Journal

Experiments in Linguistic Meaning, 3(1)

Authors

Cantin-Savoie, Ghyslain
Winterstein, Grègoire
Foucambert, Denis

Publication Date

2025-01-24

DOI

10.3765/elm.3.5820

Studying the interplay of context and semantic content in the interpretation of adversative conjunctions with eyetracking

Ghyslain Cantin-Savoie, Denis Foucambert, & Grégoire Winterstein*

Abstract. The French adversative connective *mais*, much like its English counterpart *but*, takes two conjuncts and indicates that they stand in some kind of opposition. The nature of this opposition is often discussed in the existing literature of adversatives. Using an eyetracking experiment, we look at where and when this opposition appears to be manifested in a sentence reading task, using superiority comparatives sentences that appear degraded without a supportive context, as well as inferiority comparatives that do not seem to require such specific contextual information. We presented participants ($n = 28$) with four types of sentences, differing in the form of comparative, and in the information provided by the context. Some contexts contained a pivot property to help readers access the opposition of the two conjuncts, while others were neutral in that regard. Eye movements were recorded with a 250Hz Tobii Pro Fusion eyetracker, and mixed effect models were used to analyze the following eyetracking metrics : total fixation time and regression probability for the whole sentences, as well as first-fixation duration, first-gaze duration, regression-path duration, regressions-out and regressions-in for each individual words. Even if the helping context is shown to lower negative acceptability judgment on sentences with *mais plus* (Winterstein et al. 2014), we found no context effect in online reading processing, finding instead a persisting effect of the *less/more* dichotomy in all chosen measures, sometimes before fixation of those words, pointing towards a parafoveal effect of the aforementionned dichotomy, which merits closer look in future work.

Keywords. Semantics; psycholinguistics; eyetracking; reading; adversatives

1. Introduction. This work is centred on the semantics of the French adversative connective *mais*, typically rendered in English as *but*. The semantics of adversative connectives has been the topic of much work, in which two main theoretical branches can be distinguished. Those branches can be characterized by the use of adversatives that they treat as canonical (Winterstein 2017).

On one hand, the contrastive approach considers that so-called contrastive uses of connectives like *but* exemplify best the core constraint indicated by adversatives (see a.o. Umbach 2005). What seems to be required by *but* is that its conjuncts involve predicates that are comparable, i.e. similar in some respect and dissimilar in another.

(1) **Contrastive** John is tall but Richard is small.

On the other hand, inferential approaches take argumentative examples like (2) to be more representative of the semantics of adversatives (see a.o. Blakemore 2002, Winterstein 2012). Crucially,

*The authors would like to thank the audience of ELM3 for useful comments. This work was supported by a FRQSC *Soutien à la recherche pour la relève professorale* grant (2022-NP-296699). Authors: Ghyslain Cantin-Savoie, Université du Québec à Montréak (cantin-savoie.ghyslain@courrier.uqam.ca), Denis Foucambert, Université du Québec à Montréal (foucambert.denis@uqam.ca) & Grégoire Winterstein, Université du Québec à Montréal (winterstein.gregoire@uqam.ca).

the interpretation of such examples involves some additional inference, called a *pivot*, that connects the content of both conjuncts to a third proposition. Those two connections must be dissimilar in some ways.

(2) **Argumentative** He's unattractive but he's kind so yes, I will marry him.

In section 2, we discuss these approaches in more detail, highlighting a point of contention between them : the external or internal locus of the dissimilitude effect necessary for felicitous use of adversatives. We also show in that section that comparatives like the one in (3) allow us to control whether the required dissimilitude is internal to the conjuncts' semantics (like it's the case with *less tall*) or if an external proposition is needed to introduce it (like it's the case for *taller*).

(3) John is tall but less tall/*taller than Richard.

We then go on to describe a previous study by Winterstein et al. (2014) who found that acceptability judgment discrepancy between negative (*less tall*) or positive (*taller*) comparisons can be alleviated with the introduction of an external proposition, whilst reading-time differences between positive and negative comparisons persist.

This reading-time difference is what inspired our present eyetracking study, where we look for an online reading effect of the positive/negative comparison dichotomy (that we will call **valence** from now on) and where we see if this effect persists through the introduction of a preceding **context** made to help (or not) attaining required dissimilitude between conjuncts with an external proposition. Our hypothesis is that there is a **valence** effect and that there is no effect of **context** type on reading time. In section 3, we describe the item-making, participant recruitment, and material-installing processes we went into to design an eyetracking experiment with the goal of testing the aforementioned hypothesis. Filters, data transformation, eyetracking measures and statistical models are described in section 4. The results of those analysis are shown and discussed in 5.

2. Empirical background: adversatives and comparatives.

2.1. Contrastive and argumentative cases. Let us start with a closer examination of examples (1) and (2).

In the contrastive case of (1), John and Richard refer to distinct individuals, who possess mutually exclusive properties (i.e. being small and being tall). In (2), the argumentative case oppose two properties that are not mutually exclusive (i.e. being unattractive and being kind) predicated of a single individual (the referent of *he*). Though the examples differ in many ways, the one difference relevant to us is the locus of opposition, i.e. the question of determining from where stems the opposition that is required to license the use of an adversative like *but* as connective.

For the contrastive use, the locus of opposition is internal to the conjuncts, i.e. supposed to be supported by the lexical semantics of the material in the conjuncts. This means that for a contrastive sentence with *but* to be felicitous, the second conjunct needs to deny a stated or implied proposition that is brought by the first conjunct itself. In the sentence in (1), *John is tall* brings an answer to the question *WHO is tall*, therefore implying that individuals, like John, can be in the set of tall people. It also implies the alternative possibility that other people, like Richard, are also part of the tall people set. This said eventuality is denied by the second conjunct, *Richard is small*, which means that the use of but in this context is felicitous (see Umbach 2005 for detailed account along those lines, crucially exploiting QUD information to account for such cases).

For the argumentative use of *but*, the locus of opposition is necessarily external, which means it will be found in the way both conjuncts are linked to a third, contextually given, proposition. As an example, in (4), *He's unattractive* seems to go against *I will marry him* (indicated with a red arrow), while *he's kind* does not (indicated with a green arrow).

I will marry him

(4) He's unattractive **but** he's kind

Even if *he's unattractive* and *he's kind* are not necessarily opposed, it is through the opposing kind of connections to an external proposition that felicitous use of *but* is licensed.

2.2. COMPARATIVES CASES. In this work we do not directly engage with the question of determining which of the two uses of adversatives just described is the most canonical (see Winterstein 2012 for a discussion). We will however assume that some pivot inference is involved in the interpretation of adversatives, even in contrastive cases. To see why, we now introduce the core cases of this work, which we refer to as *comparative* cases, are exemplified in (5).

(5) a. John is tall, but less tall than Richard.
b. #John is tall, but taller than Richard.

Comparative cases appear similar to contrastive ones, given how both involve the comparison of two individuals along some property, like in (5). However, the property needs to be gradable, which is not a necessity for contrastive uses.

Tall being a gradable adjective, truth-evaluation needs a comparison of some kind. So, when it is said that *John is tall*, it inherently implies that John is taller than some threshold (Kennedy & McNally 2005). Following the method described in 2.1, *John is tall* answer the question "WHO is less tall than John?" by telling us that *someone* is in the set of things that are less tall than John ; If there is no one who is less tall than John, it cannot be said that John is tall. Many people could possibly be in that set, including Richard. The second conjunct of (5-a), *[John is] less tall than Richard* denies the inclusion of Richard in the aforementioned set, predicting a correct use of *but*. Since *[John is] taller than Richard* does not deny this implication, an internal-opposition-locus view of comparative uses of *but* seems to correctly predict the degraded nature of (5-b).

When showed sentences like those in (5) in French, people judged sentences like (5-b) to be infelicitous (Winterstein et al. 2014). Reading times for sentences like those in (5-b) were also shown to be longer compared to (5-a). In the same experiment, another group of participants saw the same sentences presented after a context that introduces an external opposition pivot in the form of an equality condition, like in (6-a).

(6) a. **Context:** We are looking for a stunt double for Richard, an actor, in action scenes for a movie. The stunt double needs to be exactly the same height as Richard so that it goes unnoticed. The search is hard, because Richard is tall.

b. **Sentence:** John is tall, but less tall/taller than Richard.

John is the stunt double we are looking for

c. **Opposition:** John is tall **but** less tall/taller than Richard

When preceded by contexts like (6-a), sentences like (5-b) were not judged to be infelicitous anymore. However, reading time differences persisted, suggesting that the presentation of a context was not enough to facilitate the interpretation of the adversative connective. This suggests that those examples are somehow on a par with argumentative cases in that their interpretation rests on the identification of a contextual pivot. The reading times data also suggest that this identification comes at a later stage of interpretation.

Based on results obtained with the experiment of Winterstein et al. (2014), we hypothesize that if we show sentences like (5) to participants, reading times will be greater on sentences with *plus* (more) like in (5-b) than the reading times on sentences with *moins* (less) like in (5-a). Duration-based eye-tracking measures, described below in 4, should be sufficient to deny or confirm this hypothesis. Secondly, since we think that the locus of opposition is internal, lack of opposition internal to the conjuncts cause by positive comparison with *more* should create a problem in online processes, regardless of context manipulations.[1] Therefore, our second hypothesis is that if we preceded sentences like those in (5) with helping contexts like in (6-a), the reading time effect of the *less/more* dichotomy will persist. To test this, we created another type of context, called ”neutral”, that are similar to the one in (6-a) but does not introduce an equality condition that allows an external opposition of conjuncts. If we are right, we should not see an effect of the nature of context on the eyetracking measures we observe.

3. Method. To test these hypotheses, we ran an experiment in which participants had to read comparative sentences with *mais* preceded by contexts introducing (or not) an equality condition. The eye movements of the participants were recorded using an eyetracking device. This section describe the sentences that were shown, the hardware used, participants of the experiments, and the experimental protocol.

3.1. EXPERIMENTAL ITEMS. Each item had 3 parts : (1) a context, (2) a target sentence, and (3) a question. There were 60 items in total, of which 40 were distractors, meaning that the sentences were not comparatives with *mais*, so that subjects would not easily detect the objective of the experiment. 10 items and 30 distractors were extracted from Winterstein et al. (2014) and slightly modified to be adapted to Quebec French, the rest of the items were created for this experiment.

The experiment used a 2x2 design, with the two binary conditions as follows.

First, contexts from experimental items came in two versions (condition `Context`). One version introduced an equality condition like the one shown in (6-a), and was called helping. The other version presented information irrelevant to the interpretation of the adversative, and is referred to as the neutral version. We illustrate these version in (7).

[1]This predict that for uses where we think opposition is external like argumentative cases in (2), context manipulation should have reading time effects. Another experiment will be needed to look into this.

(7) Jeanne Doe est amnésique et ne se rappelle plus qui elle est. Les enquêteurs pensent qu'elle est la soeur jumelle/ petite soeur de Marc Leclerc, un vieil avocat porté disparu il y a 10 ans de cela. Pour s'en assurer, ils font passer à Jeanne une batterie de tests.
Translation : Jane Doe is amnesiac and doesn't remember who she is. Investigators think that she is the twin / little sister of Marc Leclerc, an old lawyer that disappeared 10 years ago. To be certain, they make Jane go through a battery of tests.

Each context was modified or created so that the total number of characters in the context would be inside a 20 character range, between 222 and 242 characters.

Second, target sentences in the experimental items were comparative uses of *mais*. Each sentence existed in two versions (condition `Valence`), depending on the kind of comparative being used: either a superiority comparative (*plus X que* / 'more X than') or inferiority one (*moins X que* / 'more X than'). Crucially, each version only differs from the other in the choice of the comparative adverb. These two versions are shown in (8).

(8) Selon les tests, Jeanne est vieille mais **moins/plus** vieille que Marc même s'ils ont tous les deux les cheveux blancs.
Translation : According to the tests, Jane is old, but less old/older than Marc even if they both have white hair.

The target sentences differ from previously shown examples in (5) in that they are surrounded by complement phrases, so that critical zones of interest (i.e. around the adversative connective) are not shown either at the extreme left or the extreme right of the screen when read. Character number for the target sentences varied between 108 and 128.

Questions were introduced as a way to distract participants into thinking the answer was the criterion we evaluate, and to determine whether participants were really reading sentences or only pressing space while moving their eyes. Questions were between 48 and 68 characters, each one asking for a yes/no answer that indicated comprehension of the previously read context-sentence pair, like shown in (9) about the sentence in (8) and the context in (7)

(9) Est-il possible que Jeanne Doe soit la soeur jumelle de Marc Leclerc?
Translation : Is it possible that Jane Doe is the twin sister of Marc Leclerc?

Every part of an item, context, sentence or question, had a mean lexical frequency maximum of 737.84 , evaluated as the maximum threshold of the interquartile range. Lexical frequency for each word was found using Lexique3 database trained on the Frantext corpus, which contains 218 French novels from 1950 to 2000 (New et al. 2004).

3.2. Equipment. Subjects sat on an adjustable desk-chair so that their heads, regardless of their individual height, would be at the same spot. Their chin rested on a table model chin-rest, as they were looking at an Asus TUF Gaming VG259Q computer screen of size 24.5" with a colour depth of 8-bit, resolution of 1920 per 1080 and a refresh rate of 59.94Hz, placed 60 centimetres away from their face. Their eye movements were observed with a Tobii Pro Fusion 250 Hz eyetracker.

3.3. Participants. This article analyses the data of 28 participants, 13 male and 15 female, aged between 19 and 68, with a mean age of 31.5 years old. We invited them by going in their university class and sending Facebook calls to student groups and friends. Participants were compensated 15

CAD for their participation. Participants were selected to have only Quebec French as their first language, and to have a normal or corrected vision. Their highest completed levels of education ranged from high school to PhD.

3.4. PROCEDURE. After arriving and signing a consent form, participants were placed in front of a screen, rested their chins on the chin-rest, underwent calibration, For which they had to follow a moving dot displayed on the screen. They then completed a first test item after which they were given the chance to ask questions before proceeding with the rest of the experimental items.

They were warned that before each item part, a fixation cross would appear and they would be asked to look directly at it before pressing the space bar to go to the next stimulus. They were asked to read each item part before pressing space, except after the question part, where they were asked to answer *yes* by pressing A or *no* by pressing L.

To instantiate the experiment, we used the Tobii Pro Lab software, which unfortunately does not offer latin-square style pseudo-randomization. We thus created four different trial groups, to ensure that contexts and sentences where shown so that every participants would see the same number of each condition, and each group would not see the same items in the same conditions than other groups, as shown in table 1.

Experimental Items	Group 1	Group 2	Group 3	Group 4
1-5	helping|less	neutral|less	helping|more	neutral|more
6-10	neutral|more	helping|less	neutral|less	helping|more
11-15	helping|more	neutral|more	helping|less	neutral|less
16-20	neutral|less	helping|more	neutral|more	helping|less

Table 1: Condition repartition amongst participant groups

After showing group 1 items to a participant, the next participant would see group 2 items, and so on to keep an even spread. After the experiment but before briefing, they would go through a reading span test in order to evaluate their respective working memory[2].

4. Analysis. Since 28 participants read 20 sentences each, we would expect 560 sentence-level data points. However, item number 19 being one character too long, it showed on 2 lines on the screen and was therefore removed. Our eyetracking device recorded valid gaze location data for 466 of the 532 remaining sentences. Fixation data was then retrieved using the Tobii-Pro I-VT fixation filter (TobiiAB 2023). 5 sentences were removed for containing less than 4 fixations each, resulting in a total of 461 sentence tokens. After removing fixation data on 10 words that were fixated less than 60 ms each, we had fixation data for a total of 5703 word tokens. Data analysis was carried out on two separate levels : sentences and words.

4.1. THE SENTENCE LEVEL. At the sentence level, we took total fixation duration and regression probability as indicators of global sentence-reading difficulty (Goldberg & Kotval 1999). Total fixation duration is defined as the sum, in ms, of the durations of fixations that happened in the reading of each sentence-part of the item triads. Regression probability is the number of regressive saccades made during the recording of a sentence, divided by its total saccade count. Each of those

[2] We don't go in details in this article about this test's procedure, it is described in Marcotte (2014).

measures were took as dependent variables (DV) in their own mixed effect model generated with lmer() (Bates et al. 2015) in R, sharing the formula shown below in (10).

(10) $\text{DV} \sim \text{Valence} + \text{Context} + \text{ScolCal} + \text{Span} + (1|\text{Participant}) + (1|\text{Media})$

Valence is a categorical variable with two possible values : *plus* ('more') or *moins* ('less'), that represents whether the word following *mais* ('but') in item sentences was *plus* (i.e. a superiority comparative) or *moins* (i.e. an inferiority comparative). Our hypothesis is directly linked to this variable.

Context is a categorical variable with two possible values, *helping* or *neutral*, that indicates whether the context preceding the sentence contains a helping opposition pivot in the form of an equality condition or not. Our hypothesis is also directly linked to that variable.

ScolCal is an ordinal variable whose values range from 1 to 4. It is used to take into account the highest school level of participants, 1 for High School, 2 for Cegep [3], 3 for a bachelor's degree from university and 4 for masters, PhDs, and higher.

Span is a continuous variable representing the result of the Reading Span Test, to control for working memory, which can affect several reading behaviours (Traxler et al. 2012), as well as other potential individual cognitive abilities related to reading (Kim et al. 2021).

Participant and Media are factors with levels representing participant Id and sentence number [4], respectively. The way they are instantiated in the model (1|Variable) means that the intercepts, but not the slopes, are random, to allow a middle ground between model complexity and statistical power. They are interpreted by the model as cross-random effects giving us only 1 intraclass correlation coefficient (ICC) per result table, which is acceptable since all of our participants saw all types of sentences.

Note that in the model presented in (10) we did not account for interactions between independent variables. Models testing for interactions between our two hypothesis-making independent variables were compared to models without interactions. Since there were no significant differences between the two in terms of fitting, we chose the simpler version to avoid problems linked to model complexity.

4.2. THE WORD LEVEL. At the word level, we considered five different measures. For each word that was fixated at least once, we capture the duration of the first fixation (in ms). First fixation duration is linked to very fast first order cognitive processes like word and letter recognition (Rayner 1998). We also included first gaze duration, which is the sum of all fixation durations in milliseconds on a given word from the first fixation until the subject fixates on something else. 87% of the time, according to our own data, first fixation duration and first gaze duration have the same value, since many words are only fixated once before a saccade brings the gaze on another one (Rayner 1998). First-gaze duration are said to be influenced by cognitive processes that are not as fast as those influencing first fixation duration (Rayner 1998), and also by plausibility of words in sentence context, which is not the case for first fixation duration (Inhoff 1984).

[3] Cegep (Collège d'Enseignement Général Et Professionnel, roughly translated to College of General and Professional Teaching) is a school level unique to Québec with students entering it at around 17 or 18 years old, created to make higher education accessible and affordable to a wider population.

[4] As an example, the sentence about Jane Doe previously shown is always number 13, regardless of whether it is in the *plus* or *moins* version.

Fixation #	Jane	is	old	but	less	old	than	Marc
1	X							
2					X			
3						X		
4				X				
5			X					
6	X							
7								X

Table 2: Fixation order of a hypothetical sentence reading

Regression-path duration for a word is the sum of all fixation durations, from the first one on the word until a fixation is made on another word to the right of it. In Table 2, showing the ordering of fixation on a hypothetical sentence reading, the regression-path duration of the word *but* would be the sum duration of fixations 4, 5 and 6.

Regression-path duration is used to indicate processing difficulties oftentimes caused by parsing effects, such as syntactic ambiguities (Hyönä et al. 2003).

For each word, we also calculated its probability of being the starting or ending point of an ocular regression (regression-out, regressions-in).

Each of these measures were taken as the dependent variable of a lmer() model shown in (11), similar to the one in (10).

(11) DV $\sim$ Valence + Context + ScolCal + Span + sFreqLex + nbCarac + (1|Participant) + (1|Media)

The model in (11) and the choices leading to its form are identical to those described in section 4.1, with the exception of 2 additional fixed effects.

nbCarac contains integer numbers representing word length by character number, which is said to influence if a word is going to be fixated on in the first place (Rayner 1998). sFreqLex represents the lexical frequency of the word, taken from the Lexique3 database made with the Frantext corpus, transformed with the scale() R function, to avoid scaling issues brought by the large value range inherent to the frequency values.

We analyzed the words in two steps. First, each word-level measure took turns being inserted as dependent variable in the model described above, allowing us to look for global effects of valence and context on words. Secondly, we indexed words by position in the sentence, taking *but* as 0, the word before it as -1, the word after it as 1, and so on. We then selected a segment of the sentences we call **critical zone** going from indexes -3 to 4, where there is no variation in word type per index. As an example, the word before *mais* (but) at position -1 is always the first iteration of the gradable adjective. However, words at the position 7 can be articles, nouns or verbs without any given link between them, so taking position as a category would be a mistake. [5]

5. Results.

[5] Some part-of-speech or semantic related tagging could alleviate this limit, and would be interesting in subsequent research.

5.1. SENTENCE LEVEL: RESULTS. Results at the sentence level in Table 3 and Table 4 show that sentences are fixated longer and that participants are more likely to look back when reading sentences with *plus* ('more'), which indicates that superiority comparatives increase sentence reading difficulty. However, context type does not affect those metrics, as found by Winterstein et al. (2014).

	logdur Fix		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	8.21	7.56 – 8.87	**<0.001**
ValenceEN [More]	0.12	0.02 – 0.22	**0.024**
ContexteEN [Neutral]	0.06	-0.02 – 0.13	0.157
ScolCal	0.03	-0.09 – 0.14	0.665
Empan	-0.05	-0.22 – 0.12	0.573
Random Effects			
σ^2	0.16		
τ_{00} Media	0.01		
τ_{00} Participant	0.10		
ICC	0.42		
N Media	38		
N Participant	28		
Observations	461		
Marginal R^2 / Conditional R^2	0.019 / 0.431		

Table 3: Model results for the log transform of total fixation duration of sentences

	prob Reg		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	0.31	0.18 – 0.44	**<0.001**
ValenceEN [More]	0.03	0.00 – 0.05	**0.026**
ContexteEN [Neutral]	0.01	-0.00 – 0.03	0.135
ScolCal	-0.00	-0.03 – 0.02	0.812
Empan	-0.01	-0.05 – 0.02	0.390
Random Effects			
σ^2	0.01		
τ_{00} Media	0.00		
τ_{00} Participant	0.00		
ICC	0.28		
N Media	38		
N Participant	28		
Observations	461		
Marginal R^2 / Conditional R^2	0.023 / 0.299		

Table 4: Model results for regression probability of sentences

Sentence-level results, visualized in Fig. 1 and Fig. 2, confirm that `Valence` has an effect on reading times. Since our question is about where and when it happens, we move to results at the word level.

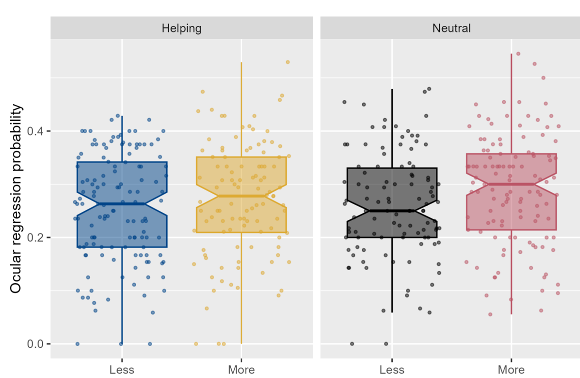


Figure 1: Regression probability on sentences, split by experimental conditions

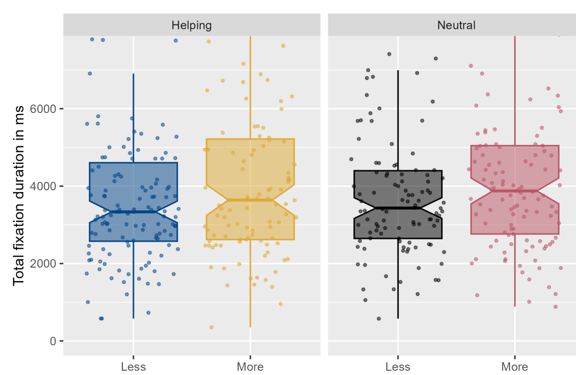


Figure 2: Total fixation duration in milliseconds on sentences, split by experimental conditions

5.2. WORD LEVEL: RESULTS. At the word level, when we take all words into account, there is no significant effect of our variables on either first fixation or first gaze duration. However, if we test on individual word positions in the critical zone described in section 4.2, we find a significant effect of `Valence` on first gaze duration on two word positions, shown on a white background on Fig. 3.

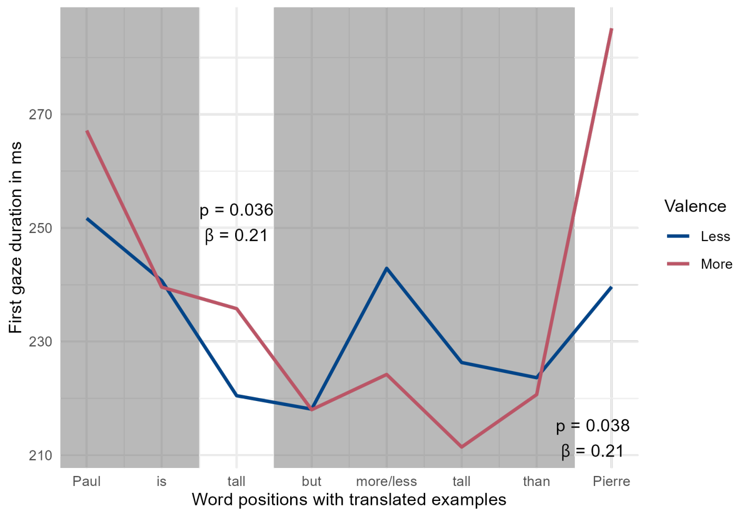


Figure 3: First gaze duration per word position, white background = significant effect of `Valence`

The fact that first gaze duration is longer on those two words cannot be explained by saying that characters in the words or their lengths are harder to decode, since if it were the case there would also be a first fixation effect, which is not the case. Also, and most importantly, the effect of the *more/less* dichotomy is seen on first gaze duration on the first occurrence of the gradable adjective ('tall'), which is before the occurrence of *more/less*. Firstly, this indicates a limitation in metrics definition : we don't know if the word *more/less* was fixated before *tall* in the cases where first gaze duration is higher. If it is not the case, it would indicate a parafoveal effect of `Valence` on the reading of the word *tall*, which would need subsequent work to confirm.

If we were to hypothesize an effect of `Context` on a given metric, it would be on regression path duration, since it is said to represent not only the time spent decoding words, but also the time spent integrating visual level information with previously known words and facts stored in memory, such as personal knowledge or information given in a preceding context (Cook & Wei 2019). However, there was no effect of `Context` on regression-path duration, not on all words taken as a whole nor on each individual word of the critical zone. There was, however, a significant `Valence` effect on the individual word position for the second occurrence of the gradable adjective, shown on Fig. 4 where the background is white.

Those results mean that when participants look at the second occurrence of the gradable adjective, they take more time and have to go back more in the superiority comparative cases.

Since we showed previously a significant `Valence` effect on regression probability at the sentence level, it is no surprise that regressions-out and regressions-in probability on individual words also have a significant effect of `Valence` when taken on all words. When we look at the

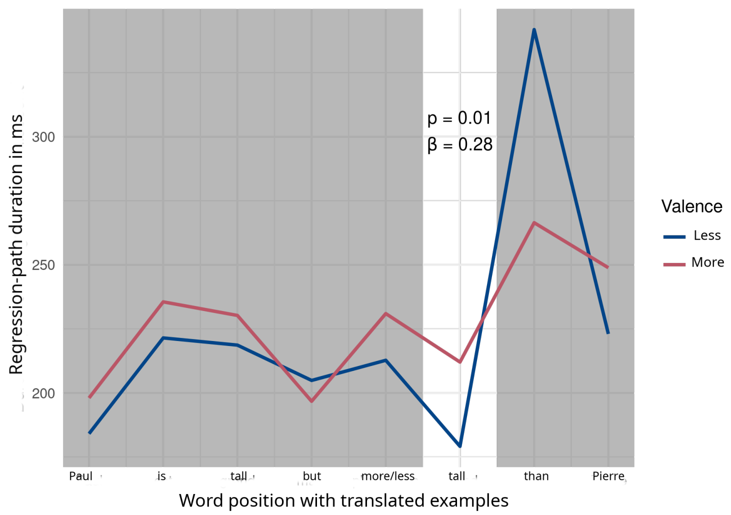


Figure 4: Regression-path duration on word position, white background = significant effect of Valence

critical zone, we can see that there are more regressions coming from and out of the first occurrence of the gradable adjective, as shown in the white background of both Fig. 5 and Fig. 6.

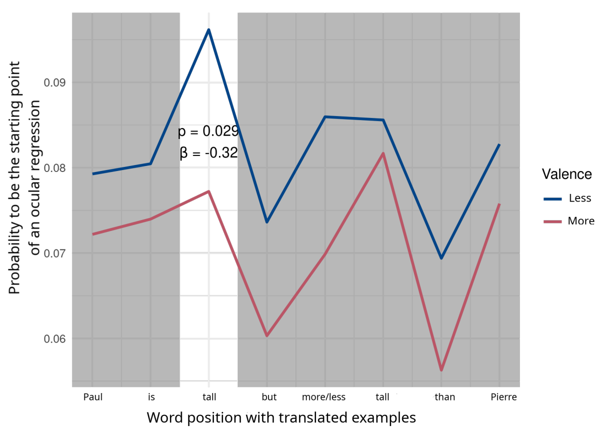


Figure 5: Regressions-out probability for each word position, white background indicating a significant effect of `Valence`.

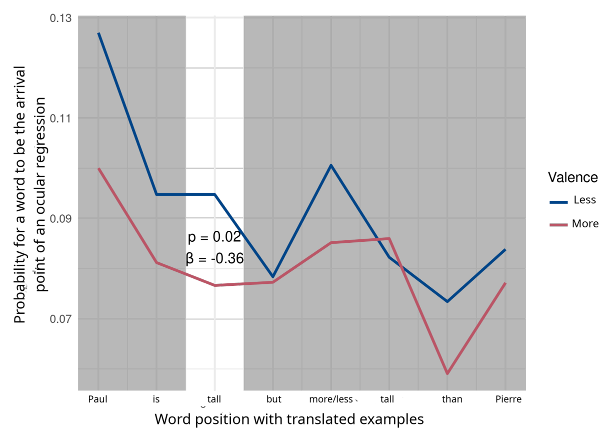
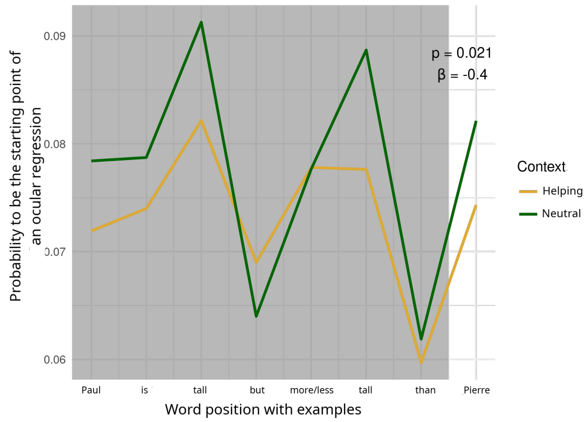


Figure 6: Regressions-in probability for each word position, white background indicating a significant effect of `Valence`.

These results are not indicative of the fact that participants are regressing from the first occurrence of *tall* to itself, and is therefore indicative of a limitation of our study: there is no chronological information about saccade for now that allows us to know where the regressions from *tall* go, and where they come from. This would requires subsequent analysis.

There is no `Context` effect on regressions-in and regressions-out probabilities when taking all words together. However, looking at the critical zone allows us to see (Fig. 7 and 8) that more regressions get out of the second named entity (*Pierre* in the example) when the context is neutral to target the first named entity (*Paul* examples). Then again, subsequent analysis is needed to discover if a significant number of saccades go from *Pierre* to *Paul*.



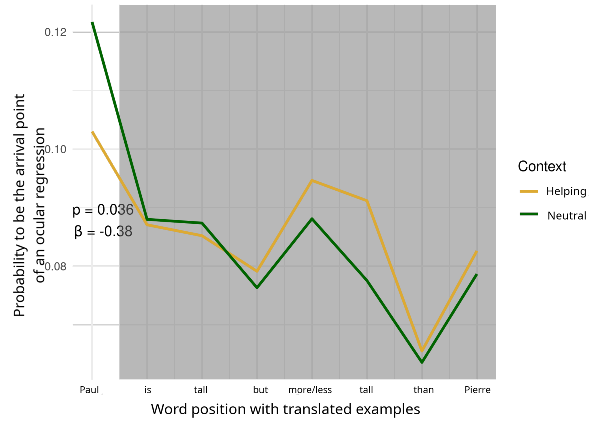


Figure 7: Regressions-out probability for each word position, white background indicating a significant effect of Context.

Figure 8: Regressions-in probability for each word position, white background indicating a significant effect of Context.

6. Conclusion and outlooks. This study tested hypotheses regarding the reading time increase of comparative sentences with *mais* (but) comparing superiority and inferiority comparatives. Theoretical work about the semantics of adversatives, as well as previous research using self-paced reading data (Winterstein et al. 2014), led us to think that (1) sentences with *plus* (more) will have higher reading times, and that (2) contexts introducing (or not) an external proposition that explicitly supports the opposition between conjuncts should not affect reading times. Results of an eyetracking study confirmed these hypotheses, additionally allowing us to point out some limits and future research avenues.

Sentences with a superiority comparative (*plus* 'more') showed an increase in total fixation time for whole sentences as well as a localized increase of first gaze duration and regression-path duration, which seems to confirm our first hypothesis. Although context type had no effect on duration-based metrics and therefore on reading times per say, confirming our second hypothesis, results from regressions patterns showed the influence of context on **where** participants decide to look back from and to. Context type seems to influence whether participants will regress from and to named entities, whilst most of the localized duration effects were found on the gradable adjectives.

Limitations to our analysis were made apparent during this study, including lack of tagging on individual words that would have allowed us to categorize them by semantic function or syntactic category. Further analysis of our data will allow us to inspect our finding that duration-based metrics show valence effects on the gradable adjectives while location-based metrics show valence effects on the named entities. A boundary paradigm study would be needed to confirm parafoveal effects influenced by the semantics of *plus/moins*. Finally, another study with different items would be key to investigate whether different kind of adversative sentences (not only comparatives) with an external locus of opposition would have duration-based effects of context manipulation.

References

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. 10.18637/jss.v067.i01.

Blakemore, Diane. 2002. *Relevance and Linguistic Meaning. The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.

Cook, Anne E & Wei Wei. 2019. What can eye movements tell us about higher level comprehension? *Vision* 3(3). 45.

Goldberg, Joseph H & Xerxes P Kotval. 1999. Computer interface evaluation using eye movements: methods and constructs. *International journal of industrial ergonomics* 24(6). 631–645.

Hyönä, Jukka, Robert F Lorch Jr & Mike Rinck. 2003. Eye movement measures to study global text processing. In *The mind's eye*, 313–334. Elsevier.

Inhoff, Albrecht Werner. 1984. Two stages of word processing during eye fixations in the reading of prose. *Journal of verbal learning and verbal behavior* 23(5). 612–624.

Kennedy, Christopher & Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 345–381.

Kim, Young-Suk Grace, Yaacov Petscher & Christian Vorstius. 2021. The relations of online reading processes (eye movements) with working memory, emergent literacy skills, and reading proficiency. *Scientific Studies of Reading* 25(4). 351–369.

Marcotte, Sylvie. 2014. Étude en temps réel de la révision de la morphographie du nombre du verbe chez les étudiants universitaires .

New, Boris, Christophe Pallier, Marc Brysbaert & Ludovic Ferrand. 2004. Lexique 2: A new french lexical database. *Behavior Research Methods, Instruments, & Computers* 36(3). 516–524.

Rayner, Keith. 1998. Eye-movements in reading and Information-Processing: 20 years of Research. *Psychological Bulletin* 124(3). 372–422.

TobiiAB. 2023. Tobii pro lab. Computer software. `http://www.tobii.com/`.

Traxler, Matthew J, Debra L Long, Kristen M Tooley, Clinton L Johns, Megan Zirnstein & Eunike Jonathan. 2012. Individual differences in eye-movements during reading: Working memory and speed-of-processing effects. *Journal of eye movement research* 5(1).

Umbach, Carla. 2005. Contrast and information structure: A focus-based analysis of *but*. *Linguistics* 43(1). 207–232.

Winterstein, Grégoire. 2012. What *but*-sentences argue for: a modern argumentative analysis of *but*. *Lingua* 122(15). 1864–1885. 10.1016/j.lingua.2012.09.014.

Winterstein, Grégoire. 2017. *Perspectives on Argumentation within Language. Theoretical, Processing, Computational and Social aspects.*: Université Paris Diderot–Paris 7 dissertation. (Habilitation à diriger les recherches).

Winterstein, Grégoire, Emilia Ellsiepen, Jaques Jayes & Barbara Hemforth. 2014. Effects of context on the processing of adversative and comparative constructions. In *Cuny*, .