anticipatory or late fixations was not influenced by name agreement or gender in either experiment.

The use of a consistent configuration also led to better classification accuracies, both when decoding items' name agreement or gender. In particular, the classifier could predict name agreement above chance level for all participants, when analysing the pattern of eye-movements or disfluency. For disfluency patterns, all features had a significant contribution. This means that the classifier found information in each feature, and that their contribution was sufficiently consistent across participants. For eye-movements patterns, onset-EVS and offset-EVS had a consistent contribution to classifications, while the number of fixations was not significant. This means that EVS is more informative across participants to predict name agreement than the number of fixations on the picture (although number of fixations was affected by name agreement in the linear mixed models). Regarding gender, the classifier could predict from the pattern of disfluency whether participants were about to name items with common gender or neuter gender. This feature was also the only one that was consistent across participants. This reinforced findings from linear mixed models and implies that, compared to Experiment 1, patterns are more stable across participants. On the contrary, the classifier could not predict from the pattern of eye-movements whether participants were about to name items with common gender or neuter gender, which reinforces the conclusions from linear mixed modelling: gender selection cannot be distinguished on the basis of eye-movements.

Onset-EVS was predicted by different types of disfluencies, similarly to Experiment 1. However, disfluencies also elicited shorter offset-EVS. Contrary to the first experiment, and because of the controlled configuration of the networks, participants probably had to shift their gaze earlier to stay synchronised with the pace of the dot after the production of a disfluency. The presence of late fixations was associated with the production of filled pauses only. Although it is unclear why the effect of self-corrections was not significant, it argues for a common origin of filled pauses and self-corrections, in relation with verbal monitoring.

General discussion

This study combined eye-movements monitoring with a network task paradigm to evaluate the effects of difficulties in isolated levels of language production on disfluencies and eye-movements. It replicated and extended previous findings on the effects of lexical and grammatical selection on disfluency and eye-movements. Additionally, eye-tracking proved to be informative about the underlying mechanisms of disfluency, beyond difficulties related to speech encoding. We will now discuss the different research questions that we outlined in the introduction, and their theoretical implications.

(i) Which pattern of disfluency occurs depending on lexical selection difficulty and grammatical selection difficulty?

We replicated the effects of low name agreement on disfluencies (Hartsuiker & Notebaert, 2010) in two experiments. We also found an effect of gender twice. More precisely, we replicated the finding that difficulties in the initial stage of lexical access elicits self-corrections and pauses in both our experiments, in line with our hypothesis. The finding that difficulties in finding a name of a picture promote pauses is compatible with the claim that pauses reflect an "act of choice" between lexical items with similar semantic features (e.g. Beattie & Butterworth, 1979). Because of this "act of choice", speakers are also more error-prone, leading to the production of self-corrections.

However, neuter gender did not elicit more disfluency than common gender. Contrary to Hartsuiker and Notebaert (2010), and because we explicitly manipulated grammatical selection, each network had four items with common gender names, and four items with neuter gender names. This has the disadvantage that the distribution of determiners for target picture names does not follow the distribution in the Dutch language. It is therefore possible that participants created opposite expectations (i.e. the expectation that the next determiner to be produced is more likely to have neuter than common gender). Additionally, disfluencies elicited by common gender items were mainly prolongations. Possibly, this effect is related to the phonological form of the common gender determiner ("de" in opposition to the neuter gender determiner "het"), which is more likely to encourage prolongations because it ends in a vowel rather than a stop consonant. This manipulation had some consequences for statistical power. Because we excluded trials where the participant used a different gender marking (e.g. plural, indefinite determiner), we excluded more trials than Hartsuiker and Notebaert.

(ii) What is the pattern of eye-movements during connected-speech production, and during lexical or grammatical selection?

As we hypothesised, we also found an effect of name agreement on onset-EVS in both experiments, which replicates and extends earlier work. More specifically, both experiments showed that Onset-EVS increases with word preparation difficulties during connected-speech production, in line with what has been shown during picture naming (e.g. Meyer et al., 1998 with low frequent items or degraded pictures) or single sentence production (e.g. Griffin, 2001 with low frequent items and low name agreement items). As they describe networks, participants' gazes reflected the difficulty related to word selection and encoding for upcoming items. However, previous literature constantly showed that about 100–300 ms before saying an object's name, speakers shift their gaze to the next item to be named (e.g. Griffin, 2001; Meyer et al., 1998), which coincides with the end of phonological encoding. On the contrary, we found that participants were still gazing at the picture while and after naming it. In all likelihood, this finding is related to our paradigm, and specifically to the fact that speech rhythm was imposed by the pace of a dot moving through the network. While this offset-EVS was shortened when a disfluency was produced (Experiment 2), it did not vary much during fluent utterances, when participants could easily stay synchronised with the pace of the dot. In that sense, the task we used is more artificial than studies based on single word or single sentence production. Indeed, the use of a marker going through the network constrained eye-movements after picture naming and altered the eye-speech lag usually described in the literature (e.g. Griffin, 2001). It also probably limited the production of anticipatory and late fixations.

(iii) What is the relationship between disfluencies and eye-movements?

The monitoring of eye-movements during a network task also revealed a strong connection between disfluencies and eye-movements in Experiment 2. Disfluencies and Onset-EVS before speech onset both increased with lexical selection difficulty, and disfluencies were predictive of longer Onset-EVS. Filled pauses predicted the production of late fixations, suggesting that some disfluencies and eye-movements are associated with other mechanisms than difficulties in speech encoding, such as self-monitoring processes. Moreover, although we showed that eye-movements' monitoring is feasible during a network task paradigm, it is crucial to control for the configuration of pictures that have to be mentioned. In particular, the use of long lines in Experiment 1 induced more pauses and fewer eye-movements towards the item to be mentioned, which indicates the use of strategies from the participants.

(iv) Can name agreement and gender be predicted based on the pattern of eye-movements or the pattern of disfluency?

Finally, we showed that lexical selection difficulty (and to a lesser extent grammatical selection) can be decoded from the pattern of disfluency or eye-movements produced by a speaker using MVPA. This means that disfluencies and eye movements are sufficiently informative about the linguistic difficulty of an item that a classifier can learn and predict the type of item a speaker was about to mention. In previous work, Coco and Keller (2014) extracted eye-movement features from visual and language tasks (i.e. visual search, object naming, and scene description) to test whether automatic classification of these tasks was possible. They showed that the three tasks were indeed associated with distinctive eye-movement patterns, which suggests that both visual and linguistic processing influence eye-movement patterns. The current experiment shows that, even within a same task, a classification algorithm can be successful in predicting the type of information being processed, based on linguistic and visual features. Although classification accuracies in the current study were just above chance level, they provided complementary findings about lexical and grammatical selection. More precisely, while some classifications reinforced findings from linear mixed models (e.g. that prolongations are the only consistent feature predicting grammatical gender in Experiment 2), they also revealed inter-individual differences (e.g. features that were not consistent across participants). Such findings call for further research on the underlying factors shaping these particularities. These methods are also particularly interesting when examining manipulations at different levels of the language production system. Indeed, previous research has demonstrated that onset-EVS or disfluency increase with linguistic difficulty (e.g. conceptual, lexical, or phonological level). However, although mean onset-EVS or proportions of disfluency can change as a function of linguistic manipulation, the distributions of these values are likely to overlap strongly across manipulations. In that sense, the use of multivariate pattern classification can bring more information, by determining whether the type of linguistic difficulties can be inferred and disentangled from the pattern of eye-movements or disfluency (i.e. by comparing patterns related to items' age of acquisition and items' name agreement for example). This could be a further step for current models of language production, to explain when and why which disfluencies occur.

Theoretical implications and future work

Speech encoding difficulties: As previously mentioned, the current findings suggest that difficulties at distinct stages of production lead to distinct patterns of disfluencies: difficulties in determiner selection led to a different pattern of disfluencies than difficulties in content word selection. They also suggest that all disfluency types considered here can – at least partly – be related to speech encoding difficulty, since they were associated with longer onset-EVS. In particular, self-corrections did not result from rushed word preparation, similarly to Griffin (2004). Future work is required to examine the effect of difficulties at other levels of the language production system using this paradigm, to get a broader view of disfluency production. Similarly to what has been done in earlier work, the conceptual generation of a message can be manipulated using blurry images (Schnadt & Corley, 2006) and time pressure could be used to hamper the monitoring system (Oomen & Postma, 2001). Future work manipulating phonological complexity is required to uncover the pattern of disfluency associated with difficulty at the phonological level. As mentioned above, MVPA could also be a complementary technique, in addition to linguistic manipulations. It could help disentangle the pattern of disfluency and eye-movements associated with difficulties at each level.

Stalling strategies: The analyses of eye-movement variables showed that not all disfluencies are related to speech encoding. Some may be related to stalling strategies. Indeed, Experiment 1 revealed the use of strategic pauses when a picture is preceded by a long line, which is consistent with Brown-Schmidt and Tanenhaus (2006). Indeed, these authors showed that speakers sometimes make use of strategic disfluency, to buy enough time to add additional information to a planned utterance (i.e. to include the name of an area of interest they gazed at in a specific time window). In the current study, we showed that speakers can also use strategic disfluency to buy time to produce an utterance while gazing at an upcoming area of interest. These findings also support work by Ferreira and Swets (2002), who showed that speakers speak more slowly when they have less time to prepare their utterances. In other words, speakers are able to adapt their speech rate or use of disfluency, and these phenomena are not always accidental.

As previously mentioned, linguistic difficulty did not affect the frequency of anticipatory fixations in the current experiments. However, the difficulties we introduced (low name agreement and infrequent grammatical gender) were not necessarily noticeable when looking at the pictures in the network. One can imagine that such fixations will increase when manipulating visual–conceptual processing, using visually degraded pictures like Meyer et al. (1998) or Schnadt and Corley (2006). Using this type of manipulation with eye-tracking will certainly contribute to revealing the use of anticipatory fixations and strategic disfluency.

Self-monitoring: In both experiments participants sometimes produced late fixations towards a picture, in particular when the picture induced a filled pause. Following Griffin's findings, we hypothesised that additional time spent on an item – and disfluency associated with it – could reflect self-monitoring processes and uncertainty about the answer. By using manipulations that hamper speech monitoring (e.g. divided attention), we will be able to test this assumption more specifically. Indeed, following this hypothesis, uncorrected errors would induce shorter onset-EVS than repaired errors.

In conclusion, this study is the first to combine eye-tracking with a network task to evaluate the effects of difficulties in isolated levels of language production on disfluencies and eye-movements. It showed that difficulties in the initial stage of lexical access result in a specific pattern of disfluency and eye-movements, whereas the effect of grammatical selection is less clear. Further work is required to analyse more precisely what patterns of disfluencies are associated with difficulties at other levels of language production.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Individual fellowship, grant agreement No 832298. Thanks to Mehdi Senoussi for his help with programming.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by H2020 Marie Sklodowska-Curie Actions: [grant number Individual fellowship/No 832298].

ORCID

Aurélie Pistono http://orcid.org/0000-0001-7363-5232
Robert J. Hartsuiker http://orcid.org/0000-0002-3680-6765

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278, https://doi.org/10.1016/j.jml.2012.11.001

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *1*(67), 1–48. https://doi.org/10.18637/jss.v067.i01

Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaenous speech. *Language and Speech*, *22* (39), 201–211. https://doi.org/10.1017/CBO9781107415324.004

Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*, *54*(4), 592–609. https://doi.org/10.1016/j.jml.2005.12.008

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 1–20. https://doi.org/10.5334/joc.10

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Coco, M. I., & Keller, F. (2014). Classification of visual and linguistic tasks using eye-movement features. *Journal of Vision*, *14*(3), 1–18. https://doi.org/10.1167/14.3.11

Coco, M. I., & Keller, F. (2015). Integrating mechanisms of visual guidance in naturalistic language production. *Cognitive Processing*, *16*(2), 131–150. https://doi.org/10.1007/s10339-014-0642-0

Craddock, R. C., Holtzheimer III, P. E., Hu, X. P., & Mayberg, H. S. (2009). Disease state prediction from resting state functional connectivity. *Magnetic Resonance in Medicine*, *62*(6), 1619–1628. https://doi.org/10.1002/mrm.22159

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dhooge, E., De Baene, W., & Hartsuiker, R. J. (2016). The mechanisms of determiner selection and its relation to lexical selection: An ERP study. *Journal of Memory and Language*, *88*, 28–38. https://doi.org/10.1016/j.jml.2015.12.004

Ferreira, F., & Rehrig, G. (2019). Linearisation during language production: Evidence from scene meaning and saliency maps. *Language, Cognition and Neuroscience*, *34*(9), 1129–1139. https://doi.org/10.1080/23273798.2019.1566562

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57–84. https://doi.org/10.1006/jmla.2001.2797

Fox Tree, J. E. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *In Journal of Memory and Language*, *34*(6), 709–738. https://doi.org/10.1006/jmla.1995.1032

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*(1), 1–16. https://doi.org/10.1016/S0010-0277(01)00138-X

Griffin, Z. M. (2004). The eyes are right when the mouth is wrong. *Psychological Science*, *15*(12), 814–821. https://doi.org/10.1111/j.0956-7976.2004.00761.x

Hartsuiker, R. J., & Notebaert, L. (2010). Lexical access problems lead to disfluencies in speech. *Experimental Psychology*, *57*(3), 169–177. https://doi.org/10.1027/1618-3169/a000021

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, *87*, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*(7), 523–534. https://doi.org/10.1038/nrn1931

Henderson, J. M., Shinkareva, S. V., Wang, J., Luke, S. G., & Olejarczyk, J. (2013). Predicting cognitive state from eye movements. *PLoS ONE*, *8*(5), 1–6. https://doi.org/10.1371/journal.pone.0064937

Levelt, W. J. M. (Willem J. M.). (1989). *Speaking: From intention to articulation*. MIT Press. https://pubman.mpdl.mpg.de/pubman/item/escidoc:67053:6

Levin, H., & Buckler-Addis, A. (1979). *The eye–voice span*. MIT Press.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, *15*(1), 19–44. https://doi.org/10.1080/00437956.1959.11659682

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*(2), B25–B33. https://doi.org/10.1016/S0010-0277(98)00009-2

Meyer, A. S., Wheeldon, L., van der Meulen, F., & Konopka, A. (2012). Effects of speech rate and practice on the allocation of visual attention in multiple object naming. *Frontiers in Psychology*, *3*(FEB), 1–13. https://doi.org/10.3389/fpsyg.2012.00039

Nozari, N., & Novick, J. (2017). Monitoring and control in language production. *Current Directions in Psychological Science*, *26*(5), 403–410. https://doi.org/10.1177/0963721417702419

Oomen, C. C., & Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self- monitoring. *Journal of Psycholinguistic Research*, *30*(2), 163–184. 11385824. https://doi.org/10.1023/A:1010377828778

Oomen, C. C. E., & Postma, A. (2002). Limitations in processing resources and speech monitoring. *Language and Cognitive Processes*, *17*(2), 163–184. https://doi.org/10.1080/01690960143000010

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://doi.org/10.5555/1953048.2078195

Peirce, J. W. (2007). PsychoPy-psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Pistono, A., Jucla, M., Barbeau, E. J., Saint-Aubert, L., Lemesle, B., Calvet, B., Köpke, B., Puel, M., & Pariente, J. (2016). Pauses during autobiographical discourse reflect episodic memory processes in early Alzheimer's disease. *Journal of Alzheimer's Disease*, *50*(3), 687–698. https://doi.org/10.3233/JAD-150408

Pistono, A., Pariente, J., Bézy, C., Lemesle, B., Le Men, J., & Jucla, M. (2019). What happens when nothing happens? An investigation of pauses as a compensatory mechanism in early Alzheimer's disease. *Neuropsychologia*, *124*, 133–143. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2018.12.018

Reilly, J., Flurie, M., & Ungrady, M. B. (2020). Eyetracking during picture naming predicts future vocabulary dropout in progressive anomia. *Neuropsychological Rehabilitation*, *28*, 1–19. https://doi.org/10.1080/09602011.2020.1835676.

Schnadt, M. J., & Corley, M. (2006). *The influence of lexical, conceptual and planning based factors on disfluency production*. 28th annual conference of the … , pp. 8–13. https://csjarchive.cogsci.rpi.edu/Proceedings/2006/docs/p750.pdf

Senoussi, M., Berry, I., VanRullen, R., & Reddy, L. (2016). Multivoxel object representations in adult human visual cortex are flexible: An associative learning study. *Journal of Cognitive Neuroscience*, *28*(6), 852–868. https://doi.org/10.1162/jocn_a_00933

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica*, *119*(2), 159–187. https://doi.org/10.1016/j.actpsy.2005.01.002

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174