

Good French isn't always best. Acceptability and linguistic prescriptivism

Gabriel Thiberge, Laboratoire de Linguistique Formelle, CNRS – Université Paris Cité, FR; Laboratoire Cognition, Langues, Langage, Ergonomie, CNRS – Université Toulouse Jean Jaurès, FR, gtac@tuta.io

Barbara Hemforth, Laboratoire de Linguistique Formelle, CNRS – Université Paris Cité, FR, barbara.hemforth@univ-paris-diderot.fr

Acceptability judgments are one of the major tools for (psycho)linguists to assess speakers' preferences for specific utterances in a given language, shedding light on the grammar of the language under study. However, it is well known that factors that are not related to grammaticality, such as frequency of exposure, cognitive constraints, and others, can influence the perceived acceptability of an utterance. We will use the system of *wh*-interrogatives in French as an example to study the impact of linguistic norms on what is considered “good” French. In three experiments, we show that adult L1 French speakers have internalized the dichotomy between variants that are considered “good French”, according to the norms, and those that are suited to more informal daily life situations. Speakers can express these differences when given the appropriate tools, but not with a unique general acceptability scale. In line with previous work, we argue that acceptability judgments are a useful task, but that they need to be refined to account for sociolinguistic factors that constrain speakers' assessments (i.e., linguistic norms, but also speaker group and formality of the context of interaction).



1 Introduction

Quantitative research in linguistics has developed considerably in recent decades, as a way to move past linguists' intuitions and to build a more systematic approach to assessing what is part – or what is not part – of a given linguistic system, based on a broader basis of naive L1 speakers (Gibson & Fedorenko, 2010, 2013). Multiple observations of the influence of minimally differing linguistic contrasts on various dependent variables (e.g., reaction times or ratings), repeated across many different participants and many different items in a counterbalanced way, simultaneously allow for robust and fine-grained inferences from the relevant linguistic and cognitive variables at a general population level of speakers and linguistic stimuli. DeBruine and Barr (2021), in their explanation of mixed models in data analysis, describe this as generalizations across encounters of participants and experimental stimuli. Of particular interest to the (psycho) linguistic community from the very beginning of cognitivist approaches and over the next decades (Chomsky, 1965; Schütze, 1996), the Acceptability Judgment Task has become one of the most widely used experimental paradigms, where ratings of naive participants are collected about a variety of linguistic forms to assess which ones they accept as part of their language. The general paradigm is easy to set up and use, and has become more standardized over the years, following calls for a more generalized approach (Cowart, 1997; Ferreira, 2005; i.a.). Because it is such a widely used paradigm, the acceptability judgment task has long been scrutinized, and a recurring worry concerns how well its results can generalize from the observed data to a wider population. Over the years, several issues have been addressed, including how informal designs without a clear systematic or inferential approach to the results (*informal* methods) compare to more constrained and somewhat artificial setups inspired by psycholinguistic methods (*formal* methods) (Sprouse et al., 2013), how different designs vary in terms of sensitivity (Marty et al., 2020; Sprouse & Almeida, 2017), and how the reliability and variability of results may depend on factors such as the scales used (Langsford et al., 2018).

Perhaps a broader and deeper issue concerns what exactly the task is measuring and the very definition of *acceptability*. Entire journal issues focus, for example, on whether or not – or in how far – (un)acceptability conflates with (un)grammaticality (Tubau et al., 2020). While grammaticality is supposed to be a property of a linguistic stimulus itself, acceptability rather touches upon the perception of stimuli by naive participants, whose judgments may be biased for reasons independent of grammaticality and possibly depending on the experimental situation (Huang & Ferreira, 2020). From what the authors consider “noisy data” (noise here being the result of participants' biases expressed during the task), Huang and Ferreira (2020) propose applying Signal Detection Theory to separate participants' capacity to distinguish acceptable from unacceptable sentences from their judgment biases (see also Dillon & Wagers, 2019). In another line of study, Hubers et al. (2020) consider, with support from eye-tracking data, some intermediate status between purely grammatical sentences and purely

ungrammatical ones, with some sentences judged acceptable while in violation of prescriptive grammar rules. This, in a sense, is reminiscent of older proposals (Blache et al., 2006; Keller, 2000; Sorace & Keller, 2005) establishing acceptability as a more continuous notion, where – in the authors’ theoretical framework – graded judgments result from the coexistence of *soft* (violable) and *hard* (inviolable) syntactic constraints. Graded judgments can also be the result of the interaction of different factors coming from linguistic and non-linguistic domains. As argued, among others, by Schütze (1996), acceptability judgments are influenced not only by the grammaticality of a construction, but also by a variety of other factors such as processing difficulty, familiarity, semantic plausibility and many more, with sometimes mixed results. On the issue of familiarity, for instance, Culicover et al. (2022) and the (extended) radical unacceptability hypothesis posit that acceptability is directly correlated with surprisal, which makes higher-frequency forms more acceptable. In contrast, earlier works (Bader & Häussler, 2010) show that while there is indeed a link between frequency and acceptability, ceiling and floor effects can be observed. Specifically, equally well-formed forms may differ in their frequencies, and equally infrequent forms can lead to divergent opinions with regard to their well-formedness.

Akin to yet another strand of studies investigating the concept of acceptability and its many dimensions, this article approaches acceptability judgment tasks from a variationist perspective, with a focus on alternation phenomena where different linguistic forms seem to coexist freely in a linguistic system while seemingly carrying the same semantic meaning. Some theoretical works (Adger, 2006, 2007) have tried to reconcile “grammaticality”, in the sense of what linguistic forms can appear within a given language, with the concepts of sociolinguistic variation and social context. Such *sociosyntactic approaches* of what is or is not *acceptable* in a language variety often consider intuitive judgments to be constrained by the frequency of the variants in the variety under study, but also by speaker identity and linguistic prescriptivism. To our knowledge, these reflections have, however, mostly extended to research on dialectal variation, where socially marginalized forms can sometimes be difficult to elicit systematically (Buchstaller & Corrigan, 2011; Jamieson et al., 2024).

This multifactorial perspective on acceptability also resonates with some experimental evidence for the direct influence of linguistic prescriptivism on participants’ evaluation of acceptability. For example, following Cowart (1997), Robbins (2019) finds that when participants are asked to rate sentences as professors, tutors, or even when “ignoring school grammars” (*intuitive* condition), they provide different assessments of the same sentences. This work seems to tap into the differentiation between normative/prescriptive grammars and more descriptive, usage-based grammars as the source of differences in acceptability judgments. A more in-depth investigation of how linguistic norms and prescriptivism bias linguistic assessment tasks, such as the acceptability judgment paradigm, seem to be important.

In line with this perspective, and inspired by works from third-wave sociolinguistics (Eckert, 2012), we are here more interested in variation *within* a given linguistic system, where the choice between alternating forms depends on social situations (i.e., we are not studying variation at the dialectal level, since all variants are available to all speakers of French). In this perspective, speakers are aware of the suitability of constructions for particular social contexts, and they use their sociopragmatic knowledge to adapt to each social situation. This level of variation may, thus, be exhibited by one and the same speaker in different contexts, but may still be influenced by demographic factors, such as age or socioeconomic background (e.g., education level or profession).

In studies on dialectology (see, for example, Cornips & Poletto, 2005), the approach taken is to combine corpus exploration with studies using acceptability or grammaticality judgments to pinpoint fine-grained differences in a more controlled way. However, as has been noted by a variety of authors going back to Labov (1996), acceptability and free usage of constructions can be very different things. Acceptability judgments are typically informed by prescribed norms so that highly used constructions that do not follow the norms are typically judged less acceptable. In line with Jamieson et al. (2024), we suggest that intuitions do not necessarily fail when sociolinguistic parameters come into play and that more fine-grained intuitions can be measured when the right questions are asked.

Here, we explore this issue by looking at the system of *wh*-interrogatives in French, and we focus on how prescriptivism and a speaker's relation to the linguistic norm might explain differences in acceptability for a given set of linguistic forms. We argue that French is particularly suited to study the sociopragmatic factors that play a role in acceptability judgments, partly because formal (mostly written) and informal (mostly spoken) French sometimes seem to follow very different rules, to the point that some researchers even consider them as two separate systems, placing French in a quasi-diglossic situation (Zribi-Hertz, 2006). Although we do not follow this diglossic approach, we think that studying acceptability judgments for variants of *wh*-interrogatives in French may help shed new light on fine-grained factors influencing acceptability. By running a series of judgment tasks and, in particular, by introducing a double scale with which participants could give their assessment of different variants available to ask a *wh*-question in French, we provide evidence for different nuances of acceptability of these sentences. Based on available corpus data exploring the same phenomenon (Thiberge et al., 2021), the design of our experiments provides evidence for an influence of contextual and sociolinguistic factors on judgments. We will show that L1 French speakers differentiate between what normatively “grammatical” French is supposed to be (which variant is better within a prescriptive view of the language) and what is actually “acceptable” in a given interaction (which variant is most often used in everyday French). To be able to express these aspects of acceptability, participants must be given the proper tools, and, on the basis of our

data, we argue that careful consideration should be given to the stimuli used in acceptability judgment tasks. Speakers seem to be aware of the difference between normative *acceptability* and what we call general *suitability*, but this difference also depends on sociolinguistic factors potentially external to the linguistic form itself, such as context formality and the speaker's age group.

This paper is organized as follows. In Section 2, we give an overview of the French *wh*-interrogative system and of the data already available with regard to this alternation phenomenon, and present our analysis of what different approaches to the notion of acceptability would predict in terms of what the acceptable variants should be for French speakers. In Sections 3 and 4, we present results from a series of three different acceptability judgment tasks with different acceptability scales and, with Bayesian inferential statistics, we give evidence for a combined influence of context formality and of the age group of participants on the results. Perhaps more crucially, the number of scales given to participants modulates how they express their judgments. In Section 5, we summarize our findings and argue for the need to go beyond a “general good-for-all acceptability” scale in (psycho)linguistic experiments, to better take into account – and maybe mitigate – the influence of prescriptive linguistic norms on acceptability judgments.

2 French *wh*-questions as a case study for acceptability

In this article, we will focus on French *wh*-interrogatives as a case study of how prescriptive linguistic norms affect acceptability and how sociolinguistic factors might influence speakers' assessments of acceptability. *Wh*-interrogatives show a fairly wide range of variation in French, with many different syntactic alternative forms, or *variants*, which basically all express the same semantic meaning. From available corpus data (Thiberge et al., 2021), it appears that these variants alternate in production depending on extra-linguistic factors such as context, age group of the speaker, or even a combination of both (not all groups of speakers prefer the same variants in a particular context). After a description of the phenomenon, we derive predictions as to how judgments of *wh*-interrogatives depending on the manipulated conditions could help to better understand aspects of *acceptability*. As we show below, our data suggests that speakers show intra-individual variation with respect to the acceptability of certain variants as they adapt to the social situation. In line with works such as Labov et al. (2011), which claim that speakers of a given language or language variety have a *sociolinguistic monitor* that enables them to detect subtle differences in sociolinguistic incongruities, we take this as evidence that speakers of a language have the sociopragmatic competence to evaluate a variant as optimal – or suboptimal – given a particular social situation. We will argue that this is an underlying dimension that may have to be added to the complexity of acceptability judgments.

2.1 A wide range of alternatives for asking *wh*-questions in French

Following Beyssade and Marandin (2006), Ginzburg and Sag (2000), and Marandin (2018), interrogation is a speech act that allows a speaker to ask about a missing piece of information. This missing piece of information can be the truth-value of a proposition (*polar* question with, roughly speaking, a *yes* or *no* answer, like : *Did you do this thing?*). This missing piece of information can also, and more specifically, be an unknown element of a set upon which a given predication is stated (*wh*- or *partial* question: *Someone took my apple and ate it. Who did this?*). Without going into a much more detailed semantic analysis of interrogation as a whole (for foundational work, see Hamblin, 1973; Karttunen, 1977), we will focus on the coexistence of multiple syntactic variants to ask such *wh*-interrogatives in French.

French *wh*-interrogatives are an already well-documented point of linguistic variation, with a wide range of corpus studies starting from the mid-20th century (Adli, 2015; Ashby, 1977; Behnstedt, 1973; Coveney, 2011; Hamlaoui, 2009; Pohl & Straka, 1965; Reinhardt, 2019a; Söll, 1983; Terry, 1970; Thiberge et al., 2021). The examples in (1) are some of the main forms available to French speakers when asking a *wh*-question. *Wh*-questions like (1a) will be described as *in situ* (IS), (1b) as *fronting* (F), (1c) as *fronting with inversion* (FINV), (1d) as *fronting with est-ce que* (FESK), (1e) as *clefting*, and (1f) as *fronting with complementizer*.

- (1) a. Tu arriveras **quand** ?
you.NOM arrive.FUT.2SG when
- b. **Quand** tu arriveras ?
when you.NOM arrive.FUT.2SG
- c. **Quand** arriveras-tu ?
when arrive.FUT.2SG-you.NOM
- d. **Quand** est-ce que tu arriveras ?
when be.PST.3SG.EXPL COMP you.NOM arrive.FUT.2SG
- e. .C'est **quand** que tu arriveras ?
EXPL.be.PST.3SG when COMP you.NOM arrive.FUT.2SG
- f. **Quand** que tu arriveras ?
when COMP you.NOM arrive.FUT.2SG
- All sentences translate to “**When** will you arrive?”

Corpus-oriented research has highlighted diachronic change in the use of the main question types, with the in-situ variant becoming more and more frequent in French in recent years (totalling more than 50% of productions around the year 2000), and fronting with inversion being used less and less (totalling less than 10% of productions) (Adli, 2015; Hamlaoui, 2009; Thiberge et al., 2021). This, however, contradicts what is still found in reference grammars for

French such as Riegel et al. (2014) (i.e., grammars used to develop teaching programs for L1 French children in school curricula), where the default interrogative variant is said to be fronting with inversion while in-situ sentences and all other variants are deemed *informal varieties*, which are purportedly used in spoken French only. In total, more than 15 pages are dedicated to French interrogative strategies in this book, but so-called “non-standard” variants (i.e., all variants but fronting with inversion) are described in a mere 10 lines of text. This is not just a niche phenomenon being exaggerated; rather, it must be considered in relation to the proportions of variants found in textbooks intended, for example, for L1 learners of the French language, where inversion is highly present (at least 40%) (Reinhardt, 2019b). In these textbooks, inversion is often presented as the “standard” way of forming interrogatives for children learning how to best use their own language.

The reasoning with respect to the use of in-situ variants, in particular, that is presented in Riegel et al. (2014) is that inversion is more complex and, therefore, disfavored when speaking, with a preference for the canonical word order of declarative sentences: Subject-Verb-Object.¹ The question of how internally “complex” each variant is, and of whether inversion (VSO order) is indeed more complex than non-inversion (SVO), has been addressed in many syntactic works. A common analysis is that the canonical word order (in-situ type) is the base order of constituents, from which all other types are derived. Generative grammars propose a “movement” explanation (Chomsky, 1981; Donati, 2006; Falk, 2012; Kayne, 1981; Ross, 1967; Stockwell, 1977). In this perspective, the interrogative element moves from its original position – leaving a trace – into the sentence’s left periphery (specifier of the CP, complementizer phrase), which is assumed to lead to increased complexity of the fronted question types compared to in situ, and even more so when a verb-subject inversion occurs. Other frameworks such as HPSG (Sag, 2010) or unification grammars (Abeillé, 2007), have described the phenomenon without movement, while still representing the non-locality in different ways that may or may not add to processing complexity. Construction grammar (e.g., Goldberg, 1995; Van Trijp, 2014) and, arguably, Combinatory Categorical Grammar (Steedman & Baldridge, 2011) can represent non-local dependencies without necessarily adding cognitive complexity, and, in this view, frequency alone could, for example, explain the use of the in-situ question type. The increased derivational complexity of fronted types postulated by some theoretical frameworks (highest for fronting with inversion; see Jakubowicz, 2011) and the “simplification” that the in-situ type represents in Riegel et al. (2014) are, we believe, relevant here. The form put forth as “standard” could be

¹ Riegel et al. (2014, p.680): “Ainsi, la langue parlée familière simplifie les structures pour aboutir à une certaine unité de l’interrogation : l’intonation joue un rôle essentiel et la phrase garde l’ordre habituel de la phrase déclarative.” Rough translation (our own): “Thus, the informal spoken language simplifies structures to bring some harmony in interrogation: prosody plays a crucial part and the sentence keeps the usual word order from declarative sentences.”

argued to be cognitively more demanding, which would make it more prestigious to master and use.

2.2 Using acceptability judgments to target the discrepancy between so-called standard French and corpus frequency of the variants

To shed light on the usage of the different interrogative variants, Thiberge et al. (2021) analyzed 617 French *wh*-interrogatives that were extracted from the ESLO project (Eshkol-Taravella et al., 2011; LLL, 2017). The ESLO corpus gathers around 5 million tokens, for more than 400 hours of recording in a wide variety of contexts (from scripted interviews to primary school classes, public conferences, or street recordings), with two time periods of data collection: one in the 1960s and the other around 2010. A key finding, among others, was that in the more recent period (second period of data collection), speakers aged 35–55 used fewer in-situ sentences than speakers aged 15–25 (~ 55% of productions vs. ~ 70% of productions). Moreover, a very marginal proportion of interrogatives were fronting-with-inversion sentences (less than 9% for both age groups).

This finding was modulated by the *context* of the recording: the difference between age groups, while visible when comparing interviews of L1 speakers by a researcher (46% of in-situ sentences for 35–55 y.o. speakers, ~ 60% for 15–25 y.o.), disappears in recordings conducted during family gatherings at meal-time (72% vs. 69% of in-situ sentences, respectively). These two contexts differ with respect to the presence of a researcher in the room during the recording, and also with respect to the topics that are discussed. During interviews, the researcher is asking questions following a scripted questionnaire relating to the personal lives of the interviewees, while during mealtimes, the topics cover daily activities of all members of the family. The formality of the social situation is clearly different between contexts, and we argue that this is a key element to understanding the different linguistic behaviors across age groups: all speakers use fewer in-situ variants in formal contexts (interviews) than in informal contexts (mealtime gatherings), but speakers aged 35–55 seem more prone to adapt their productions to the change in formality (as a reviewer points out, the reverse view might hold, with younger speakers being less sensitive to formality differences and seeing more situations as informal than older speakers). In a view where normative pressure is taken into account, this is consistent with sociolinguistic work on *age-grading* (Wagner, 2012), according to which this age group is most likely to be affected by linguistic prescriptivism as a consequence of work-related needs to appear well educated and literate.

All in all, recent corpus data show that only a small proportion of the interrogatives found in modern corpora are fronting-with-inversion constructions while in-situ sentences account for around 50% of the productions, or even more depending on the context. This contrasts sharply with the variant that is promoted as the “standard” way of speaking French (i.e., fronting with inversion) in reference grammars of French (e.g., Riegel et al., 2014). Moreover, this discrepancy

between corpus data and the prescriptive linguistic norm makes French *wh*-interrogatives a particularly interesting test case. Acceptability is typically correlated with the probability of the occurrence of an utterance. However, depending on the social situation and the speaker, this prediction may not hold here (Lau et al., 2017).

The experiments we report below shed light on different factors relevant to the acceptability of linguistic variants. Although we acknowledge that reference grammars are far from being a consistently useful tool for linguists, they are used for L1 and L2 teaching and provide insights into what is considered standard in a given language. In fact, by exploring the issue of *why a reference grammar would describe a language (here, French) in a view so inconsistent with what is observed in corpus data*,² we provide evidence that L1 speakers have the competence to apply a more variable take on the notion of overall acceptability, which is readily available when given the proper tools to express it. Using the acceptability judgment paradigm, we focus on the issue of how speakers perceive and use linguistic variants when, prescriptively, one variant is considered “standard”. With three acceptability judgment tasks, we assessed the preferences of French L1 speakers with regard to the four main variants of interrogatives: in situ, fronting, fronting with inversion, fronting with *est-ce que*. This approach will also contribute to resolving a dearth of experimental data on the use of interrogative variants in contemporary French. It builds on recent fine-grained corpus work on the usage of French *wh*-interrogatives in different contexts of interaction (Thiberge et al., 2021).³

Based on the available data, we anticipate that one-dimensional acceptability judgments will translate to higher ratings for the normative standard variant, that is to say fronting-with-inversion-type interrogatives, if the relationship between the prescriptive norm and acceptability is deeply entrenched in the speaker’s mind. If, on the contrary, acceptability assessments correlate with frequency (in a radical approach to frequency, such as Culicover et al. (2022), but see Bader and Häussler (2010) for a more nuanced account, as mentioned above), we would expect the most common form in contemporary French, the in-situ-type interrogatives, to yield higher results. This is what we aimed for in the first task (Section 3), where we contrasted fronting-with-inversion, fronting, and in-situ interrogatives, and where participants only had one general acceptability scale to provide their assessment.

With two other acceptability judgment tasks (Section 4), we aimed at further distinguishing between what “good French” and “suitable French” are for L1 speakers, which we extended to the fronting with *est-ce que* variant. For this, we used two different scales. The first scale targeted the normative well-formedness of the variants: i.e., “how does a sentence reflect what the ‘rules’

² See Abeillé and Godard (2021) for a more recent take on the diverse uses of French interrogative variants.

³ The experiments in this paper also build on preliminary studies where small experiments gave some insight into the preferences of L1 French adult speakers concerning some of the available variants in this language. See Thiberge (2018) for more details.

say French should be?”. The other scale targeted the suitability of the variants in a (short) context we provided and manipulated: i.e., “is it okay to use this particular sentence given the social situation?”, in line with Labov et al.’s (2011) sociolinguistic monitor, which assesses the incongruity of linguistic variants in socially situated interactions. We predicted that, with these two different scales, the judgments would better reflect how the linguistic norm weighs on the participants’ preferences. More specifically, we predicted that with these dichotomized tools, the general acceptability pattern exhibited in the first experiment would project onto the well-formedness scale, which corresponds to the prescriptive view of what the standard way of speaking French is or should be, while a more nuanced pattern would be seen on the suitability scale, which better fits with patterns seen in corpus data. As such, fronting-with-inversion sentences should yield higher ratings on the first scale, compared to other variants, while in-situ sentences should yield higher ratings than fronted variants on the second scale, with modulations from external factors such as context (manipulated independent variable) and age group.

3 Experiment 1: Written acceptability judgment task

The first experiment we conducted was an acceptability judgment task in which we assessed the general preferences of adult French L1 speakers with respect to the three main interrogative variants found in the corpus data (Adli, 2015; Hamlaoui, 2009; Thiberge et al., 2021). This was our first experimental factor, *TYPE*, with three levels (fronting, fronting with inversion and in situ). We added a *FORMALITY* factor, in which we contrasted *formal* and *informal* contexts. The manipulation of both these factors was meant to reflect the variability found, for example, in Thiberge et al. (2021), with a design allowing for a more systematic approach to the alternation phenomenon. Here, we anticipated an effect of *TYPE* consistent with previous findings (fronting-with-inversion sentences yielding higher acceptability ratings than the other two types). For *FORMALITY* we hypothesized, based on the corpus data mentioned above, that the difference between fronting-with-inversion sentences, on the one hand, and both fronting and in-situ sentences would be smaller in informal contexts. We also integrated a binary *AGE* parameter into the analysis (participants younger than 30 vs. older than 30) to explore whether our results would parallel those from the corpus data, where different age groups exhibited different linguistic preferences in contexts that differ in terms of formality.⁴

⁴ Moreover, in our materials, two variables that will not be explored further in the main text are present (academic level of the participant and, as suggested by a reviewer, argument type of the *wh*-element). Both of these variables were uncontrolled for in the construction of the experiments. There is a wide variety of academic profiles throughout the three experiments, with no satisfying way of grouping participants in meaningful categories, so it is, in our opinion, not usable (even though it could, indeed, have offered a deeper view of age effects and of our participants’ relationship(s) to the linguistic norm, depending on their education profile). The second variable (argument/adjunct *wh*-element) is more balanced across items, but shows no meaningful influence on the judgments. All this is accessible in the materials and scripts.

3.1 Notes on statistical analyses

With respect to the inferential statistics, for the statistical analyses of all experiments, we ran Bayesian regressions. A more detailed explanation can be found in the Supplementary Materials, but the main motivation for this is that we wanted to stay as close to the data as possible (thus, not working on a null hypothesis to be falsified, as is done in frequentist modeling.) (See Sorensen et al. (2016) for a more detailed view on the advantages of Bayesian statistics for linguists, psychologists and the cognitive sciences in general.) Bayesian cumulative link models were run for each experiment (see the Supplementary Materials for full specification and outputs), to adequately assess the link between our dependent variable (ordinal ratings, in all cases) and our independent variables (TYPE, FORMALITY, and AGE). Maximal models were run with a full-fledged random-effects structure, integrating participants and items as potential sources of variability in the data as well as random slopes (Barr et al., 2013). All independent variables were mean-center coded to allow a more direct interpretation of the model estimates (Brehm & Alday, 2022). In the present study, we used the *brms* package (Bürkner, 2017, 2018; Carpenter et al., 2017) in R version 4.2.1 (Posit team, 2023; R Core Team, 2023).

Finally, a crucial part of the Bayesian analysis lies in how credible and how probable it is that the estimates ($\hat{\beta}$) are higher or lower than 0 (= no effect when looking at the difference in conditions). Since there is no official consensus on what the thresholds should be in Bayesian inferential analyses, and since a high inter-participant variability can be expected with regard to how normative pressure influences acceptability judgments, we decided to take slightly larger thresholds for the interpretation of the output of our models. As such, and inspired by, a.o., Burnett and Pozniak (2021) and Granlund et al. (2019), we will report probabilities ($P(\beta < 0)$ or $P(\beta > 0)$) that are $\geq .90$ as reliably indicative of an effect of the independent variables on the dependent variables (“robust” evidence for an effect) and probabilities that are $\geq .80$ simply as evidence for an effect. Probabilities lower than .80 will be taken as not supporting the existence of an effect.⁵ For each relevant estimate, we will also report 95% Credibility Intervals (95%CrI), which are the values of the posterior distribution calculated by the model between which there is a 95%-chance that the true value for the estimated effect lies.

3.2 Materials, protocol and participants

For this 3×2 design, we constructed 30⁶ fictitious dialogues between a person A and a person B, all of which followed the same pattern: person A sets a very short illocutionary context varying in formality, then person B asks a *wh*-question relevant in this context. Context formality was modulated by three elements, two of which were combined in each A-sentence: dislocation of

⁵ Crucially, they will not be considered as supporting the absence of an effect, either.

⁶ See the Supplementary Materials for a full list.

an element in the sentence (*Jean arrive demain* vs. *Eh, Jean il arrive demain.*), colloquial/neutral lexical alternatives (*mon frère* vs. *mon frangin*), and absence/presence of a full-fledged negation (*Personne ne m'a rien donné* vs. *Personne m'a rien donné*). We created these alternating contexts⁷ to mimic the contrasting contexts examined by Thiberge et al. (2021). In the first context, which we call *informal* (corresponding to family interactions in the ESLO corpus), there is minimal social distance between interlocutors. This is achieved through the use of colloquial or low-register nouns (e.g., *frangin* for *frère*), the breaking of prescriptive norms (such as the absence of full-fledged negation), and orality markers (like subject or object doubling, which is typically avoided in standard French). In this context, we anticipated that non-standard interrogative variants would be less incongruent, as per the sociolinguistic monitor model proposed by Labov et al. (2011). By contrast, in the second context, which we call *formal* (corresponding to interviews with researchers in the ESLO corpus), no proximity markers are present. Instead, neutral nouns, full-fledged negation, and the absence of doubling create the right conditions for greater stigmatization of the non-standard variants and, conversely, greater adequacy of the standard default form. **Table 1** shows an item in all its conditions.

Table 1: Example item from Experiment 1.

TYPE	FORMALITY	
	<i>informal</i>	<i>formal</i>
Fronting	A: Ouais, moi je me barre demain.	A: Je pars demain.
(F)	B: Où tu vas ?	B: Où tu vas ?
Fronting with	A: Ouais, moi je me barre demain.	A: Je pars demain.
inversion (FINV)	B: Où vas-tu ?	B: Où vas-tu ?
In situ	A: Ouais, moi je me barre demain.	A: Je pars demain.
(IS)	B: Tu vas où ?	B: Tu vas où ?
	A: <i>Yeah, me, I'm leaving tomorrow.</i>	A: <i>I'm leaving tomorrow.</i>
	B: Where are you going?	B: Where are you going?

The 30 target dialogues were mixed with 30 filler dialogues with the same A-B structure, containing no *wh*-interrogative and no variation in formality. We opted for a 1:1 ratio for fillers and target items, so as to not lengthen the experiment and to avoid satiation effects. After 3 practice items, the 60 items were presented in a Latin Square design with full randomization. The experiment was carried out on the Internet via the IbexFarm platform (Drummond, 2016). After a short introduction to the participants' rights, a description of the procedure, and a short questionnaire about their linguistic profile, explicit consent was required to proceed.

⁷ See the Supplementary Materials for the results of a norming experiment for this formality parameter.

The participants were then presented with the 63 (practice + target + filler) items. Below each sentence, participants rated the acceptability of the B-sentence on a 0–10 numeric scale, with each point represented in a numbered box (0-1-2..8-9-10). *Acceptability* was here defined as both the general understandability of the sentence and its well-formedness, given the short context provided. Once a participant clicked on the scale or entered a digit on their keyboard, the next item was presented. There was no time limit set for each trial, and the whole experiment was estimated to take roughly 15 minutes. Participants were informed that there was no good or bad decision and that we were only interested in their intuitions.

Participants were recruited via the RISC network (CNRS, UMR 3352) and social networks. 44 self-identified L1 speakers of French took part (age: 19–80, mean = 39.8, median = 27.5). The data for target items amounted to 1320 (30*44) ratings.

3.3 Results

Figure 1 presents a violin plot of the answers for the three TYPE conditions, and shows the general pattern of results (error bars = standard error; white lines = quartiles).⁸

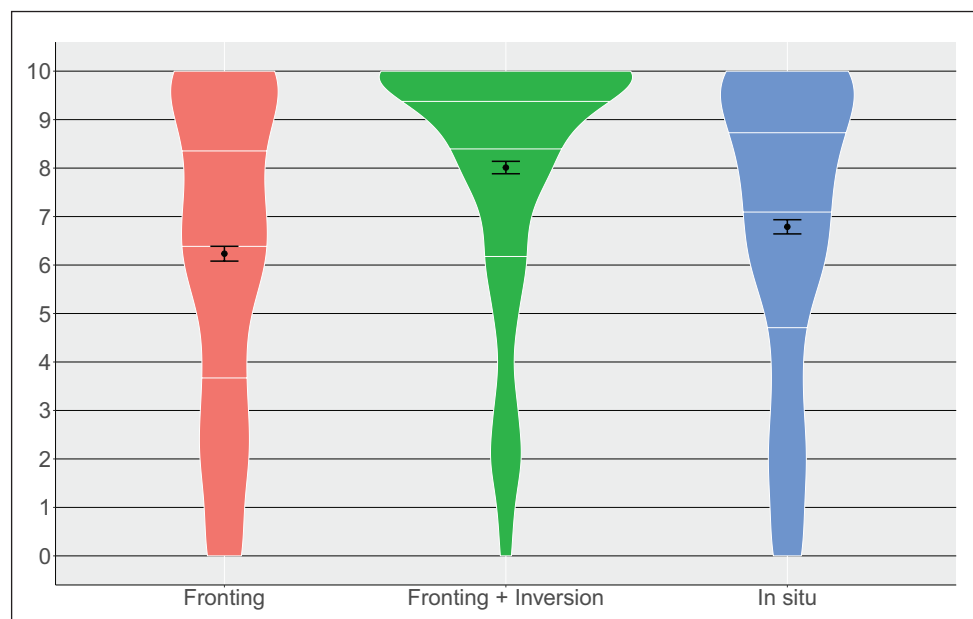


Figure 1: Mean acceptability and distributions of ratings from Experiment 1.

⁸ For ease of presentation, and because we are more interested in how the different variants pattern with each other than in absolute numbers, we will mostly present graphical representations of the distribution of judgments yielded by the different scales in the experiments. Although the scales we use are of ordinal nature, we follow the tradition in much psycholinguistic work in representing judgments as continuous on the graphs (with means, error bars, and violin plots to keep track of the distribution of answers across all possible levels of the answer scales). The detailed and more precise numeric values behind these graphs can be found in the Supplementary Materials.

Fronting-with-inversion-type interrogatives generally received higher ratings than fronting-type and in-situ-type interrogatives. The Bayesian cumulative-link modeling of the data (see **Figure 4** below for more, and model mb1 in the Supplementary Materials) yields robust evidence for an effect of the TYPE predictor here, both for the fronting vs. fronting-with-inversion contrast ($\hat{\beta} = -2.01$, $P(\beta < 0) = 1$, 95%CrI = [-2.91,-1.15]) and for the in-situ vs. fronting-with-inversion contrast ($\hat{\beta} = -1.52$, $P(\beta < 0) = 1$, 95%CrI = [-2.27,-0.81]).

Figure 2 details the results depending on the two levels of the manipulated FORMALITY variable. Here, the overall pattern distinguishing the three syntactic types is similar across conditions, but the contrast between fronting-with-inversion sentences and the other two types is larger in the formal condition than in the informal condition.

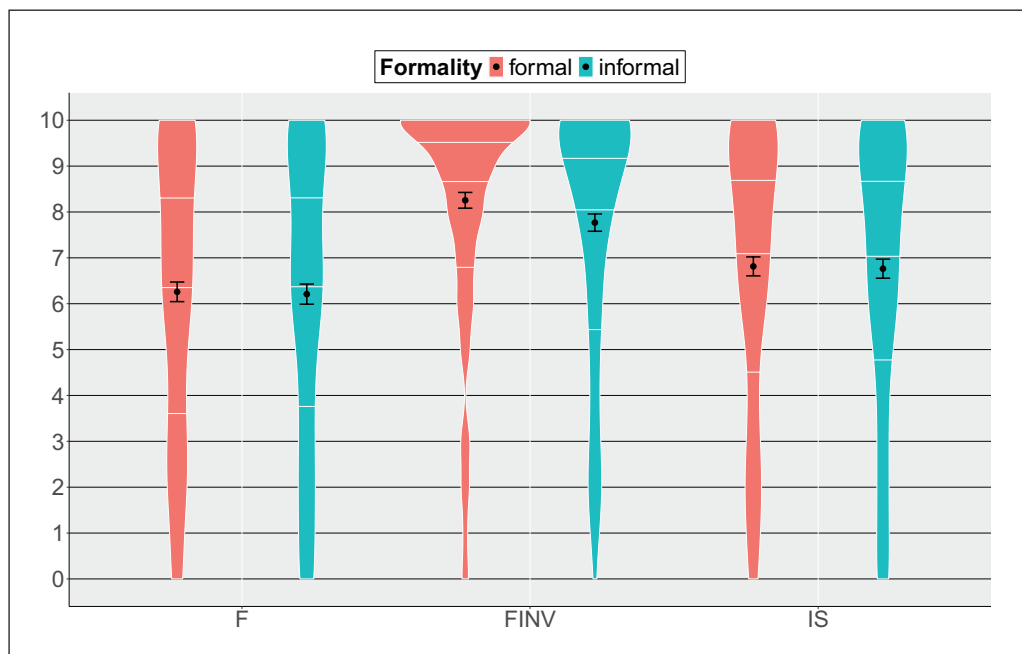


Figure 2: Mean acceptability and distributions of ratings from Experiment 1, by context.

While ratings associated with F-type and in-situ-type sentences stay roughly equivalent between contexts, fronting-with-inversion-type sentences are associated with higher ratings in the formal condition than in the informal one. The Bayesian model yields evidence for a TYPE*FORMALITY interaction for the fronting vs. fronting-with-inversion contrast ($\hat{\beta} = -0.37$, $P(\beta < 0) = 0.87$, 95%CrI = [-1.02,0.29]) and robust evidence for the in-situ vs. fronting-with-inversion contrast ($\hat{\beta} = -0.45$, $P(\beta < 0) = 0.93$, 95%CrI = [-1.06,0.17]).

Figure 3 details the results depending on the age group of participants. As an important note, we took the split into a +30 and a -30 group and not based on the median age. This

follows from corpus studies, in which age group was a meaningful predictor of interrogative type production (Thiberge et al., 2021), but also more generally from the sociolinguistic literature. Absolute age is often an important predictor of attitudes toward linguistic variation in itself, because it correlates with many other socioeconomic circumstances or group affiliation. As mentioned above, this is an important element of *age-grading* phenomena: people falling into the age category where professional obligations heavily constrain a person's relationship to the linguistic norm (Wagner, 2012). However, since the age variable was not controlled for, the groups are not fully balanced in number ($N_{+30y.o.} = 19/44$). The overall contrast between the three syntactic types of interrogatives remains highly similar across age groups, but the contrast is larger between fronting-with-inversion sentences and the other two types for participants older than 30.

The difference between mean ratings associated with fronting vs. fronting-with-inversion sentences is larger in the +30 group for which the model yields evidence (for the TYPE*AGE interaction, ($\hat{\beta} = -1.57$, $P(\beta < 0) = 0.97$, $95\%CrI = [-3.23, 0.04]$)). This contrast is even larger for ratings associated with in-situ vs. fronting-with-inversion sentences, for which the model yields robust evidence (TYPE*AGE interaction at ($\hat{\beta} = -1.67$, $P(\beta < 0) = 0.99$, $95\%CrI = [-3.09, -0.26]$)).

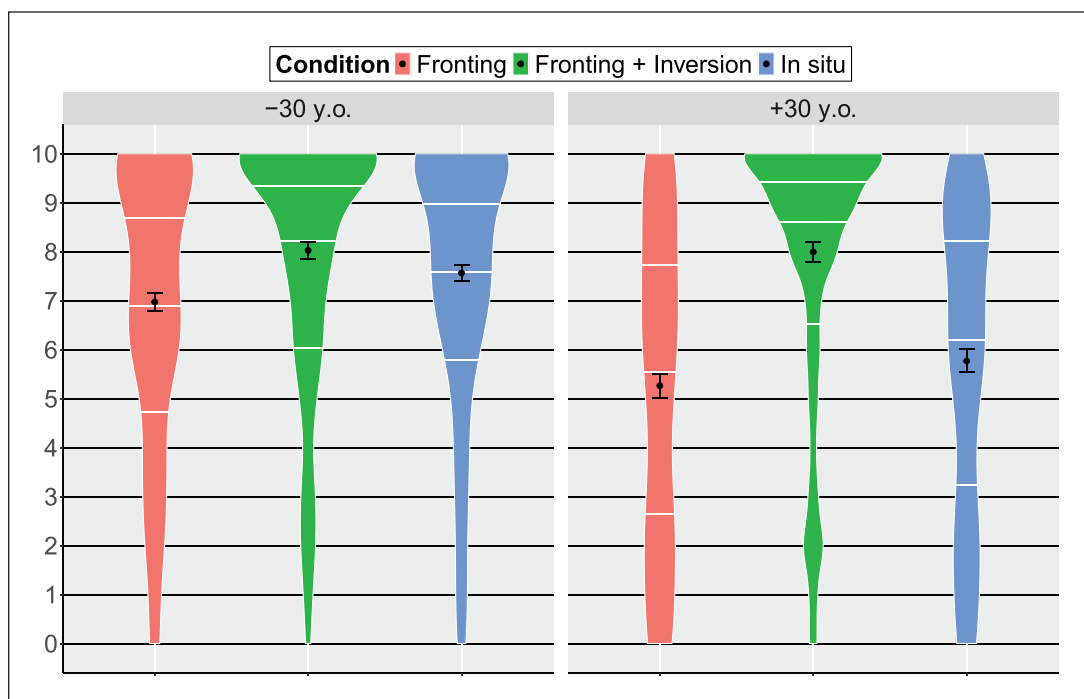


Figure 3: Mean acceptability and distributions of ratings from Experiment 1, by age group.

Figure 4 shows the posterior distributions for the maximal mb1 model fit to the data from Experiment 1. In Bayesian inferential statistics, posterior distributions are simulated for each predictor (parameter) specified in the model, which help characterize the influence of said predictor on the dependent variable. Each posterior distribution is described by its position relative to 0 (= no influence of the predictor on the dependent variable), its range (extreme values and credible intervals, within which the true value of the estimated coefficient lies), and its mean value (estimated coefficient, just as in frequentist models). As such, posterior distributions where the true value of the estimated coefficient has a high probability of being different than 0 give evidence for an effect of the predictor. In addition to the results described above, the model shows robust evidence for a main effect of FORMALITY ($\hat{\beta} = 0.26$, $P(\beta > 0) = 0.94$, 95%CrI = [-0.08,0.60]), meaning that, overall, the ratings given by participants were higher in the formal than in the informal contexts. From the TYPE*FORMALITY interactions described before, this follows from the higher ratings associated with fronting-with-inversion-type sentences in formal contexts.

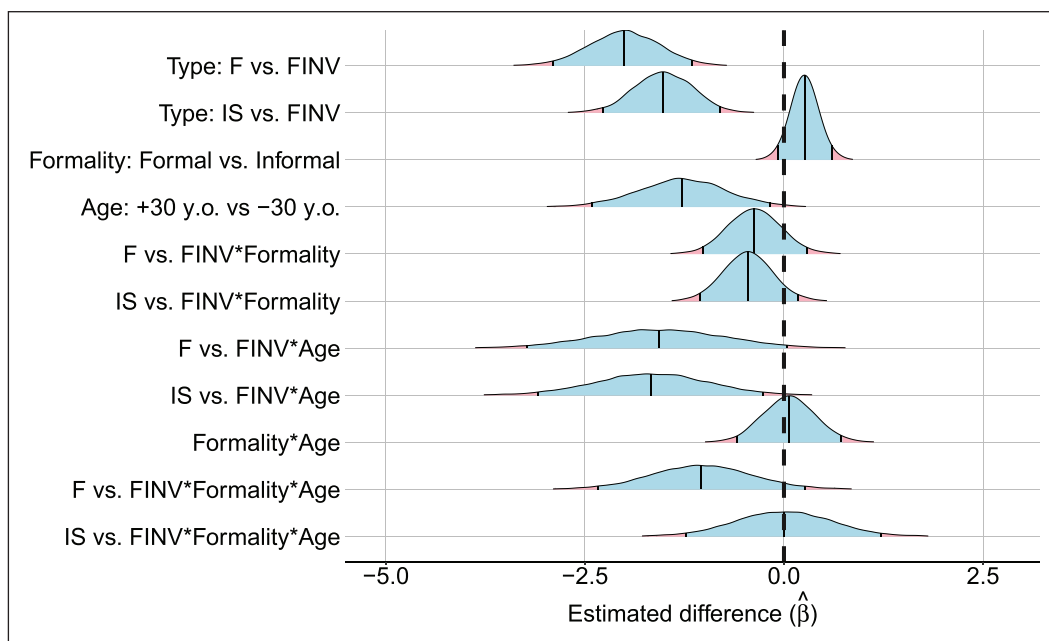


Figure 4: Posterior distributions for the maximal model ran for Experiment 1.

Perhaps of greater interest is the TYPE*FORMALITY*AGE interaction for fronting vs. fronting-with-inversion sentences ($\hat{\beta} = -1.04$, $P(\beta < 0) = 0.95$, 95%CrI = [-2.33,0.26]). This means that there is robust evidence for a higher sensitivity to context variation among the +30 y.o. group: F interrogatives are rated even lower in formal contexts by this population than by the -30 y.o. participants, compared to fronting-with-inversion sentences.

3.4 Interim summary

Experiment 1 offers a deeper look into the role of context, based on the general acceptability of different *wh*-interrogative variants of French. A general pattern of acceptability, compatible with a normative take on what standard French should be, is confirmed here: fronting with inversion yielded higher ratings than in situ, and in situ yielded higher ratings than simple fronting. These results are compatible with earlier preliminary experimental work (Thiberge, 2018) on this topic but they are mainly on a par with many prescriptive or reference grammars of French (e.g., Riegel et al., 2014), where the fronting-with-inversion variant is held as the standard for French.

With respect to FORMALITY, we found evidence for an influence of context on this pattern: in formal contexts, fronting with inversion yields higher ratings than in informal contexts. This is directly reminiscent of corpus findings, such as Thiberge et al. (2021), where the uses of each variant were different depending on the context (more fronting-with-inversion sentences in formal contexts with an interviewer or at school than at family meals). Preferences were, however, not reversed, with in-situ sentences becoming the preferred option. We also find evidence for an effect of age, with participants older than 30 years being more sensitive to the difference between variants overall, and giving lower ratings than participants younger than 30 to the two prescriptively “non-standard” variants (simple fronting and in situ). A difference in linguistic production between participants aged 15–25 vs. participants aged 35–55 was also observed for the *wh*-interrogative variants in the ESLO corpus.

Also in line with these corpus findings, we find in Experiment 1 an interaction between FORMALITY and the AGE GROUP of participants. When compared with participants younger than 30, participants older than 30 seem to more strongly contrast some variants (simple fronting and fronting with inversion) when the context is formal than when it is informal. This could come from a different approach to the linguistic norm as a function of age, and possibly as a function of professional circumstances. **Figure 5** illustrates how this could work: when considering age as a continuous variable, the in-situ variant seems more acceptable in formal⁹ contexts for participants younger than 30 and older than 60–70. There is no such phenomenon for the simple fronting variant, which becomes less and less acceptable with increasing age; hence, the TYPE*FORMALITY*AGE interaction described above. This U-shaped curve for in-situ sentences (also visible for fronting-with-inversion sentences, but reversed and to a lesser degree) could be linked to the notion of age-grading, as proposed in sociolinguistic works such as Wagner (2012). Speakers aged between 30 and 60/70 are confronted with the necessity of abiding by linguistic prescriptive norms in their

⁹ As pointed out by a reviewer, it is somewhat surprising that the age-grading result is not mirrored in the informal condition. For instance, it could be expected that middle-aged participants would be less affected by normative pressures in the informal condition, leading to a flatter line there, which does not happen. The drop in acceptability for in-situ sentences in the informal condition for the older participants is surprising. However, the age variable was not controlled for, and a more systematic approach to this is needed in further research (in particular, there are only 5 people older than 70 here, and this low number calls for caution when comparing both sides of the U-shaped curve.)

everyday professional lives (i.e., “fronting-with-inversion type is the standard variant”), and this may be reflected in their acceptability ratings. The age effect found in this first experiment is also a reason for not overestimating the representativity of studies that generally include a majority of young adults (university students) as participants, in particular, when normativity may play a role.

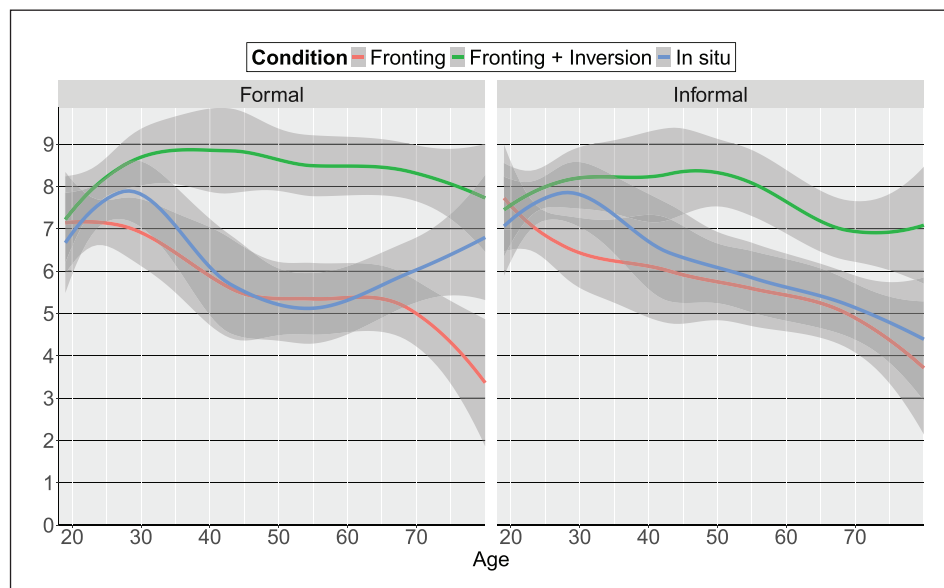


Figure 5: Mean acceptability ratings from Experiment 1, by age (continuous).

4 “Good” French and “suitable” French – Experiments 2 & 3

The results from Experiment 1 are consistent with the hypothesis that preferences for different variants vary depending on the context. Experiment 1 also provided evidence for a general confusion between *acceptability* and *compatibility with the normative standard for a language*. Contrary to frequency-driven accounts of acceptability (e.g., Culicover et al.’s (2022) account, which links (un)acceptability to surprisal), the preferences we found also directly contradict evidence from corpora, with the least used variant in spoken corpora (fronting with inversion) yielding the highest ratings. We devised Experiment 2 as a way to better explore this apparent contradiction.

4.1 Materials, protocol and participants (Experiment 2)

With the exact same items as in Experiment 1 (practice, targets and fillers), we investigated the acceptability of different interrogative types with three main changes. First, we changed the modality of stimulus presentation. Experiment 1 was a written text-based experiment, with participants reading sentences (dialogues between person A and B) and then having all the time they needed to develop a judgment and provide it on a unique response scale. In Experiment 2, we opted for an auditory presentation, with the reasoning that it made for a more ecological

setting, more consistent with the everyday interactions found in spoken corpora (e.g., ESLO). Instead of the written dialogue, participants saw an audio player that they had to trigger to hear the stimuli. They could replay them as often as they wished.

The second part we adjusted was the response scale, which this time was divided into two different questions:

- Is this sentence “good French”? (*Est-ce que c’est du “bon français”?*)
- Is this sentence “suitable French”? (*Est-ce que c’est du “français adapté”?*)

Both sub-scales were defined in the previously displayed instructions as, respectively, “well-formed French” and “you could speak like this in this context”.

A third adjustment relates to the number of response options on the scales. Experiment 1 was run with a 11-point scale (0–10), while this time the scales were both 6-point (1–6) scales. We chose this option for two reasons: Since participants had to provide judgments on two scales, we wanted to make their task easier by providing fewer options. We also wanted to exclude a neutral position to force participants to make a choice. From previous experience, we did not expect these changes to systematically affect our results. Answers were not mandatory, so participants who felt unable to decide on a scale could skip it for a particular item and continue with the experiment.

The auditory stimuli were recorded with 4 adult French L1 speakers (F1, F2, M1, M2), whom we recorded in a soundproof room. During the recording phase, speakers were instructed to go for “reserved naturalness”, without putting too much expressivity in their voice, so as to minimize prosodic contrasts across, and even within, the sounds played by participants in the actual experiment. They read all possible combinations of contexts and sentence types for all items, and we selected the recordings that we judged most “neutral” in terms of retrievable social cues. All sentences in all conditions were recorded and then manipulated in Audacity (AudacityTeam, 1999–2021) to create dialogues absolutely parallel to the written stimuli of Experiment 1. We crossed voices so that all 12 combinations (F1-F2, F1-M1, F1-M2, F2-F1, F2-M1, F2-M2, M1-F1, M1-F2, M1-M2, M2-F1, M2-F2, M2-M1) were presented for all items and all participants of Experiment 2. We also carried out a norming study to assess the “a priori gender” people generally assigned to voices (so as to minimize potential rating differences in terms of how women or men are supposed to speak to each other¹⁰), and to check that no voice was more associated with a colloquial tone than others, so as to not bias the FORMALITY variable. Data from this norming study can be found in the Supplementary Materials.

¹⁰ In no way do we mean here that women or men are supposed to talk to each other in a specific way, but that there might exist some internalized expectations in participants in this regard, due to age or to how an individual perceives what a formal or informal setting is, and we wanted to minimize this possibility across items, and, as such, to have a balanced presentation of stereotypical feminine/masculine voice combinations.

64 self-identified L1 speakers of French took part in total, both via the Internet and in the lab. They were recruited by the RISC network and social networks. Due to some technical issues with the presentation of some audio stimuli for some participants (usually a sound that could not be played by the server), we only kept answers from 54 participants who saw and answered at least 90% of all practice + filler + target items on both scales (age: 19–70, mean = 35.8, median = 30). This amounts to 1603 observations on each scale for target items.

4.2 Results (Experiment 2)

We analyzed data from Experiment 2 in the same manner as Experiment 1 (see 3.1 and the Supplementary Materials on OSF). However, this time we ran two models: one (mb2g) for the first scale, with “well-formedness” ratings as the dependent variable, and one (mb2s) for the second scale, with “suitability to the context” ratings as the dependent variable.

Figure 6 shows a pattern similar to that seen in Experiment 1. On the first scale, fronting-with-inversion-type interrogatives generally received higher ratings than both F-type and in-situ-type interrogatives. The Bayesian cumulative-link modeling of the data (see **Figure 12** at the end of the section for more, and model mb2g in the Supplementary Materials) yields robust evidence for an effect of the TYPE predictor here, for both the fronting vs. fronting-with-inversion contrast ($\hat{\beta} = -3.16$, $P(\beta < 0) = 1$, 95%CrI = $[-3.92, -2.42]$) and the in-situ vs. fronting-with-inversion contrast ($\hat{\beta} = -2.81$, $P(\beta < 0) = 1$, 95%CrI = $[-3.56, -2.11]$).

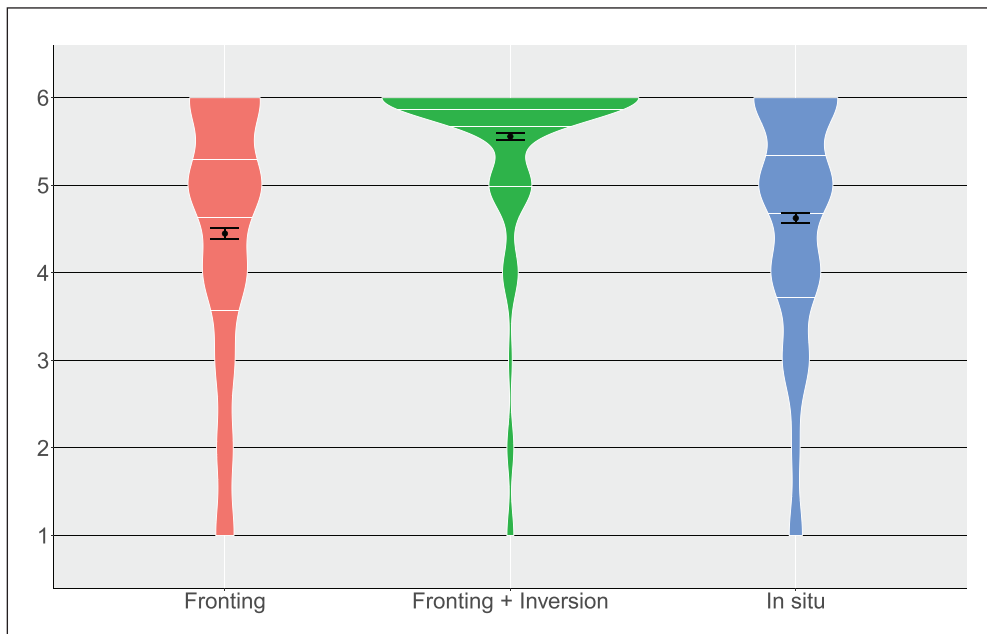


Figure 6: Mean acceptability ratings from Experiment 2 (good French).

However, **Figure 7** displays a different pattern. On the second scale, fronting-with-inversion-type interrogatives received ratings similar to F-type interrogatives, and in-situ-type interrogatives yielded higher ratings overall. The Bayesian cumulative-link modeling of the data (see **Figure 13** below, and model mb2s in the Supplementary Materials) yields robust evidence for an effect of the TYPE predictor for the in-situ vs. fronting-with-inversion contrast only ($\hat{\beta} = 0.45$, $P(\beta > 0) = 0.97$, 95%CrI = [-0.03,0.95]).

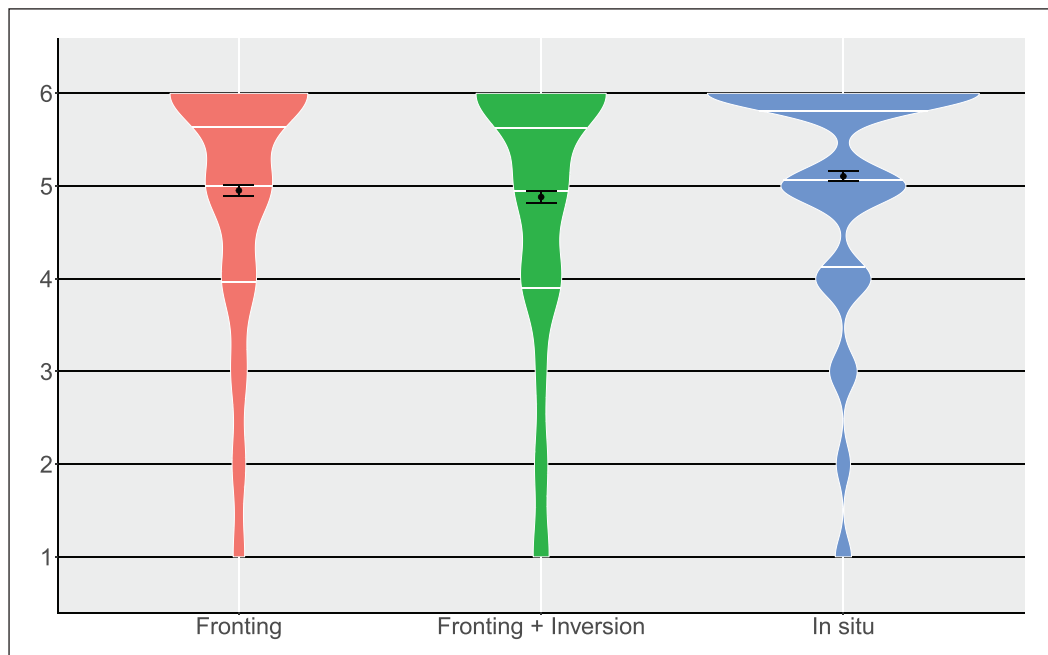


Figure 7: Mean acceptability ratings from Experiment 2 (suitable French).

For FORMALITY, the two scales also appear to capture different dimensions of the preferences of French speakers. For the “good French” scale, the effects of FORMALITY appear to be very similar to those observed in Experiment 1 (**Figure 8**): fronting-with-inversion sentences are rated higher than both other types, overall, and the contrast is larger in formal contexts (evidence for the fronting vs. fronting-with-inversion contrast ($\hat{\beta} = -0.35$, $P(\beta < 0) = 0.83$, 95%CrI = [-1.08,0.36]). The pattern is different on the “suitability” scale (**Figure 9**), with the three variants yielding ratings around 5/6 on this scale, whatever the context. However, when looking at the results in more detail, the type yielding the highest ratings in the formal contexts is in-situ, when fronting and fronting-with-inversion sentences yield roughly similar ratings, and in the informal contexts, fronting-with-inversion sentences yield the lowest ratings. The model run for this scale yields robust evidence that the fronting vs. fronting-with-inversion contrast is reduced in the formal contexts (TYPE*FORMALITY interaction, $\hat{\beta} = -0.42$,

$P(\beta < 0) = 0.90$, 95%CrI = [-1.08,0.24]), and also evidence with respect to the in-situ vs. fronting-with-inversion contrast (TYPE*FORMALITY interaction, $\hat{\beta} = -0.46$, $P(\beta < 0) = 0.87$, 95%CrI = [-1.29,0.36]). One could wonder why judgments on the “good French” scale would be influenced by context formality at all, if all it captures is the participants’ knowledge of prescriptive norms. One possible answer is that these norms allow for some degree of style-shifting, as a reviewer suggests. Another perspective on this would be that formal contexts can sometimes reinforce the perception of well-formedness of the standard form when it is used, with a near ceiling effect in that condition for the fronting-with-inversion type (mean rating of 5.66/6 and a smaller standard deviation than all other type*context combinations, on this scale and on the other scale).

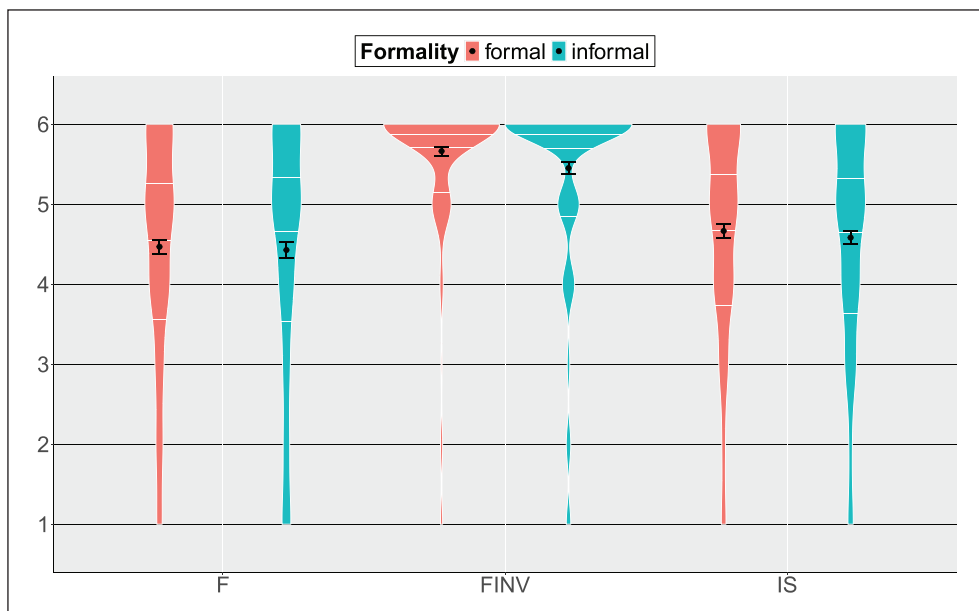


Figure 8: Mean acceptability ratings from Experiment 2, by context (good French).

For the AGE variable and on the “well-formedness” scale, the overall pattern does not change between age groups, but participants younger than 30 gave higher ratings to fronting-with-inversion sentences, compared to older participants (**Figure 10**). This results in evidence for a TYPE*AGE interaction for the in-situ vs. fronting-with-inversion contrast ($\hat{\beta} = 0.64$, $P(\beta > 0) = 0.82$, 95%CrI = [-0.78,2.06]). On the “suitability scale” (**Figure 11**), participants younger than 30 exhibit a stronger dispreference for fronting-with-inversion sentences (rated lowest) than participants older than 30 (who rated fronting-with-inversion types about the same as fronting types, but still lower than in-situ types). This is captured by robust evidence for TYPE*AGE interactions for both the fronting vs. fronting-with-inversion ($\hat{\beta} = -0.67$,

$P(\beta < 0) = 0.96$, 95%CrI = [-1.45,0.10]) and the in-situ vs. fronting-with-inversion ($\beta = -0.80$, $P(\beta < 0) = 0.96$, 95%CrI = [-1.75,0.13]) contrasts.

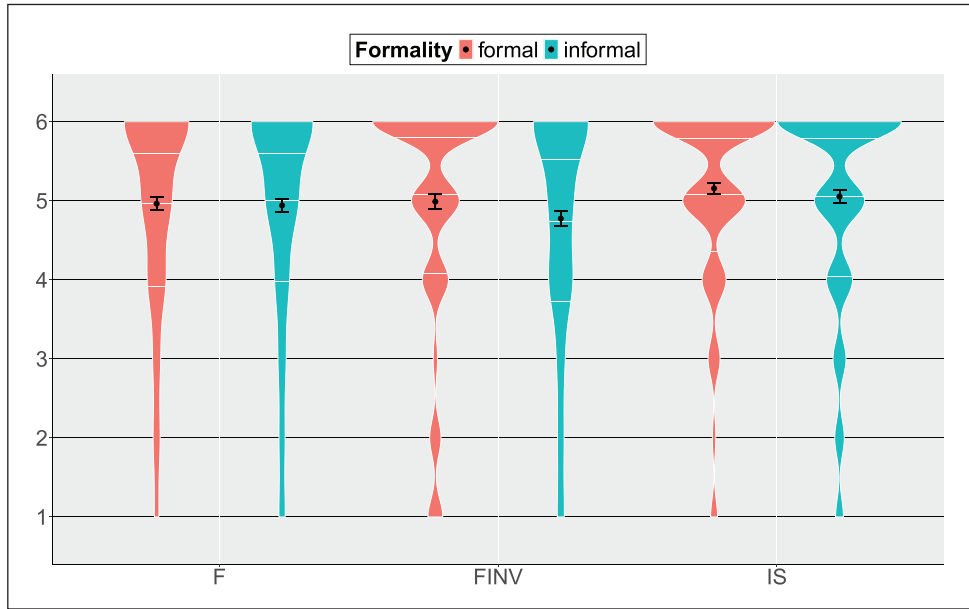


Figure 9: Mean acceptability ratings from Experiment 2, by context (suitable French).

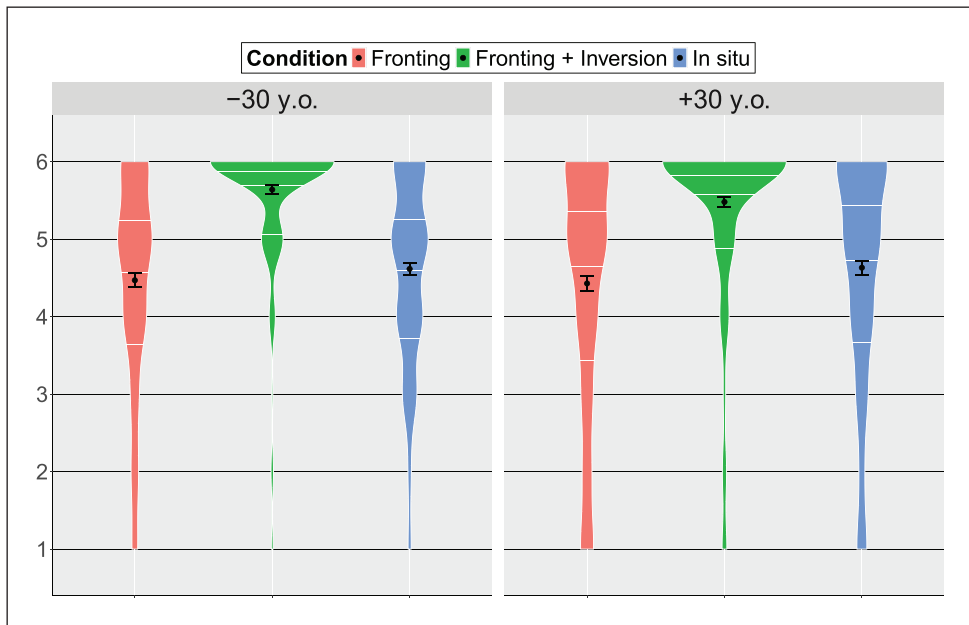


Figure 10: Mean acceptability ratings from Experiment 2, by age group (good French).

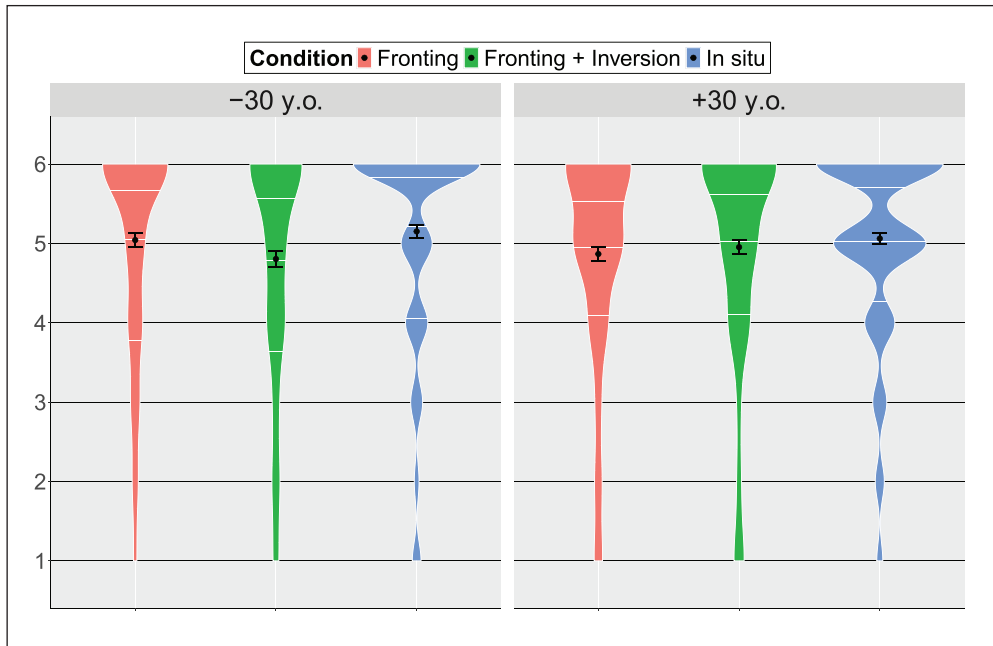


Figure 11: Mean acceptability ratings from Experiment 2, by age group (suitable French).

Figures 12 and 13 show the posterior distributions for the two models run for both scales of Experiment 2.

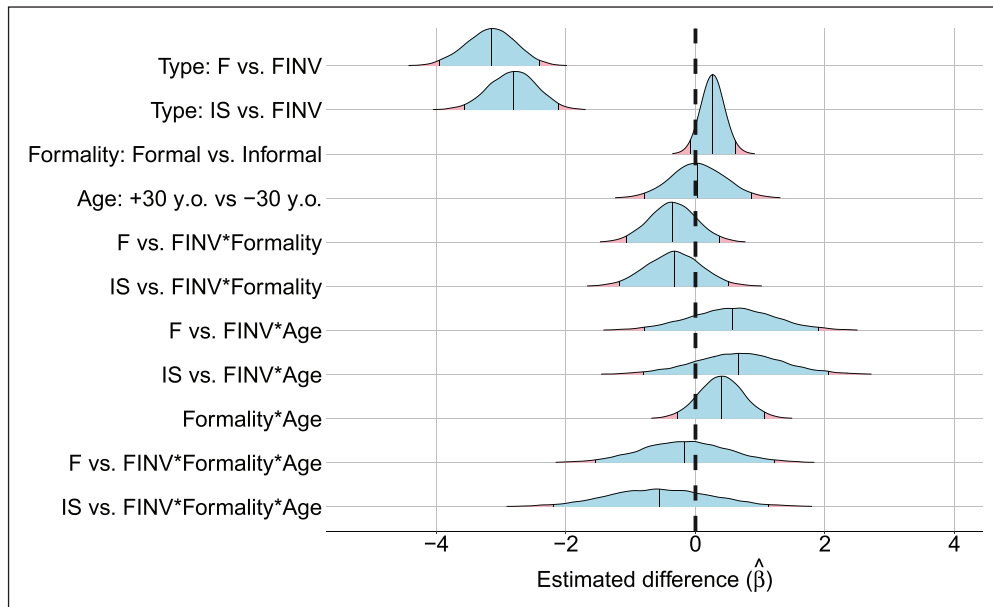


Figure 12: Posterior distributions for the maximal model run for Experiment 2 (good French).

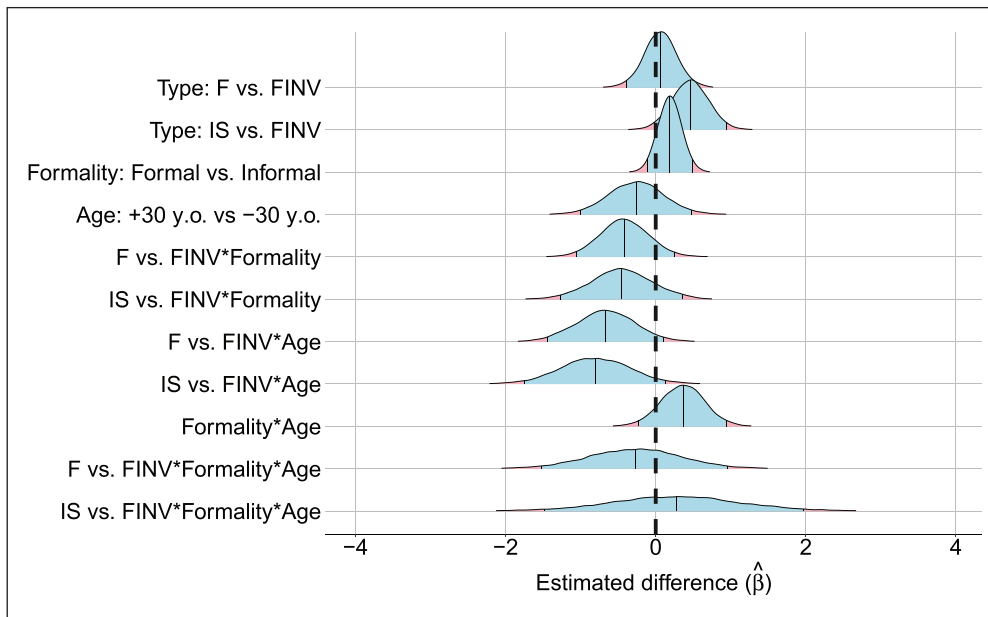


Figure 13: Posterior distributions for the maximal model run for Experiment 2 (suitable French).

4.3 Interim summary

Contrary to Experiment 1, there seems to be no 3-way interaction $TYPE*FORMALITY*AGE$ for either the fronting vs. fronting-with-inversion or the in-situ vs. fronting-with-inversion contrasts on either scale. Presenting participants with explicit scales where they could disentangle their judgments on well-formedness and suitability for a specific context (in terms of formality) seems to have neutralized, in some way, the differences between age groups with regard to formality assessments and, more precisely, with regard to preferences for some variants in different contexts.

Overall, Experiment 2 provides no evidence for an influence of the number of response options on the scales, as seen from the very similar patterns on the well-formedness scale here and on the unique scale used in Experiment 1. On this scale, fronting-with-inversion sentences are preferred to both fronting and in-situ sentences, and this preference is modulated by context formality.

More importantly, this experiment provides evidence for a differentiated approach by participants to the concept of acceptability. Well-formedness does not follow frequency of use in corpus data, and corpus frequency does not correlate with well-formedness alone. The pattern on the first scale (“good/proper French”) is close to what can be observed in a general acceptability task (i.e., Experiment 1) and closely reflects what can be found in a reference grammar of the language. However, this scale alone would miss participants’ ability to select which variant

would best suit a “real-life” everyday context of interaction. This is more clearly captured by the second scale (“suitability”), which participants seem to have rightly interpreted as “Would I, myself, use this sentence in this specific situation?”

Interestingly, the TYPE*AGE interactions are reduced on the well-formedness scale. This could be construed as evidence that participants have a normative view of what proper French should be (with regards to at least, but most probably not only, *wh*-interrogation). These interactions are still present on the second scale, which would indicate that not all age groups consider all variants as “suitable” for a given context. Although there is no evidence for a three-way interaction with the FORMALITY predictor in Experiment 2, this is still reminiscent of Thiberge et al.’s (2021) results, where different age groups use different variants in different contexts. Further exploration of what makes two contexts different and of how different age groups define and interpret formality would be useful for future work.

4.4 A quick look at the *est-ce que wh*-interrogatives (Experiment 3)

Experiment 2 gave a more fine-grained take on the acceptability of the three main variants available to French speakers for asking a *wh*-question. Like Experiment 1 however, it ignored the FESK variant, where an *est-ce que* sequence is introduced between the fronted *wh*-element and the Subject-Verb sequence. This variant is actually quite frequently used in everyday speech (about as much as the fronting-with-inversion variant in ESLO data, for example). At the same time, it does not seem much more complex than simple fronting, since it does not involve Verb-Subject inversion. We conducted a third experiment on the basis of the paradigm applied in Experiment 2 (two different 11-point slider scales, one for well-formedness and the other for suitability in a context, with written stimuli, however) to pinpoint how this variant compares to the others, by comparing it to in-situ and fronting-with-inversion sentences. The details of this experiment (participants, materials and detailed analyses) are presented in the Supplementary Materials, and, for the sake of clarity, we will only report the main results here, to provide a more exhaustive picture of the alternation phenomenon with respect to interrogative variants in French.

Figure 14 shows a pattern quite similar to that seen in Experiments 1 and 2. On the first scale, fronting-with-inversion-type interrogatives generally received higher ratings than both FESK-type and in-situ-type interrogatives. Interestingly, FESK interrogatives appear in an intermediate position between fronting-with-inversion and in-situ sentences.

Figure 15 displays yet a different pattern. On the second scale, fronting-with-inversion-type interrogatives received ratings similar to in-situ-type interrogatives, and F-type interrogatives yielded lower ratings overall. Although they are rated “better French” than in-situ sentences, FESK sentences generally seem not to be suitable for many contexts. It can also be noted that there is no difference here between fronting-with-inversion and in-situ sentences in terms of

suitability, which contrasts with what was observed in Experiment 2. A possible explanation could be that the presence of FESK sentences in the stimuli made the difference between fronting-with-inversion and in-situ less salient to participants.

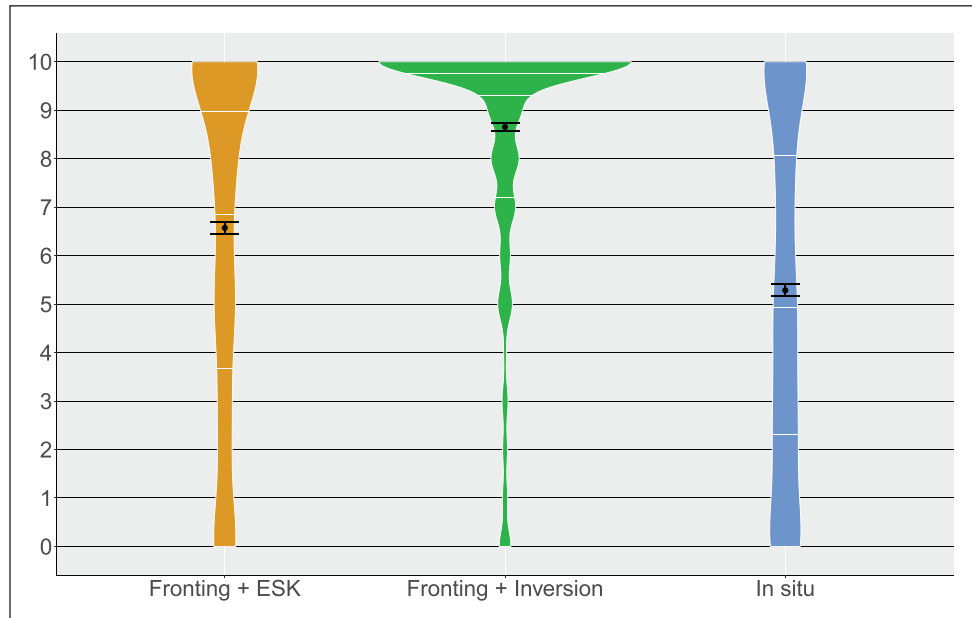


Figure 14: Mean acceptability ratings from Experiment 3 (good French).

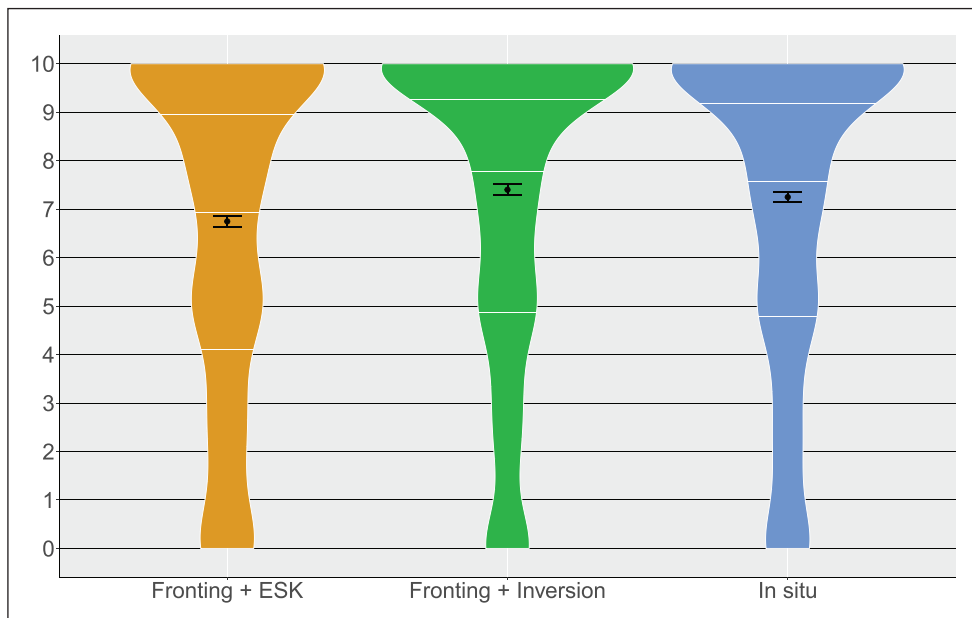


Figure 15: Mean acceptability ratings from Experiment 3 (suitable French).

For FORMALITY, and on the “good French” scale (**Figure 16**), the pattern appears similar again to that observed in the previous experiments: fronting-with-inversion sentences are rated higher than both other types, overall. However, this time, there is no evidence of an interaction TYPE*FORMALITY, for both the FESK vs. fronting-with-inversion and the in-situ vs. fronting-with-inversion contrasts. The reason may be that since FESK sentences are in an intermediate position between fronting-with-inversion and in-situ sentences, participants are less sensitive to the differences between contexts when they consider the overall “well-formedness” of the three variants, which differs from Experiment 2 where fronting and in-situ sentences were behaving in a similar way, for example.

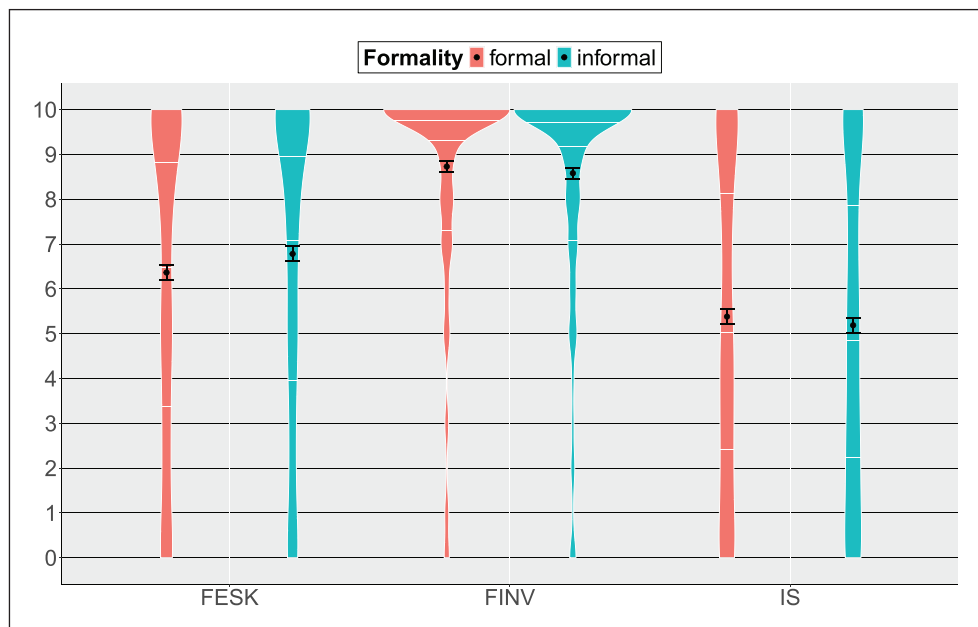


Figure 16: Mean acceptability ratings from Experiment 3, by context (good French).

On the “suitability” scale (**Figure 17**), the pattern is different, with the three variants rated between 6.56/10 (FESK sentences in informal contexts) and 7.74/10 (fronting-with-inversion sentences in formal contexts), so less contrast than on the first scale. Fronting-with-inversion sentences yield higher ratings in formal contexts, and FESK sentences yield lower ratings in informal contexts, while in-situ sentences appear to not be sensitive to differences in formality.

Also to be noted here, the fronting-with-inversion sentences receive similar ratings across contexts on the “good French” scale. On the “suitability” scale, however, the ratings are quite different for this type, which we take as further evidence that, although some style-shifting is allowed (see the results from Experiment 2 above), the first scale is actually more capturing participants’ knowledge of the linguistic norms.

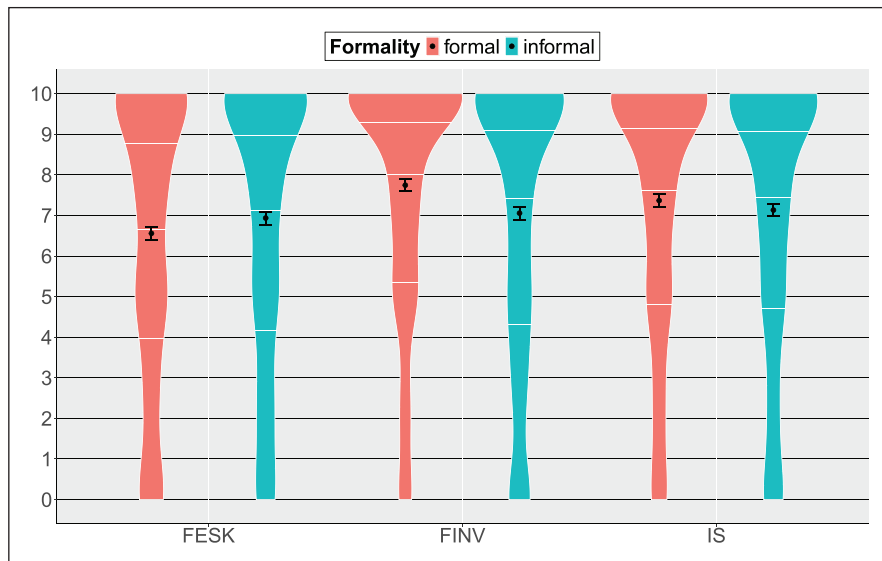


Figure 17: Mean acceptability ratings from Experiment 3, by context (suitable French).

As in Experiment 2, for the AGE variable and on the “well-formedness” scale, the general pattern does not change between age groups (Figure 18), but participants older than 30 make a greater distinction between fronting-with-inversion sentences and both other types, compared to younger participants. On the “suitability scale” (Figure 19), participants younger than 30 exhibit a slight dispreference for FESK sentences (rated lowest) compared to the two other types, but this difference is larger in participants older than 30.

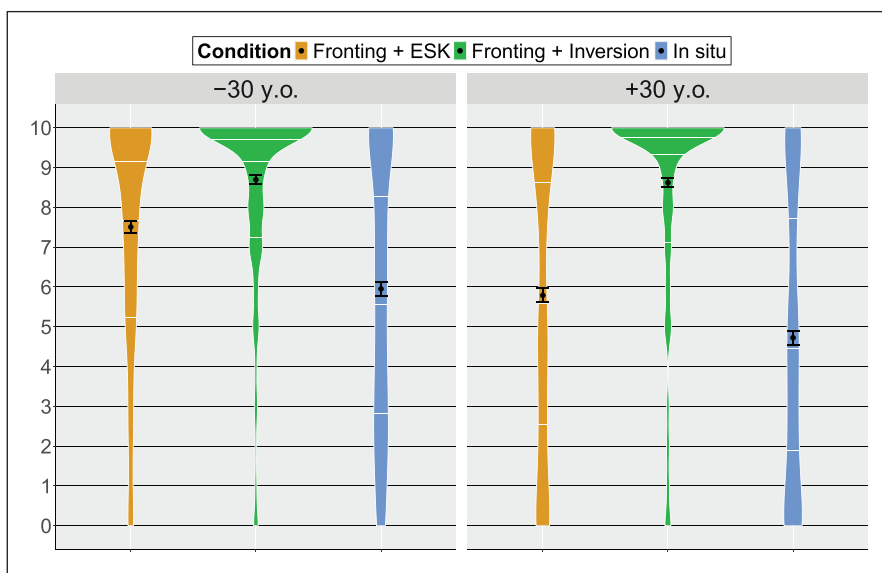


Figure 18: Mean acceptability ratings from Experiment 3, by age group (good French).

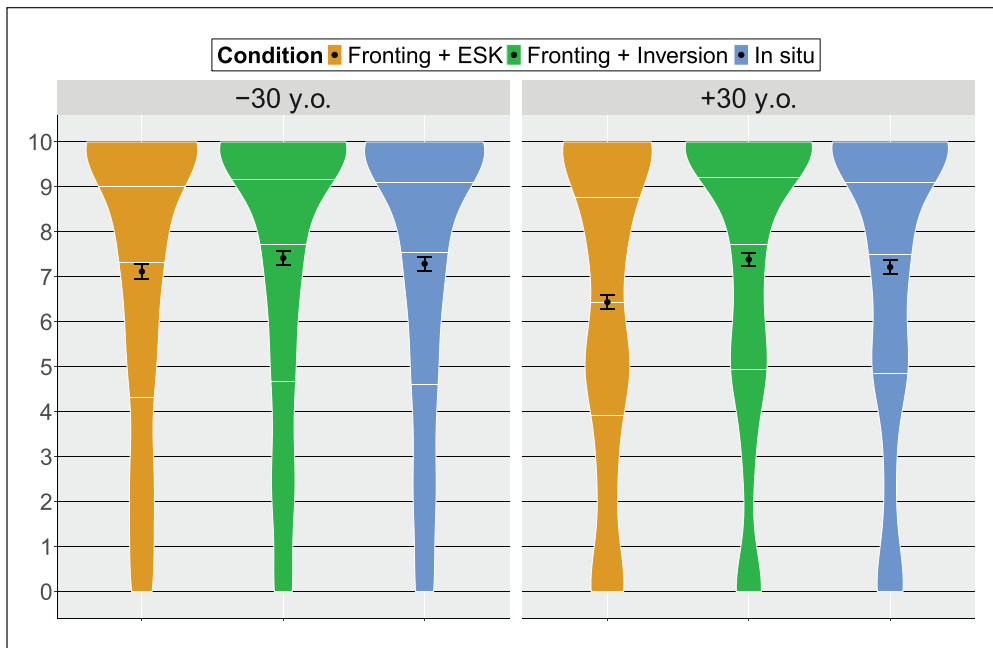


Figure 19: Mean acceptability ratings from Experiment 3, by age group (suitable French).

In addition, there is evidence for a TYPE*FORMALITY*AGE interaction for in-situ vs. fronting-with-inversion sentences ($\hat{\beta} = 0.42$, $P(\beta > 0) = 0.82$, $95\%CrI = [-0.48, 1.32]$), on the “well-formedness” scale. This means that participants older than 30 appear to have rated in-situ sentences a bit higher in formal contexts, when compared to fronting-with-inversion sentences, than participants younger than 30. Although this is surprising, because there is no evidence for such 3-way interactions on the “suitability” scale, this is, nonetheless, tentative evidence for different age groups evaluating different variants differently in different contexts.

When aggregating all these results from Experiment 3, it seems FESK interrogatives are in a somewhat intermediate position between fronting-with-inversion and in-situ sentences in terms of well-formedness, but, overall, they are not as suitable for formal contexts in particular. We also observe a different assessment of this variant in different contexts across age groups, which is in line with both previous experiments and corpus work. FESK seems to be a variant that works in any context without being specialized for any, thus making it fairly acceptable in general, but also generally less suitable than the other two variants in their preferred context.

5 General discussion

The data we presented above calls for a gradient approach to the very notion of acceptability but also highlights the fact that providing participants with a one-dimensional good-for-all acceptability scale may not be appropriate. With Experiment 1, we show that, given only one

general scale, participants conflate acceptability with the normative view of what the standard for their language is or should be. In this sense, acceptability seems strongly related to conformity to the prescriptive linguistic norm, with one form deemed better than the others, or “best”. With Experiments 2 and 3, we show a more nuanced picture of the alternation phenomenon. On the first scale we provided participants with, the fronting-with-inversion variant was the “best-formed” variant (as described in reference grammars), above FESK sentences, which, in turn, were judged as better-formed than in-situ sentences and fronting sentences. This follows the recommendations from prescriptive linguistic norms. For “suitability”, however, all variants seem to have a more unique and specific purpose, with fronting-with-inversion sentences yielding lower ratings in informal contexts and FESK sentences yielding lower ratings in formal contexts. Even if fronting-with-inversion sentences are still rated high in all contexts, this is an overall picture more in line with the frequencies observed in everyday ordinary interactions, such as are found in the the ESLO corpus (Thiberge et al., 2021).

In this article, we focused on one specific alternation phenomenon of a specific linguistic system (*wh*-interrogatives in adult French L1 speakers). We found that preferences for the different available forms varied across speaker groups (age) and also across contexts (non-formal vs. formal). We also found that these preferences are far from optimally captured with a unique and general acceptability scale, which rather seems to reflect a general normative view of what a “proper sentence” should look like in a given language, on a par with what can be found in reference grammars (e.g., for French, Riegel et al., 2014) and in textbooks directed at French children practicing their L1 at school.

From a broader perspective, the difference in modality between Experiment 2 (audio) and Experiment 3 (written), as well as the wider range used for the scales (6-point for Experiment 2, 11-point for Experiment 3), do not seem to have influenced the perception of what is “proper French” and what is “French one can use in a particular context”. Both sets of participants from both experiments were able to distinguish between the two in a similar fashion. We take this as further evidence that French speakers are aware of the discrepancy between “normative French”, as described in reference grammars and taught at school, and “everyday French”, as used in everyday interactions, and they are able to express it when given the proper tools to do so (i.e., different scales, and not a unique good-for-all general acceptability scale). A question arises as to whether participants would give similar responses if they were presented with only one of the two scales. We intend to follow up on this in further research. This, of course, is particularly salient in a language, such as French, where the written modality is often idealized and even defended by an institution such as the *Académie Française*, whose very job consists in setting what can and cannot be considered as “proper French”.¹¹

¹¹ See Abeillé et al. (2023) for an overview of the current debates on French and linguistic norms.

This subtle distinction between normative *acceptability* (as in “acceptability closely defined by the prescriptive linguistic norm”) and a more general *suitability* for a given situation (as in “adaptation to a given social interaction and not detected as incongruent with it by the sociolinguistic monitor”) is relevant from a sociolinguistic perspective and particularly in a third-wave approach to linguistic alternation phenomena (Eckert, 2012). It draws a better picture of how participants react to different linguistic forms and how well they connect with the linguistic norm or so-called “standard”, compatible, for instance, with the perspective of the Social Meaning Games (Burnett, 2017, 2019), where language users dynamically modify their linguistic behavior across and even within interactions, according to which social *persona* (Ochs, 1992) they want to convey.

In line with many other works trying to pinpoint the nature of acceptability and how acceptability judgments help characterize it (Huang and Ferreira, 2020; Hubers et al., 2020; Tubau et al., 2020; i.a.), we argue that this distinction is relevant for psycholinguistic studies in general. Of course, the difference between alternative variants may be related to syntactic complexity, and different facilitatory processes may have historically filtered some variants out of language use (be it because of syntactic complexity by itself, prosodic balancing, or information load and pragmatic uses, for instance). On top of these processes, however, another layer should be accounted for: the sociolinguistic demand for a form, or variant, that is suitable for the particular context it is produced in. Consciously or not, participants in language studies use the nuanced distinctions between what is said, what can be said, what should be said, and what sounds best in terms of well-formedness and with regard to the sociolinguistic norms that apply in a given situation. This awareness is part of the mechanisms that allow speakers to dynamically chose which variant will best allow them to convey the social *persona* they are trying to build during interaction, depending on the context and the social groups they or their interlocutor(s) belong to.¹²

The sociopragmatic ability to evaluate the suitability of a given linguistic production for the particular context/interaction it appears in is still not systematically taken into account in many current-day experimental and quantitative works in the syntactic and, more generally, in the linguistic domain. Language competence is, beyond the knowledge of the components of the language system, a context-dependent interaction device for transmitting and acquiring information. In this view, acceptability is more than a reflection of grammatical experience or an inverse correlate to surprisal (Culicover et al., 2022). Acceptability is a socially situated judgment that should be assessed accordingly. Not only are there ceiling and floor effects (Bader

¹² This *persona*, or “social mask”, is built in social interaction by using linguistic variants according to different sets of social stereotypes that are attached to them (see Beltrama, 2020; Podesva, 2011). The relationship between linguistic forms and social *persona(e)* can be assessed in many different ways.

& Häussler, 2010) where frequency and well-formedness do not follow from one another, but a relatively rare linguistic form, such as the fronting with inversion interrogative variant in French, can be deemed highly “acceptable” in a sociolinguistically compatible context, and a much more frequent variant, such as the in-situ question type, might not be “acceptable” in another (e.g., a formal dialogue in which respecting the prescriptive norm is essential, such as a job interview).

Sociosyntactic approaches such as Cornips and Poletto (2005, p. 952) or Buchstaller and Corrigan (2011, p. 33), and even Schütze (1996, p. 151), have already underscored the importance of contextualizing stimuli in order to assess acceptability. But this recommended contextualization mostly aims at ensuring a pragmatically consistent interpretation of the target sentence by participants, to minimize possible sources of variability in their answers. In support of these preexisting considerations for context, we also argue that experiments where stimuli are presented in isolation are not sufficient to clearly establish the current status of a linguistic phenomenon. We suggest that this contextualization needs to go further and that, when assessing the preferences of speakers, the context of interaction needs to be controlled and accounted for even more carefully. Drawing from the experiments we presented, this can also be done by manipulating context formality. This, in a way, also relates to Labovian worries about the influence of a potential observer on people’s linguistic behavior (*Observer’s Paradox*; Labov, 1972), but, again, it goes deeper. What we manipulated here was not the experimenter’s potential influence on acceptability assessments by the participants, who mostly never even interacted with the experimenter, because the studies were run online. What matters, and what can also change participants’ assessments of acceptability, is the social relationship between the persons who are actually involved in the interaction in which different variants could be used (in a dialogue, for instance), whether the interaction reduces the social distance between interlocutors (use of colloquial terms and orality markers) or whether it does not (no proximity markers). In parallel, experiments that only seek to assess the general acceptability of a given linguistic phenomenon may miss an entire dimension of the use of said form(s) in everyday language, and risk introducing biases in subsequent analyses. Looking for the one best scale for acceptability judgments may, thus, be misleading when multiple dimensions have to be accounted for.

Abbreviations

VSO: Verb-Subject-Object

SVO: Subject-Verb-Object

ESLO: Enquêtes sociolinguistiques à Orléans (Eshkol-Taravella et al., 2011; LLL, 2017)

Glosses

2: second person

COMP: complementizer

EXPL: expletive

FUT: future tense

NOM: nominative

PST: present

SG: singular

Interrogative variants

F: Fronting (simple)

FESK: Fronting with *est-ce que*

FINV: Fronting with inversion

IS: in situ

Statistical analyses

CrI: Credibility Interval

Data accessibility statement

All supplementary materials are available at <https://osf.io/gjdc4/> (detailed analyses, R scripts, graphs and materials + norming data for all experiments). DOI: <https://doi.org/10.17605/OSF.IO/GJDC4>.

Ethics and consent

This experimental research was performed in accordance with general ethical guidelines aiming to protect the rights and health of all participants. In particular, all participants were fully informed of their legal rights before taking part in the experiments: no known risk was associated with taking part in the experiments, they could withdraw from the study at any time without prejudice, and they could access, revise or ask for the deletion of the data they provided before publication. All participants gave full consent based on this information. All the published data

is anonymous and no link can be established back to the original participants. All participants could also ask for additional information on how their data were processed or what the purpose of the studies was.

Acknowledgements

We thank the editor at *Glossa Psycholinguistics*, Meredith Tamminga, for her many comments, which made this publication a possible bridge between psycholinguistics and sociolinguistics. We also thank the three anonymous reviewers, whose very insightful comments helped make this article more impactful, and more integrated within the current literature on both acceptability and prescriptivism. Finally, we thank all anonymous participants in our studies.

Funding information

This work is supported/ partially supported by a public grant overseen by the IdEx Université Paris Cité (ANR-18-IDEX-0001) as part of the Labex *Empirical Foundations of Linguistics* – EFL.

Competing interests

The authors have no competing interests to declare.

Authors' contributions

GT conducted the experiments described in this article as well as the statistical analyses, and redacted the first draft of this paper. BH advised on the building of the experimental paradigms as well as the statistical analyses, and proposed revisions of the first draft. Both authors contributed to the revisions and agreed upon the final version of the manuscript.

ORCID IDs

Gabriel Thiberge: <https://orcid.org/0000-0002-9727-7860>

Barbara Hemforth: <https://orcid.org/0000-0001-8371-1323>

References

- Abeillé, A. (2007). *Les grammaires d'unification* [978-2-7462-1251-0]. Hermès. <https://halshs.archives-ouvertes.fr/halshs-00748332>
- Abeillé, A., Auger, J., Benzitoun, C., Burnett, H., Candea, M., Gadet, F., Gasquet-Cyrus, M., Gautier, A., Hoedt, A., Klinkenberg, J.-M., Launey, M., Neveux, J., Panckhurst, R., Piron, J., Filstroff, R., Rossari, C., Siouffi, G., Raemdonck, D. V., & Véron, L. (2023, May). *Le français va très bien, merci* (Vol. Tracts). Gallimard. <https://hal.science/hal-04122818>

- Abeillé, A., & Godard, D. (2021). *La grande grammaire du français*. Éditions Actes Sud.
- Adger, D. (2006). Combinatorial variability. *Journal of Linguistics*, 42(3), 503–530. <https://doi.org/10.1017/S002222670600418X>
- Adger, D. (2007). Variability and modularity: A response to Hudson. *Journal of Linguistics*, 43(3), 695–700. <http://www.jstor.org/stable/40058017>. <https://doi.org/10.1017/S0022226707004811>
- Adli, A. (2015). What you like is not what you do: Acceptability and frequency in syntactic variation. In A. Adli, M. G. García, & G. Kaufmann (Eds.), *Variation in language: Usage-based vs. system-based approaches* (pp. 173–200). Mouton de Gruyter. <https://doi.org/10.1515/9783110346855-008>
- Ashby, W. (1977). Interrogative forms in Parisian French. *Semasia*, 4, 35–52.
- AudacityTeam. (1999–2021). Audacity(r): Free audio editor and recorder [computer application] [Version 2.4.2, retrieved Sep 20th 2020]. <https://audacityteam.org/>
- Bader, M., & Häussler, J. (2010). Toward a model of grammaticality judgments. *Journal of Linguistics*, 46(2), 273–330. <https://doi.org/10.1017/S0022226709990260>
- 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>
- Behnstedt, P. (1973). *Viens-tu? est-ce que tu viens? tu viens?* (Vol. 41). Tübinger Beiträge zur Linguistik.
- Beltrama, A. (2020). Social meaning in semantics and pragmatics. *Language and Linguistics Compass*, 14(9), e12398. <https://doi.org/10.1111/lnc3.12398>
- Beysade, C., & Marandin, J.-M. (2006). The speech act assignment problem revisited: Disentangling speaker's commitment from speaker's call on addressee. *Empirical Issues in Syntax and Semantics*, 6, 37–68. www.cssp.cnrs.fr/eiss6/beysade-marandin-eiss6.pdf
- Blache, P., Hemforth, B., & Rauzy, S. (2006, July). Acceptability prediction by means of grammaticality quantification. In N. Calzolari, C. Cardie, & P. Isabelle (Eds.), *Proceedings of the 21st International Conference on Computational Linguistics and 44th annual meeting of the Association for Computational Linguistics* (pp. 57–64). Association for Computational Linguistics. <https://doi.org/10.3115/1220175.1220183>
- Brehm, L., & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, 125, 104334. <https://doi.org/10.1016/j.jml.2022.104334>
- Buchstaller, I., & Corrigan, K. (2011). How to make intuitions succeed: Testing methods for analysing syntactic microvariation. In W. Maguire & A. McMahon (Eds.), *Analysing variation in English* (pp. 30–48). Cambridge University Press. <https://doi.org/10.1017/CBO9780511976360.003>
- Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1), 395–411. <https://doi.org/10.32614/RJ-2018-017>

- Burnett, H. (2017). Sociolinguistic interaction and identity construction: The view from game-theoretic pragmatics. *Journal of Sociolinguistics*, 21(2), 238–271. <https://doi.org/10.1111/josl.12229>
- Burnett, H. (2019). Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy*, 42(5), 419–450. <https://doi.org/10.1007/s10988-018-9254-y>
- Burnett, H., & Pozniak, C. (2021). Political dimensions of gender inclusive writing in Parisian universities. *Journal of Sociolinguistics*, 25(5), 808–831. <https://doi.org/10.1111/josl.12489>
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1). <https://doi.org/10.18637/jss.v076.i01>
- Chomsky, N. (1965). *Aspects of the theory of syntax*. The MIT Press. <https://doi.org/10.21236/AD0616323>
- Chomsky, N. (1981). *Lectures on government and binding*. Foris.
- Cornips, L., & Poletto, C. (2005). On standardising syntactic elicitation techniques (Part 1). *Lingua*, 115, 939–957. <https://doi.org/10.1016/j.lingua.2003.11.004>
- Coveney, A. (2011). L'interrogation directe. *Travaux de Linguistique*, 2, 112–145. <https://doi.org/10.3917/tl.063.0112>
- Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. Sage.
- Culicover, P. W., Varaschin, G., & Winkler, S. (2022). The radical unacceptability hypothesis: Accounting for unacceptability without universal constraints. *Languages*, 7(2), 96. <https://doi.org/10.3390/languages7020096>
- DeBruine, L. M., & Barr, D. J. (2021). Understanding mixed-effects models through data simulation. *Advances in Methods and Practices in Psychological Science*, 4(1). <https://doi.org/10.1177/2515245920965119>
- Dillon, B., & Wagers, M. (2019). Approaching gradience in acceptability with the tools of Signal Detection Theory. In G. Goodall (Ed.), *The Cambridge handbook of experimental syntax* (pp. 62–96). Cambridge University Press. <https://doi.org/10.1017/9781108569620.004>
- Donati, C. (2006). On wh-head movement. *Current Studies in Linguistics Series*, 42, 21–46. https://www.researchgate.net/profile/Caterina_Donati/publication/261932598_On_Wh_Head_Movement/links/02e7e535f810234c01000000.pdf. <https://doi.org/10.7551/mitpress/7197.003.0006>
- Drummond, A. (2016). *Ibexfarm* (version 0.3. 9).
- Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, 41(1), 87–100. <https://doi.org/10.1146/annurev-anthro-092611-145828>
- Eshkol-Taravella, I., Baude, O., Maurel, D., Hriba, L., Dugua, C., & Tellier, I. (2011). Un grand corpus oral “disponible”: le corpus d’Orléans 1 1968–2012. *Traitement Automatique des Langues*, 53(2), 17–46. <https://halshs.archives-ouvertes.fr/halshs-01163053>

- Falk, Y. (2012). *Superiority effects*, The Hebrew University of Jerusalem. <http://pluto.huji.ac.il/~msyfalk/Superiority.pdf>
- Ferreira, F. (2005). Psycholinguistics, formal grammars, and cognitive science. *The Linguistic Review*, 22, 365–380. <https://doi.org/10.1515/tlir.2005.22.2-4.365>
- Gibson, E., & Fedorenko, E. (2010). Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences*, 14(6), 233–234. <https://doi.org/10.1016/j.tics.2010.03.005>
- Gibson, E., & Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28(1–2), 88–124. <https://doi.org/10.1080/01690965.2010.515080>
- Ginzburg, J., & Sag, I. (2000). *Interrogative investigations*. CSLI publications.
- Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.
- Granlund, S., Kolak, J., Vihman, V., Engelmann, F., Lieven, E. V., Pine, J. M., Theakston, A. L., & Ambridge, B. (2019). Language-general and language-specific phenomena in the acquisition of inflectional noun morphology: A cross-linguistic elicited-production study of Polish, Finnish and Estonian. *Journal of Memory and Language*, 107, 169–194. <https://doi.org/10.1016/j.jml.2019.04.004>
- Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language*, 10(1). <https://www.jstor.org/stable/25000703>
- Hamlouï, F. (2009). *La focalisation à l'interface de la syntaxe et de la phonologie: Le cas du français dans une perspective typologique* [Doctoral dissertation]. Université Paris 3.
- Huang, Y., & Ferreira, F. (2020). The application of signal detection theory to acceptability judgments. *Frontiers in Psychology*, 11, 498268. <https://doi.org/10.3389/fpsyg.2020.00073>
- Hubers, F., Redl, T., De Vos, H., Reinartz, L., & De Hoop, H. (2020). Processing prescriptively incorrect comparative particles: Evidence from sentence-matching and eye-tracking. *Frontiers in Psychology*, 11, 498796. <https://doi.org/10.3389/fpsyg.2020.00186>
- Jakubowicz, C. (2011). Measuring derivational complexity: New evidence from typically developing and SLI learners of L1 French [Specific Language Impairment]. *Lingua*, 121(3), 339–351. <https://doi.org/10.1016/j.lingua.2010.10.006>
- Jamieson, E., Smith, J., Adger, D., Heycock, C., & Thoms, G. (2024). ‘When intuitions (don’t) fail’: Combining syntax and sociolinguistics in the analysis of Scots. *English Language and Linguistics*, 29(1), 1–34. <https://doi.org/10.1017/S1360674323000679>
- Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, 1(1), 3–44. <https://doi.org/10.1007/BF00351935>
- Kayne, R. S. (1981). ECP extensions. *Linguistic Inquiry*, 12(1), 93–133. <https://www.jstor.org/stable/4178206?seq=1>
- Keller, F. (2000). *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality* [Doctoral dissertation]. University of Edinburgh.

- Labov, W. (1972). Some principles of linguistic methodology. *Language in Society*, 1(1), 97–120. <https://doi.org/10.1017/S0047404500006576>
- Labov, W. (1996). When intuitions fail. *Papers from the 32nd Regional Meeting of the Chicago Linguistics Society*, 76–106.
- Labov, W., Ash, S., Ravindranath, M., Weldon, T., Baranowski, M., & Nagy, N. (2011). Properties of the sociolinguistic monitor. *Journal of Sociolinguistics*, 15(4), 431–463. <https://doi.org/10.1111/j.1467-9841.2011.00504.x>
- Langsford, S., Perfors, A., Hendrickson, A. T., Kennedy, L. A., & Navarro, D. J. (2018). Quantifying sentence acceptability measures: Reliability, bias, and variability. *Glossa*, 3(1), 1–34. <https://doi.org/10.5334/gjgl.396>
- Lau, J. H., Clark, A., & Lappin, S. (2017). Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Science*, 41(5), 1202–1241. <https://doi.org/10.1111/cogs.12414>
- LLL. (2017). Eslo [ORTOLANG (Open Resources and TOols for LANGuage) –www.ortolang.fr]. <https://hdl.handle.net/11403/eslo/v1>
- Marandin, J.-M. (2018). *Les types de phrases en français contemporain*. <https://hal.archives-ouvertes.fr/hal-01882110>
- Marty, P., Chemla, E., & Sprouse, J. (2020). The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa*, 5(1), 72. <https://doi.org/10.5334/gjgl.980>
- Ochs, E. (1992). Indexing gender. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 335–358). Cambridge University Press.
- Podesva, R. J. (2011). Salience and the social meaning of declarative contours: Three case studies of gay professionals. *Journal of English Linguistics*, 39(3), 233–264. <https://doi.org/10.1177/0075424211405161>
- Pohl, J., & Straka, G. (1965). Observations sur les formes d'interrogation dans la langue parlée et dans la langue écrite non littéraire. *Actes du Xe Congrès International de Linguistique et Philologie Romanes*, 501–513. <https://api.semanticscholar.org/CorpusID:164230515>
- Posit team. (2023). *Rstudio: Integrated development environment for r*. Posit Software, PBC. <http://www.posit.co/>
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reinhardt, J. (2019a). Les interrogatives directes tirées de dix romans policier [ORTOLANG (Open Resources and TOols for LANGuage) –www.ortolang.fr]. <https://hdl.handle.net/11403/interrogatives-in-novels/v1>
- Reinhardt, J. (2019b). *Regularity and variation in French direct interrogatives : The morphosyntax and intonation of question forms in reality tv shows, audio books and teaching materials* [Doctoral dissertation]. Universität Konstanz. <http://nbn-resolving.de/urn:nbn:de:bsz:352-2-y00meiao8m342>

- Riegel, M., Rellat, J.-C., & Rioul, R. (2014). *Grammaire méthodique du français*. Presses Universitaires de France.
- Robbins, P. T. (2019). *The effect of prescriptive rules and instructions on the grammaticality judgment task*. https://cos.northeastern.edu/wp-content/uploads/2016/07/Robbins_Final.pdf
- Ross, J. R. (1967). *Constraints on variables in syntax* [Doctoral dissertation]. Massachusetts Institute of Technology.
- Sag, I. A. (2010). English filler-gap constructions. *Language*, 86(3), 486–545. <https://www.jstor.org/stable/40961690?seq=1>. <https://doi.org/10.1353/lan.2010.0002>
- Schütze, C. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. University of Chicago Press.
- Söll, L. (1983). L'interrogation directe dans un corpus de langage enfantin. In F.-J. Haussman (Ed.), *Études de grammaire française descriptive* (pp. 45–54). Groos.
- Sorace, A., & Keller, F. (2005). Gradience in linguistic data. *Lingua*, 115(11), 1497–1524. <https://doi.org/10.1016/j.lingua.2004.07.002>
- Sorensen, T., Hohenstein, S., & Vasishth, S. (2016). Bayesian linear mixed models using stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology*, 12(3), 175–200. <https://doi.org/10.20982/tqmp.12.3.p175>
- Sprouse, J., & Almeida, D. (2017). Design sensitivity and statistical power in acceptability judgment experiments. *Glossa*, 2(1). <https://doi.org/10.5334/gjgl.236>
- Sprouse, J., Schütze, C. T., & Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from *Linguistic Inquiry* 2001–2010. *Lingua*, 134, 219–248. <https://doi.org/10.1016/j.lingua.2013.07.002>
- Steedman, M., & Baldridge, J. (2011). Combinatory categorial grammar. In R. D. Borsley & K. Börjars (Eds.), *Non-transformational syntax* (pp. 181–224). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781444395037.ch5>
- Stockwell, R. P. (1977). *Foundations of syntactic theory*. Prentice Hall.
- Terry, R. M. (1970). *Contemporary French interrogative structures* (Vol. 5). Editions Cosmos.
- Thiberge, G. (2018). Position du syntagme WH- en français : Réelle optionnalité ou biais sociolinguistique? *ELIS – Échanges Linguistiques en Sorbonne, Université Paris Sorbonne*, 5, 64–91. <https://halshs.archives-ouvertes.fr/halshs-01802590>
- Thiberge, G., Badin, F., & Liégeois, L. (2021). French partial interrogatives: A microdiachronic corpus study of variation and new perspectives in a refined pragmatics framework. *Faits de Langue*, 51(2), 179–202. <https://doi.org/10.1163/19589514-05102010>
- Tubau, S., Etxeberria, U., Déprez, V., & Espinal, M. T. (2020). What are (un)acceptability and (un)grammaticality? How do they relate to one another and to interpretation? *Frontiers in Psychology*, 11, 621267. <https://doi.org/10.3389/978-2-88966-374-3>
- Van Trijp, R. (2014). Long-distance dependencies without filler-gaps: A cognitive-functional alternative in fluid construction grammar. *Language and Cognition*, 6(2), 242–270. <https://doi.org/10.1017/langcog.2014.8>

Wagner, S. E. (2012). Age grading in sociolinguistic theory. *Language and Linguistics Compass*, 6(6), 371–382. <https://doi.org/10.1002/lnc3.343>

Zribi-Hertz, A. (2006). Français standard et francilien commun: Conséquences du phénomène diglossique pour la description et l'enseignement du Français. *Actes du colloque de linguistique de l'Université d'Okayama*.

