

Good French isn't always best.

Acceptability between normativity and usage.

Word count: 8428

Author 1
Affiliation 1
xxx@xxx.xxx

Author 2
Affiliation 2
xxx@xxx.xxx

Abstract Acceptability judgments are one of the major tools for (psycho-)linguists to assess speakers' preferences for specific utterances in any given language, shedding light on their internal grammar. It is however well known that factors unrelated to grammaticality such as frequency of exposure, cognitive constraints and others can influence the perceived acceptability of an utterance. Through the study of the partial interrogative system of French, a language that exhibits great variation in the forms that are available to ask a question, but also where the linguistic norms weigh heavily on what is considered "good" French, we add to the evidence that acceptability is a notion that needs to be taken with caution. In three experiments, we show that L1 adult speakers of French have internalized the dichotomy between "good French" and "real-life" French, which they are able to express when given the proper 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).

Keywords: French; partial interrogatives; acceptability judgments; experimental sociolinguistics; quantitative studies; Bayesian modeling

1 Introduction

Quantitative research in linguistics has developed considerably in the past 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 (Gibson & Fedorenko, 2010, 2013). Multiple observations of the influence of minimally differing linguistic contrasts on various depending variables (e.g. reaction times or ratings), repeated across many different participants and many different items in a counterbalanced way, allow for robust inferences on what goes on at a general population level (i.e. all the speakers of a given linguistic community). Of particular interest to the (psycho-)linguistic community from the very beginning of the cognitivist approaches and over the next decades (Chomsky, 1965; Schütze, 1996), the Acceptability Judgment Task has become one of the most used experimental paradigms, where ratings of naive participants are collected about a variety of linguistic forms to assess which ones they accept as parts of their language. The overall paradigm is easy to set up and easy to 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. A number of

issues have been addressed over the years, for example the issue of how informal designs without a clear systematic or inferential approach to the results (*informal* methods) compare to more constrained set-ups inspired by psycholinguistic methods (*formal* methods) (Sprouse, Schütze, & Almeida, 2013), the issue of how different designs vary in terms of sensitivity (Marty, Chemla, & Sprouse, 2020; Sprouse & Almeida, 2017), or the issue of how reliable and variable the results of such a task may be depending on – for example – the scales that are used (Langsford et al., 2018).

Perhaps a broader and deeper issue concerns what exactly the task is measuring, and the very definition of *acceptability*. Whole journal issues focus for instance 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, who may each have their own biases with regard to how well-formed a sentence is in a given experimental settings (Huang & Ferreira, 2020). From what the authors consider “noisy data” (noise being here 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 and 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 being judged acceptable while in violation of prescriptive grammar rules. This in a sense is reminiscent of older proposals (Blache, Hemforth, & Rauzy, 2006; Keller, 2000; Sorace & Keller, 2005) establishing acceptability as a more continuous notion, where – in the authors’ theoretical framework – this results from some syntactic constraints being *softer* than others. Finally, this nuanced perspective on acceptability also resonates with some experimental evidence for an 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 either 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 a differentiation of normative or prescriptive grammars or 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 this paper we explore this issue by looking at the system of partial interrogatives in French. 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 partial question in French, we provide evidence for different nuances of acceptability of these sentences. Based on available corpus data exploring the same phenomenon (Thiberge, Badin, & Liégeois, 2021), the design of our experiments provides evidence for an influence of contextual and sociolinguistic factors on the assessments. 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 actually 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*, which might be conflated with the more traditional notion of (prescriptive) *grammaticality*, and what we call general

suitability, but this difference is nuanced by sociolinguistic factors potentially external to the linguistic form in 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 partial interrogative system and of the data already available with regard to this alternation phenomenon, and we present our analysis as to 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. In section 5, we summarize our findings and we argue for the integration of sociolinguistic concepts to better characterize the very notion of acceptability.

2 French partial interrogation as a case study for acceptability

In this paper, we will focus on French partial interrogatives as a case study of how acceptability might or might not be conflated with grammatical norms, and of how sociolinguistic factors might influence speakers' assessments of acceptability. Partial 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, Badin, & Liégeois, 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 make predictions as to how judgments of different partial interrogatives depending on different manipulated conditions could help better understanding different aspects of *acceptability*. As will be shown 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. We will take this as evidence that speakers of a language have a more or less developed socio-pragmatic competence to choose a variant that is optimal given a particular social situation. Following the norm may not always be the best way for a speaker to fit in. We will argue that this is an underlying dimension of acceptability that psycholinguistic approaches should take into account.

2.1 A wide range of alternatives for asking partial questions in French

Following Beyssade and Marandin (2006), Ginzburg and Sag (2000), 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 (*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 (see Hamblin (1973), Karttunen (1977) for foundational work, and all that has been built upon it), we will focus on the coexistence of multiple syntactic variants to ask such partial interrogatives in French.

French partial 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, 2019; Söll, 1983; Terry, 1970; Thiberge, Badin, & Liégeois, 2021). The examples in (1) are some of the main forms available to French speakers when asking a partial question. Partial 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 (more than 50% of productions around the year 2000), and *fronting with inversion* being used less and less (less than 10% of productions) (Adli, 2015; Hamlaoui, 2009; Thiberge, Badin, & Liégeois, 2021). This, however, contradicts what is still found in reference grammars for French such as Riegel, Rellat, and Rioul (2014), 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. All in all, more than 15 pages are dedicated to interrogative strategies of French in this book, but so-called “non-standard” variants (i.e. all variants but *fronting with inversion*) are described on a mere 10 lines of text.

The reasoning with respect to the use of in particular *in situ* variants that is put forward in Riegel, Rellat, and Rioul (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 and the minimalist framework 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 (CP, complementizer phrase), which is assumed to lead to increased complexity of the fronting + inversion question type compared to the others. Other frameworks such as HPSG (Sag, 2010) or unification grammars (Abeillé, 2007) have described the phenomenon without movement explana-

¹ Riegel, Rellat, and Rioul (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.”

tions, and in this view frequency alone could for example explain the use of the *in situ* question type.

2.2 Using acceptability judgments to target the discrepancy between reference grammars and real-world usage of the variants

To shed light on the usage of the different interrogative variants, Thiberge, Badin, and Liégeois (2021) analyzed 617 partial interrogatives of French, 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 less *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 everyday people 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, but also 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 the everyday lives of all members of the family. 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 less *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.

All in all, recent corpus data show that only a small proportion of the interrogatives found in modern corpora are *fronting-with-inversion*-sentences while *in-situ*-sentences account for around 50% of the productions, or even more depending on the context. This contrasts sharply with what could be predicted from a reference grammar for French (e.g. Riegel, Rellat, and Rioul (2014)), often used by teachers for example in French as a Foreign Language settings to prepare their materials. Moreover, this discrepancy between corpus data and the prescriptive grammar makes French partial interrogatives a particularly interesting test case. Frequency and acceptability typically correlate 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, Clark, & Lappin, 2017).

The experiments we report below shed light on the complex approach to acceptability by L1 French speakers. 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. By means of the acceptability judgment paradigm, we target the issue of how speakers of a given language perceive and use linguistic variants in a normative setting where, prescriptively, one variant is considered the “standard” variant.

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

With three acceptability judgment tasks, we assessed French native speakers' preferences with regard to the four main variants: *in situ*, *fronting*, *fronting with inversion*, *fronting with est-ce que*. This approach aims to resolve a dearth in experimental data regarding the use of interrogative variants in contemporary French, and builds upon recent fine-grained corpus work on the usage of French partial interrogatives in different contexts of interaction (Thiberge, Badin, & Liégeois, 2021).³

Based on available data, we could anticipate a one-dimensional “acceptability” to translate to higher ratings for the normative *standard* variant, that is to say *fronting-with-inversion*-type interrogatives, if the relationship between prescriptive *grammaticality* and *acceptability* is deeply entrenched in the speakers' mind. If on the contrary acceptability assessments are associated with real-usage and frequency effects (see for instance the (extended) radical unacceptability hypothesis put forth by Culicover, Varaschin, and Winkler (2022)), we would expect the most common form in contemporary French, i.e. the *in-situ*-type interrogatives, to yield higher results. This is what we target 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, which we extended to the *fronting with est-ce que* variant. For this, we used two different scales, one targeting the well-formedness of the variants to assess, and the other targeting the suitability of the variants to a (short) context we provided. We predicted that participants would more easily express a difference between “acceptability” – which by itself could be conflated with prescriptive “grammaticality” or “well-formedness” – and “suitability” – which would correspond more to real-world usage and frequency aspects of the different forms available. As such, *fronting-with-inversion*-sentences should yield higher ratings on the first scale, while *in-situ*-sentences should yield higher ratings on the second scale.

3 First Experiment: written Acceptability Judgment Task

The first experiment we ran was a traditional acceptability judgment task, where we assessed the general preferences of French L1 adult speakers with regard to the three main interrogative variants found in corpus data (Adli, 2015; Hamlaoui, 2009; Thiberge, Badin, & Liégeois, 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* to *informal* contexts. The manipulation of both these factors was meant to reflect the variability found for example in Thiberge, Badin, and Liégeois (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 both other types). As 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 could parallel those from the corpus data, where different age groups exhibited different linguistic preferences in contexts that differ in terms of formality.

³ And also preliminary studies where small experiments gave some preliminary insights on the preferences of L1 French adult speakers on some of the available variants in this language. See Thiberge (2018) for more details.

3.1 A note on statistical analyses

As an important note, we mention here that for the statistical analyses of all experiments we ran Bayesian regressions. An in-depth explanation can be found in the supplementary material (see [here](#)), but the main motivation for this is two-fold: first, 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) and, second, we wanted to get a fine-grained view of the data, which the binary outcome of frequentist models does not allow (see Sorensen, Hohenstein, and Vasishth (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 Supplementary Materials for full specification and outputs), to 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 effect 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 for a more direct interpretation of the model estimates (Brehm & Alday, 2020). In the present study, we used the *brms* package (Bürkner, 2018; Bürkner et al., 2017; 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 upon 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). Inspired by a.o. Burnett and Pozniak (2021), 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 IV on the DV (“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.⁴ 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 3x2 design, we constructed 30⁵ 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,⁶ then person B asks a partial question relevant in this context.

⁴ Crucially, they will not be considered as not supporting the absence of an effect either.

⁵ See Supplementary Material for a full list.

⁶ Context formality was modulated by three elements, two of which were combined in each A-sentence: dislocation of 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*). See the Supplementary Material for the results of a norming experiment of this formality parameter.

Table 1: Example item from Experiment 1

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

The 30 target dialogues were mixed with 30 filler dialogues with the same A-B structure, containing no partial interrogative and no variation in formality. After 3 practice items, the 60 items were presented in a Latin square design with full randomization. The experiment took place on the Internet via the IbexFarm platform (Drummond, 2016). After a short introduction of the participants' rights, a description of the procedure and a short questionnaire about their linguistic profile, explicit consent was required to move along. Then, participants were presented with the 63 (practice + target + filler) items. Below each sentence, participants rate the acceptability of the B-sentence on a 0–10 numeric scale. *Acceptability* was here broadly 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, 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).

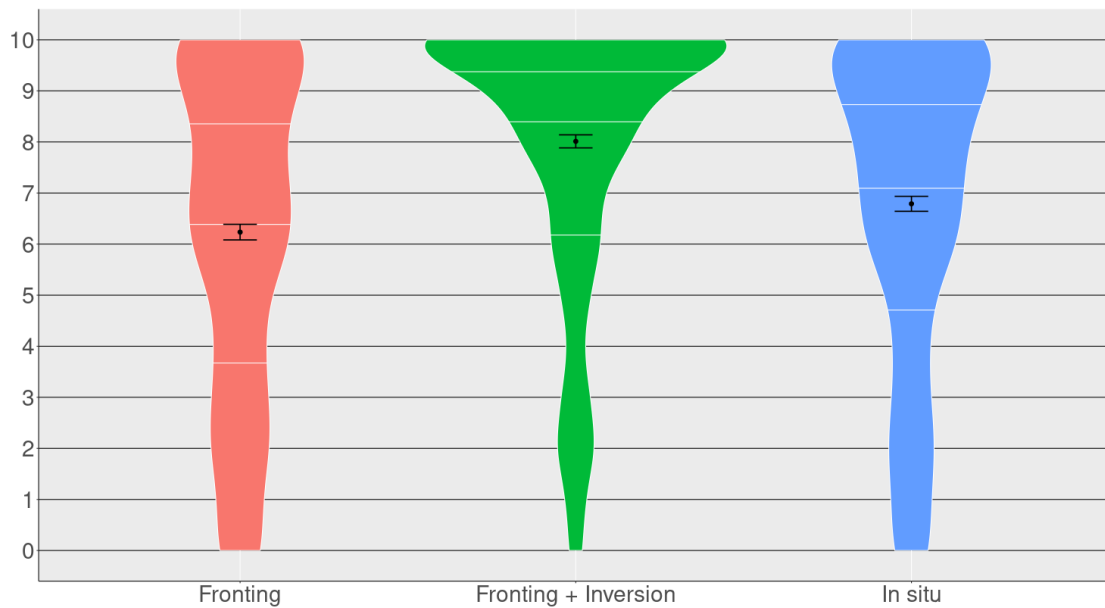


Figure 1: Mean acceptability and distributions of ratings from Experiment 1.

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 Material) 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.

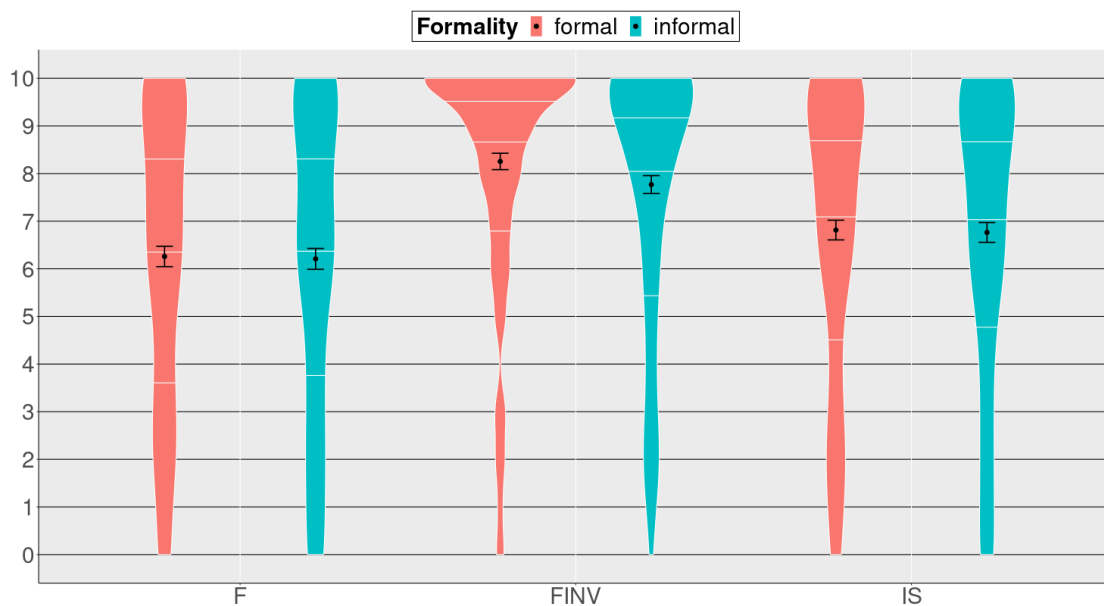


Figure 2: Mean acceptability and distributions of ratings from Experiment 1, by context.

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. We took the split into a 30+ and a <30 group from corpus studies Thiberge, Badin, and Liégeois (2021). 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 bigger for ratings associated with *in-situ* vs. *fronting-with-inversion* sentences, for which the model yields robust evidence (TYPE*AGE interaction at ($\hat{\beta} = -1.67$, $P(\beta < 0) = 0.99$, $95\%CrI = [-3.09, -0.26]$)).

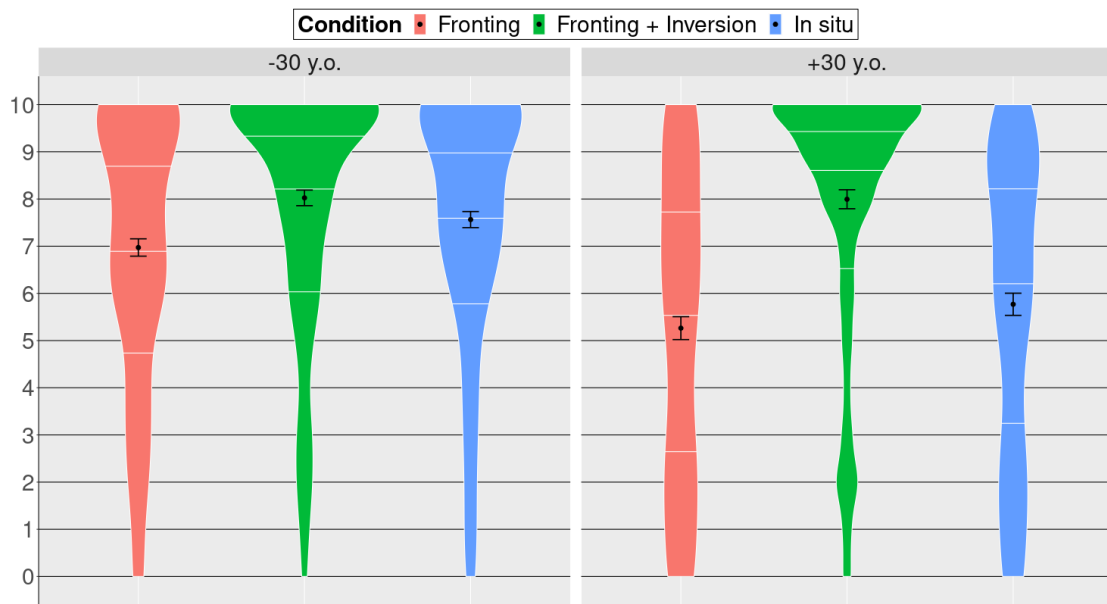


Figure 3: Mean acceptability and distributions of ratings from Experiment 1, by age group.

Figure 4 shows the posterior distributions for the maximal mb1 model fit to the data from Experiment 1. Apart from 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* contexts than in informal ones. From the TYPE*FORMALITY interactions described before, this follows from the higher ratings associated with *fronting-with-inversion*-type sentences in formal contexts.

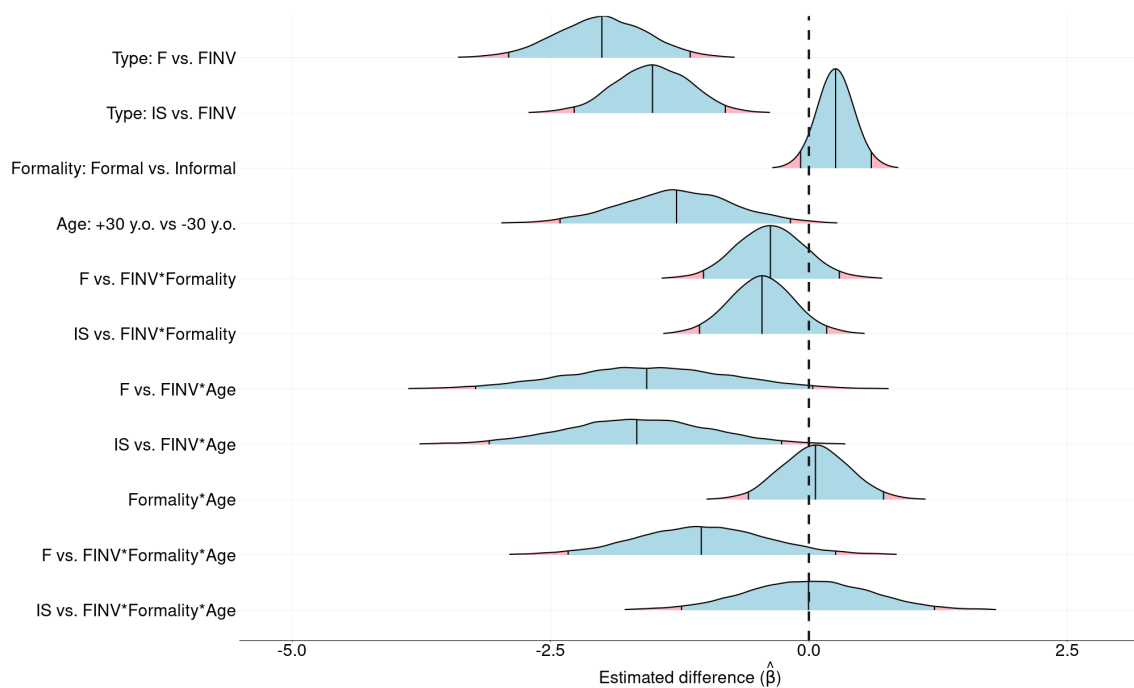


Figure 4: Posterior distributions for the maximal model run for Experiment 1.

Perhaps of more 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 partial 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 the subject but mainly it is in par with many prescriptive or reference grammars of French (e.g. Riegel, Rellat, and Rioul (2014)), where the *fronting + 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, Badin, and Liégeois (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 in 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 y.o. 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 y.o. was also observed for the partial interrogative variants in the ESLO corpus.

Even more 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 y.o. 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 integration. Figure 5 illustrates how this could work: when considering age as a continuous variable, the *in situ* variant seems more acceptable in formal contexts for participants younger than 30 y.o. and older than 60–70. There is no such phenomenon for the *simple fronting* variant, which becomes less and less acceptable when age increases, 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 Wagner (2012). Speakers aged between 30 and 60/70 y.o. are confronted to the necessity of abiding by the linguistic prescriptive norm in their 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.

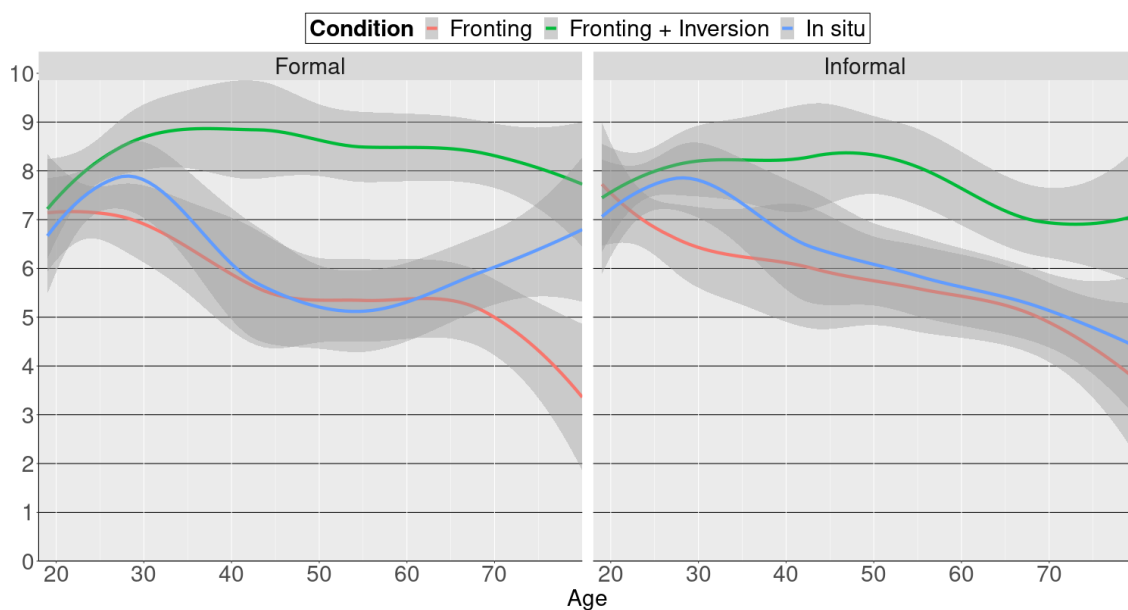


Figure 5: Mean acceptability ratings from Experiment 1, by age (continuous).

4 Nuancing between Good French and Suitable French - Experiments 2 & 3

Results from Experiment 1 are consistent with the hypothesis that preferences for different variants vary depending on the context, and it provided evidence for a general confusion between *acceptability* and prescriptive *grammaticality* (as in *compatibility with the normative standard for a language*). The preferences we found 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 understand 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 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 in 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 0–10 11-point scale, while this time the scales were both 1–6 6-point 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 less 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 affect our results systematically. Answers were not mandatory so participants who felt unable to decide on a scale could skip it for a particular item and still continue with the experiment.

Voices were provided by 4 French adult L1 speakers (F1, F2, M1, M2), who we recorded in a soundproof room. All sentences in all conditions were recorded and then manipulated in Audacity (AudacityTeam, 1999-2021) to create dialogues absolutely parallel with 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 overall for all items and all participants of Experiment 2. We also ran a norming study to assess the “a priori gender” people generally assigned to voices (so as to minimize potential rating differences overall in terms of how women or men are supposed to speak to one another⁷), and to check that no voice was more associated with a colloquial tone than the other, so as to not bias the FORMALITY variable. All the data from this norming study can be found in the Supplementary Material.

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 subsection 3.1 and 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 = $[-3.56, -2.11]$).

⁷ 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.

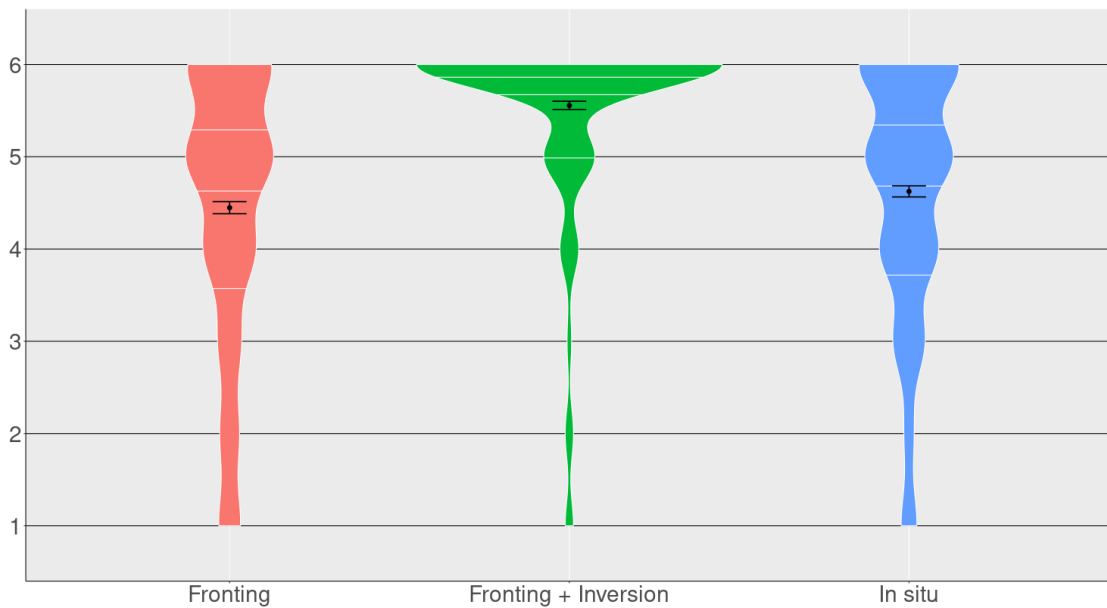


Figure 6: Mean acceptability ratings from Experiment 2 (good French).

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 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 = $[-0.03, 0.95]$).

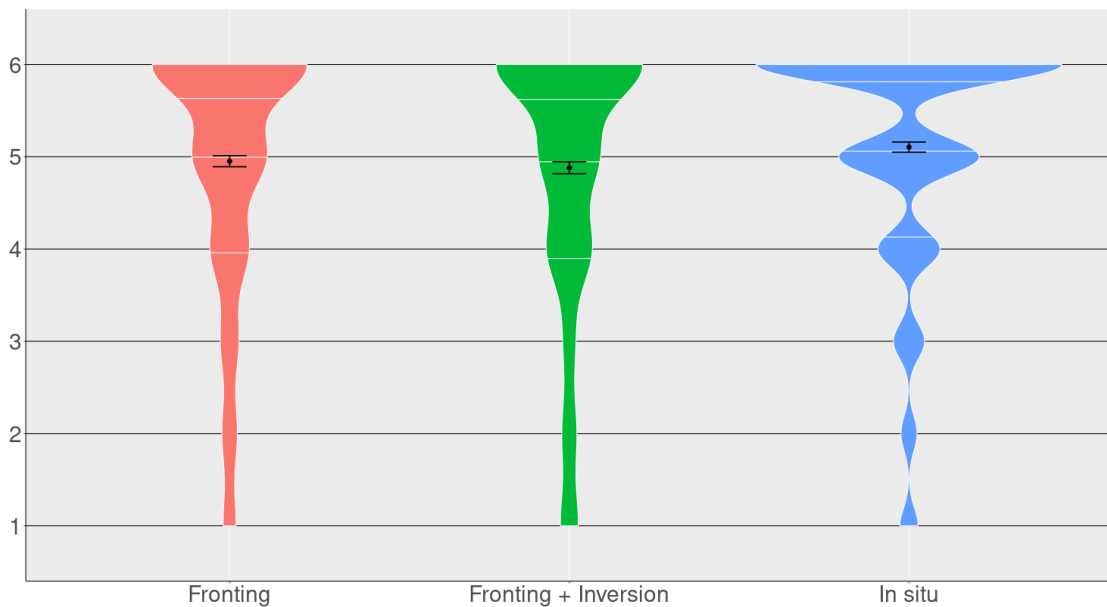


Figure 7: Mean acceptability ratings from Experiment 2 (suitable French).

As for FORMALITY and AGE, the two scales also appear to capture different dimensions of French speakers' preferences. For the "good French" scale, 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

bigger 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, with all three variants yielding ratings around 5/6 on this scale, whatever the context. When looking at the results in more detail, however, 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 = $[-1.29, 0.36]$).

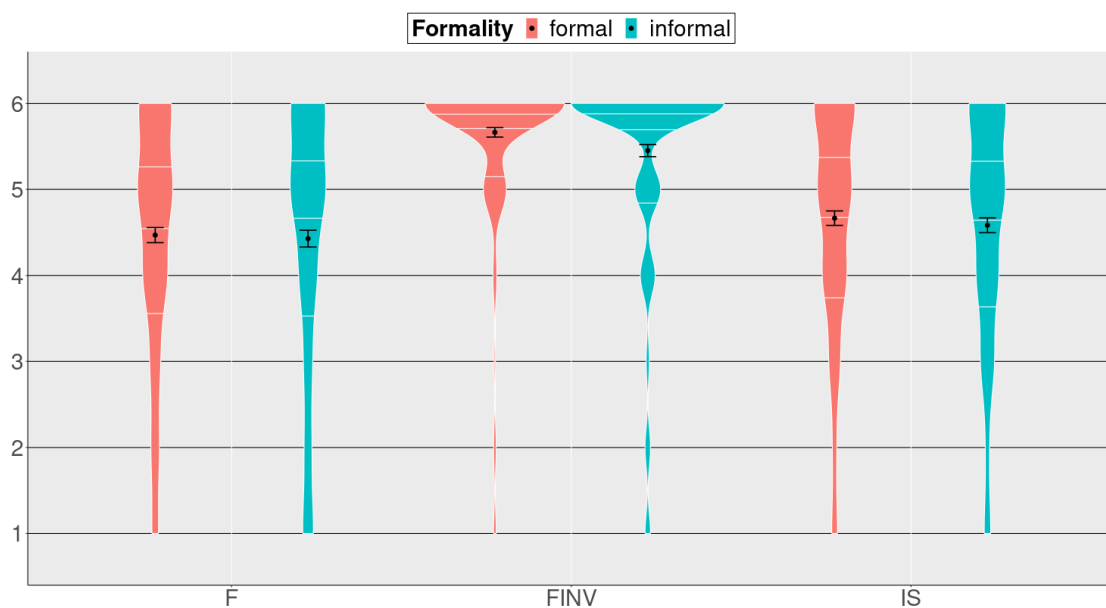


Figure 8: Mean acceptability ratings from Experiment 2, by context (good French).

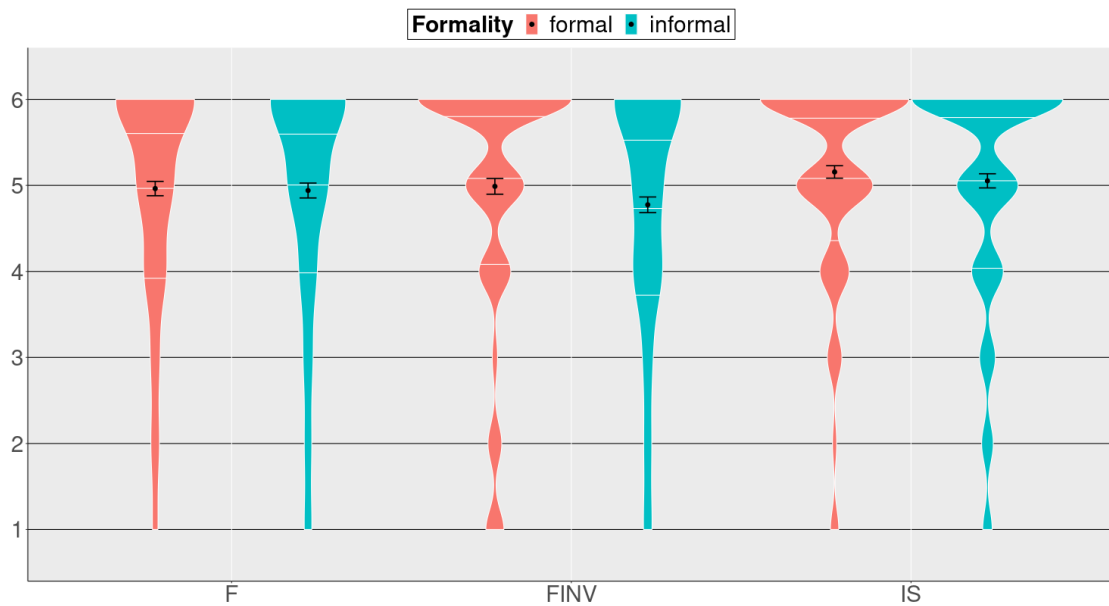


Figure 9: Mean acceptability ratings from Experiment 2, by context (suitable French).

For the AGE variable and on the “well-formedness” scale, the overall pattern does not change across 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 = [-.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* ($\hat{\beta} = -0.80$, $P(\beta < 0) = 0.96$, 95%CrI = [-1.75, 0.13]) contrasts.

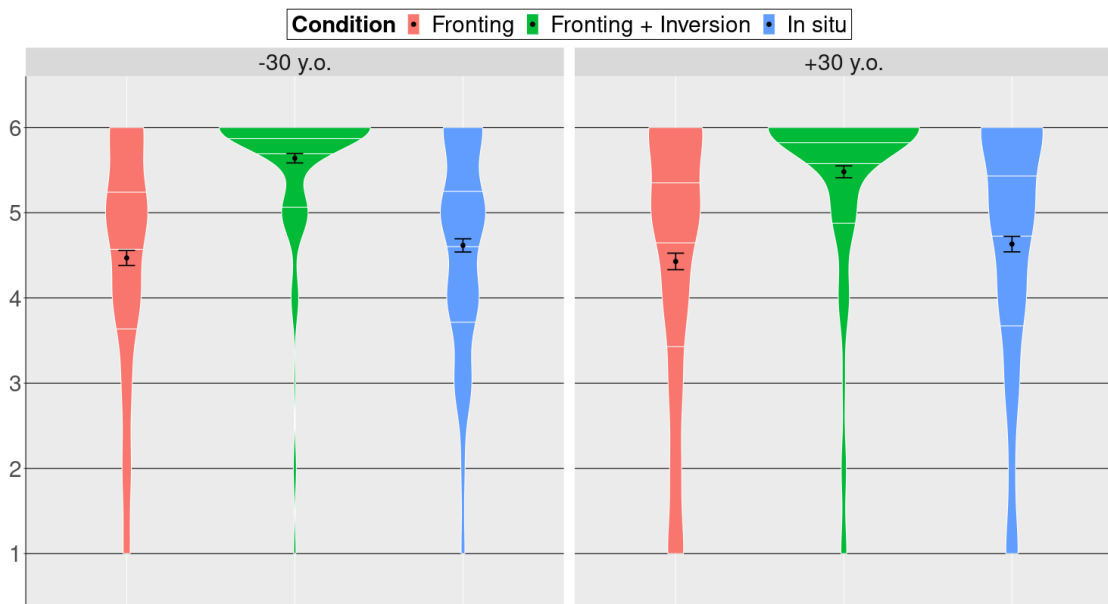


Figure 10: Mean acceptability ratings from Experiment 2, by age group (good French).

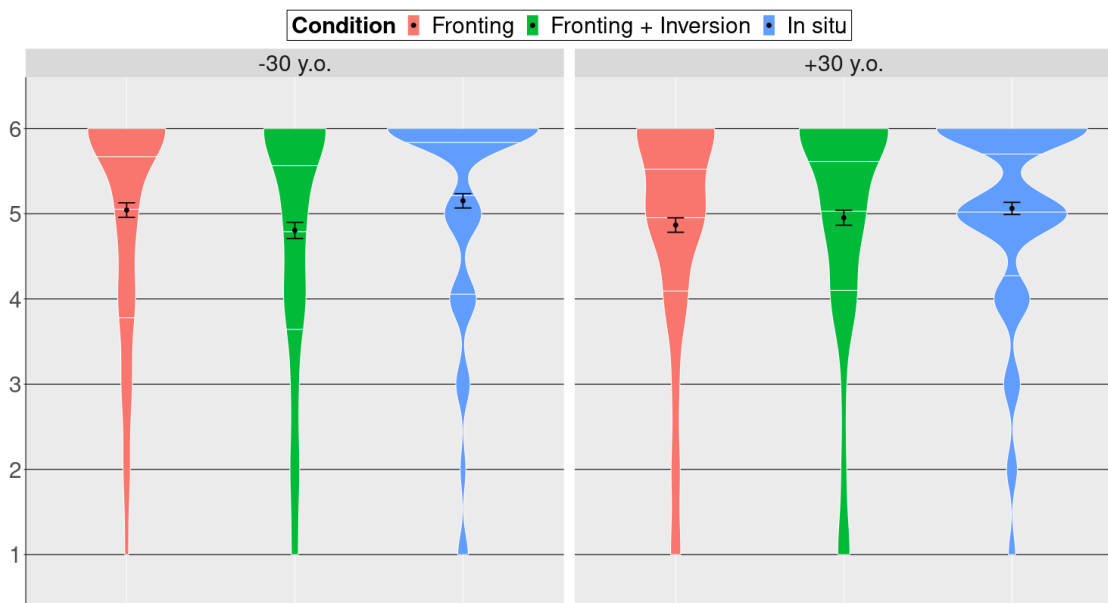


Figure 11: Mean acceptability ratings from Experiment 2, by age group (suitable French).

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

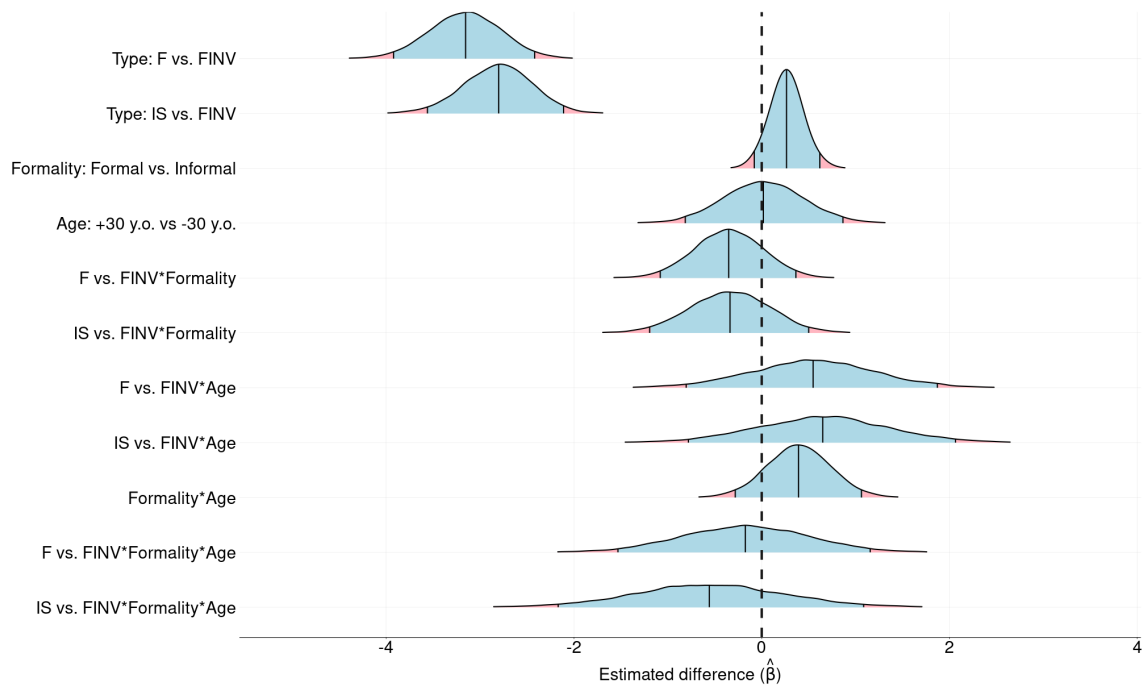


Figure 12: Posterior distributions for the maximal model run for Experiment 2 (good French).

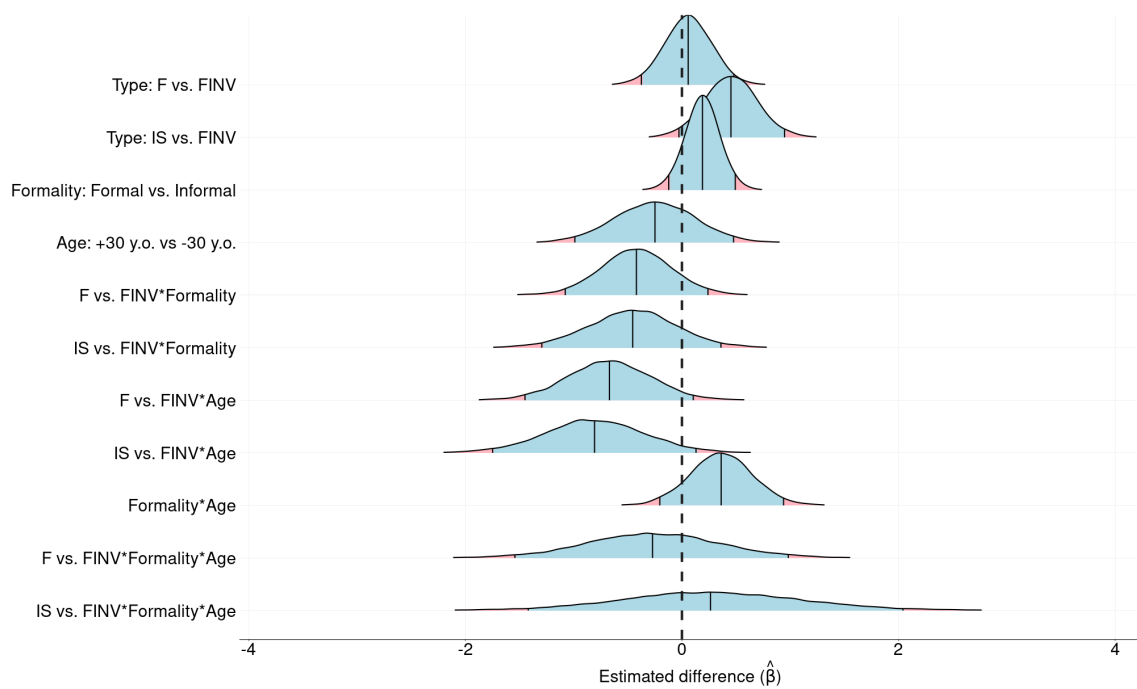


Figure 13: Posterior distributions for the maximal model run for Experiment 2 (suitable French).

Different from previous results, there seems to be no 3-way interaction $TYPE*FORMALITY*AGE$ for both the *fronting* vs. *fronting-with-inversion* and 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 to 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 in the very similar patterns on the well-formedness scale here and on the unique scale used in Experiment 1. *Fronting-with-inversion* sentences are on this scale preferred to both *fronting* and *in-situ* sentences, and this preference is modulated by context formality.

More importantly, this experiment provides evidence for a nuanced approach by participants to the concept of “acceptability”. Well-formedness does not entail real language usage, and real language usage does not follow from well-formedness alone. The pattern on the first scale (“good/proper French”) is close to that observed in a general acceptability task (i.e. Experiment 1) and closely reflects what can be found in a reference grammar for the language. However, this scale alone would miss participants’ proficiency in selecting which variant would best suit a real-life context of interaction. This is more finely 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 partial interrogation at least, but most probably not only). These interactions are still present on the second scale, which would indicate that not all age groups deem all variants as “suitable” to a given context. While there is no evidence for a three-way interaction with the FORMALITY predictor in Experiment 2, this is still reminiscent of Thiberge, Badin, and Liégeois (2021)’s 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.

4.3 A quick look at the “est-ce que” partial interrogatives (Experiment 3)

Experiment 2 gave a more fine-grained take on the acceptability of the three main variants available to French speakers to ask a partial question. Like Experiment 1 however, it ignored the *fronting* + *ESK* 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 ran a third experiment on the basis of the paradigm applied in Experiment 2 (two different scales, one for well-formedness and one for suitability to 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 *fronting* + *ESK*-type and *in-situ*-type interrogatives. Interestingly, *fronting* + *ESK* interrogatives appear in an intermediate position between *fronting-with-inversion* and *in-situ* sentences.

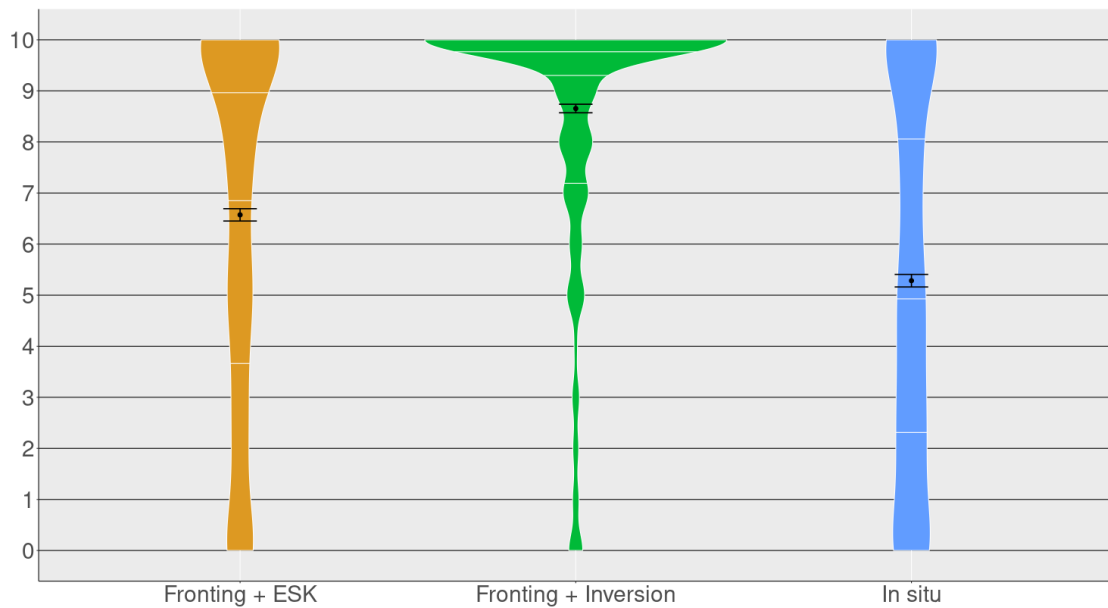


Figure 14: Mean acceptability ratings from Experiment 3 (good French).

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 yielded lower ratings overall. Even though they are rated “better French” than *in-situ* sentences, *fronting + ESK* sentences overall seem to not be suitable for many contexts. Also to be noted, 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. One explanation could be that the presence of *fronting + ESK* sentences in the stimuli made the difference between *fronting-with-inversion* and *in-situ* less salient to participants.

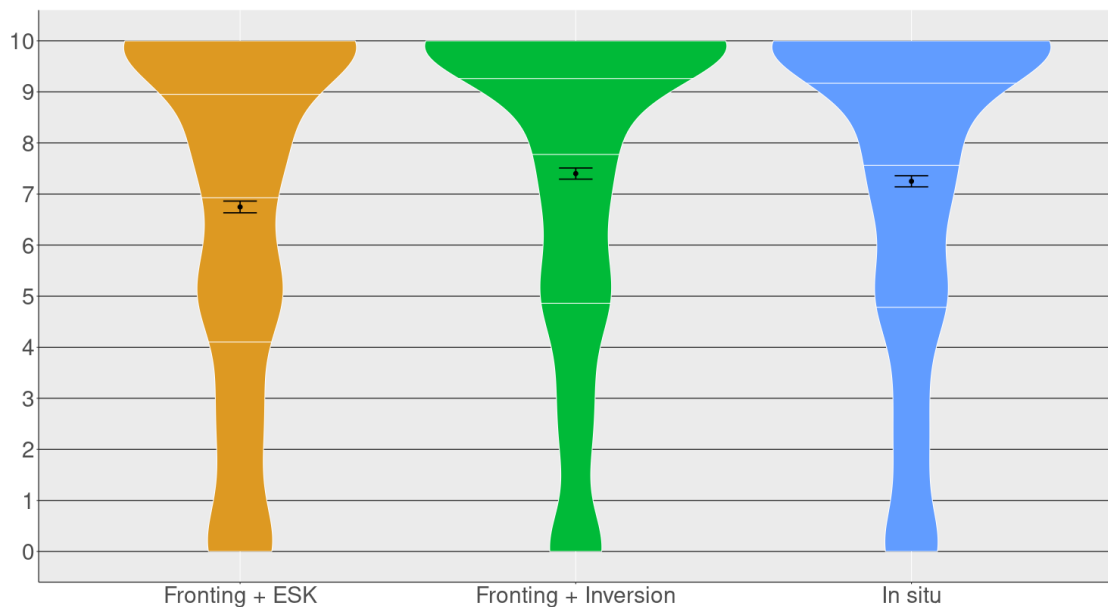


Figure 15: Mean acceptability ratings from Experiment 3 (suitable French).

For FORMALITY, and on the “good French” scale (Figure 16), the pattern appears once more similar to that observed in previous Experiments: *fronting-with-inversion*- sentences

are rated higher than both other types overall. This time however, there is no evidence of an interaction TYPE*FORMALITY, for both the *fronting* + *ESK* vs. *fronting-with-inversion* and the *in-situ* vs. *fronting-with-inversion* contrasts. An explanation could be that since *fronting* + *ESK* 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 3 variants, which differs from Experiment 2 where *fronting* and *in-situ* sentences were more or less alike in Experiment 2 for instance.

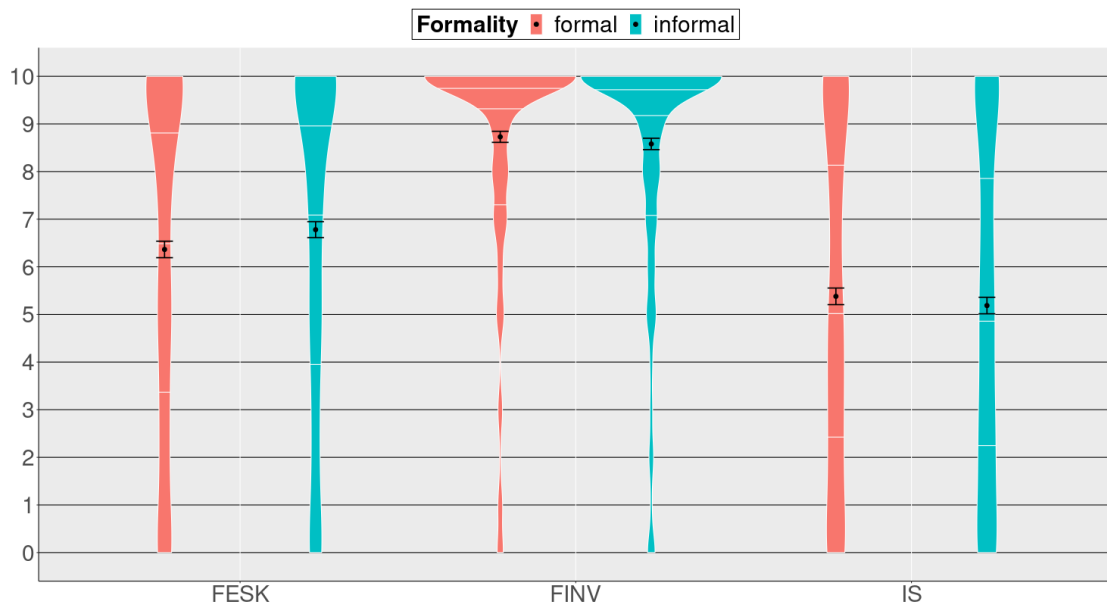


Figure 16: Mean acceptability ratings from Experiment 3, by context (good French).

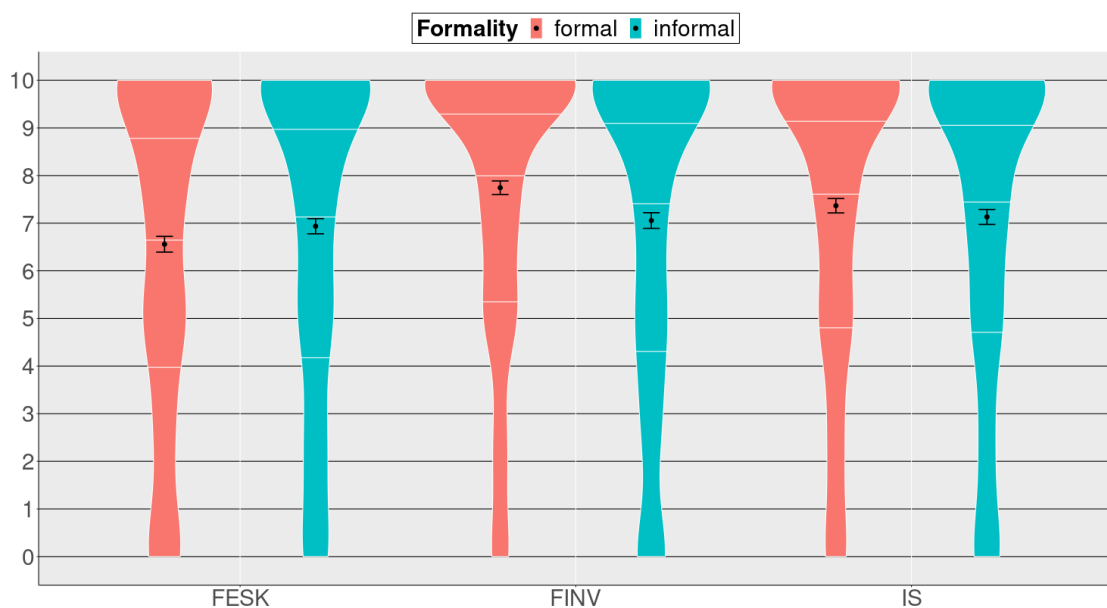


Figure 17: Mean acceptability ratings from Experiment 3, by context (suitable French).

On the “suitability” scale (Figure 17), the pattern is different, with all 3 variants rated between 6.56/10 (*fronting* + *ESK* 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 *fronting* + *ESK* yield lower ratings in informal contexts, while *in-situ* sentences appear to not be sensitive to differences in formality.

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

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 appreciating different variants differently in different contexts.

When aggregating all these results from Experiment 3, it seems *fronting* + *ESK* 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 to, in particular, formal contexts. We also observe a different appreciation of this variant in different contexts across age groups, which is in line with both previous experiments and corpus work. *fronting* + *ESK* seems to be a variant that works more or less 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 that a one-dimensional approach to acceptability may not be appropriate. With Experiment 1 we show that, given only one general scale, participants conflate “acceptability” with a normative view on what the standard for their language is. In this sense, acceptability becomes somewhat the same as prescriptive grammaticality, 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. The *fronting-with-inversion* variant was the “best-formed” variant (as described in reference grammars), above *fronting* + *ESK* sentences, which in turn were better-formed than *in-situ* sentences and *fronting* sentences. As for “suitability” however, all variants seem to serve their proper purpose, with *fronting-with-inversion* sentences yielding lower ratings in informal contexts and *fronting* + *ESK* 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 distribution of usage found in real-life interactions such as the ESLO corpus.

In this paper, we focused on one specific alternation phenomenon of a specific linguistic system (partial interrogatives in French L1 adult speakers). We found that preferences for the different available forms varied across speaker groups (age) but 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 equate a general normative view of what a “proper sentence” should look like in said language, in par with what can be found in reference grammars (e.g., for French, Riegel, Rellat, and Rioul (2014)).

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 changed participants’ ability to distinguish between what is “proper French” and what is “French one can use in a particular context”. We take this as further evidence of the discrepancy between “normative French” as described in reference grammars and taught at school and “everyday French” as used in everyday interactions. French adult speakers have perfectly internalized this difference, 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). 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”.⁸

This subtle distinction between “normative *acceptability*” and a more general “*suitability*” to a given situation is relevant from a sociolinguistic perspective and particularly in a third-wave approach to 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, in accordance with what social *persona* (Ochs, 1992) they want to convey.

In line with other works trying to pinpoint the nature of *acceptability* and how *acceptability* judgments help characterize it (Huang & Ferreira, 2020; Hubers et al., 2020; Tubau et al., 2020, i.a.), we argue that this distinction has to be taken into account in psycholinguistic studies more generally. 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: sociolinguistic demand for a form, or variant, that is suitable to the particular context it is produced in. Consciously or not, participants in language studies use the distinctions and nuances 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.⁹

The sociopragmatic ability to evaluate the suitability of a given linguistic production to the particular context 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 ac-

⁸ See Abeillé et al. (2023) for an overview of the current debates on French and linguistic norms.

⁹ 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.

quiring information. In this view, acceptability is more than a reflection of grammatical experience or an inverse correlate to surprisal (see for instance the (extended) radical unacceptability hypothesis put forth by Culicover, Varaschin, and Winkler (2022)). Acceptability is a socially situated judgment that should be assessed accordingly. 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 where respecting the linguistic norm is essential, such as a job interview).

Based on the results of our experiments, we suggest that, when assessing the preferences of speakers, the context of interaction needs to be controlled and accounted for very carefully. Drawing from the experiments we presented, this can be done by manipulating context formality, but also the social relationship between the persons present in the interaction (whether it is a balanced relationship with people from the same social group, or an unbalanced one with people who belong to different groups and have different normative assessments of the linguistic forms they use). And above all else, experiments where stimuli are presented in isolation are not sufficient to clearly establish the current status of a linguistic phenomenon. In parallel, experiments which 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.

Abbreviations

VSO: Verb-Subject-Object SVO: Subject-Verb-Object ESLO: Enquêtes sociolinguistiques à Orléans (Eshkol-Taravella et al., 2011; LLL, 2017)

Gloss

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 availability/Supplementary files

All supplementary materials are available at https://osf.io/gjdc4/?view_only=97693d587ddb4a50878cbd4b97cf1ae1 (detailed anal-

yses, R scripts, graphs and materials + norming data for all experiments). DOI: <https://doi.org/.../...> (will be added for non-anonymous version)

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 on their legal rights before taking part in the experiments: no known risk was associated with taking part in the experiments, they could withdraw of the study at anytime 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.

Funding information

Will be added for non-anonymous version

Acknowledgements

Will be added for non-anonymous version

Competing interests

The authors have no competing interests to declare.

Authors' contributions

Author1 conducted the experiments described in this paper as well as the statistical analyses, and redacted the first draft of this paper. Author2 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.

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., Et al. (2023). *Le français va très bien, merci*. Gallimard.
- Abeillé, A., & Godard, D. (2021). *La grande grammaire du français*. Éditions Actes Sud.
- Adli, A. (2015). What you like is not what you do: Acceptability and frequency in syntactic variation. *Variation in language: Usage-based vs. system-based approaches*, 173–200. https://romanistik.phil-fak.uni-koeln.de/sites/romanistik/Mitarbeiter/Adli/Adli_2015.pdf

- 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/>
- 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://www.ncbi.nlm.nih.gov/pmc/articles/PMC3881361/>
- 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.
- 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). Acceptability prediction by means of grammaticality quantification (N. Calzolari, C. Cardie, & P. Isabelle, Eds.). 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*, Sydney, Australia, Association for Computational Linguistics. <https://doi.org/10.3115/1220175.1220183>
- Brehm, L., & Alday, P. M. (2020). A decade of mixed models: It's past time to set your contrasts, In *The 26th architectures and mechanisms for language processing conference (amlap 2020)*. <https://amlap2020.github.io/a/131.pdf>
- Bürkner, P.-C. (2018). *Advanced bayesian multilevel modeling with the r package brms. the r journal*, 10 (1), 395–411. doi: 10.32614 (tech. rep.). RJ-2018-017. <https://arxiv.org/pdf/1705.11123.pdf>
- Bürkner, P.-C. Et al. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1), 1–28.
- Burnett, H. (2017). Sociolinguistic interaction and identity construction: The view from game-theoretic pragmatics. *Journal of Sociolinguistics*, 21(2), 238–271. http://www.heatherburnett.net/uploads/9/6/6/0/96608942/jofs_burnett.pdf
- Burnett, H. (2019). Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy*, 42(5), 419–450.
- Burnett, H., & Pozniak, C. (2021). Political dimensions of gender inclusive writing in parisian universities. *Journal of Sociolinguistics*, 25(5), 808–831. <https://doi.org/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://www.osti.gov/servlets/purl/1430202>
- Chomsky, N. (1965). Aspects of the theory of syntax. *Cambridge, MA: MIT Press*.
- Chomsky, N. (1981). Lectures on government and binding (dordrecht: Foris). *Studies in generative grammar*, 9.
- Coveney, A. (2011). L'interrogation directe. *Travaux de linguistique*, 2, 112–145. <https://www.cairn.info/revue-travaux-de-linguistique-2011-2-page-112.htm>
- Cowart, W. (1997). *Experimental syntax: Applying objective methods to sentence judgments*. Thousand Oaks, CA: Sage.

- Culicover, P. W., Varaschin, G., & Winkler, S. (2022). The radical unacceptability hypothesis: Accounting for unacceptability without universal constraints. *Languages*, 7(2), 96.
- Dillon, B., & Wagers, M. (2019). Approaching Gradience in Acceptability with the Tools of Signal Detection Theory (G. Goodall, Ed.). In G. Goodall (Ed.), *The Cambridge Handbook of Experimental Syntax*. Cambridge University Press. <https://www.cambridge.org/core/books/abs/cambridge-handbook-of-experimental-syntax/approaching-gradience-in-acceptability-with-the-tools-of-signal-detection-theory/F9CC8B04075505E7AAC1F6BEC4A50D8E>
- 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
- Drummond, A. (2016). Iboxfarm (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, 87–100. <https://pdfs.semanticscholar.org/89d3/de3d9d756fd39cd0b2b1270c309feb4a49a7.pdf>
- 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. Ms. 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.
- Gibson, E., & Fedorenko, E. (2010). Weak quantitative standards in linguistics research. *Trends in cognitive sciences*, 14(6), 233–234.
- Gibson, E., & Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes*, 28(1-2), 88–124.
- Ginzburg, J., & Sag, I. (2000). *Interrogative investigations*. Stanford: CSLI publications. <https://www.bookdepository.com/Interrogative-Investigations-Jonathan-Ginzburg/9781575862781>
- 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/https://doi.org/10.1016/j.jml.2019.04.004>
- Hamblin, C. L. (1973). Questions in montague english. *Foundations of Language*, 10(1).
- Hamlaoui, F. (2009). *La focalisation à l’interface de la syntaxe et de la phonologie: Le cas du français dans une perspective typologique* (Doctoral dissertation). Paris 3.
- Huang, Y., & Ferreira, F. (2020). The application of signal detection theory to acceptability judgments. *Frontiers in Psychology*, 11, 498268.
- Hubers, F., Redl, T., De Vos, H., Reinarz, L., & De Hoop, H. (2020). Processing prescriptively incorrect comparative particles: Evidence from sentence-matching and eye-tracking. *Frontiers in psychology*, 11, 498796.
- Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and philosophy*, 1(1), 3–44.
- 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*. Phd Dissertation, University of Edinburgh.
- Langsford, S., Perfors, A., Hendrickson, A. T., Kennedy, L. A., & Navarro, D. J. (2018). Quantifying sentence acceptability measures: Reliability, bias, and variability. *Glossa: a journal of general linguistics*, 3(1), 1–34.
- Lau, J. H., Clark, A., & Lappin, S. (2017). Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Science*, 41, 1202–1241.
- 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: a journal of general linguistics (2016-2021)*, 5(1), 72.
- Ochs, E. (1992). 14 indexing gender. *Rethinking context: Language as an interactive phenomenon*, 11, 335. http://www.sscnet.ucla.edu/anthro/faculty/ochs/articles/Ochs_1993_Indexing_Gender.pdf
- 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.
- Pohl, J., & Straka, G. (1965). Observations sur les formes d'interrogation dans la langue parlée et dans la langue écrite non littéraire, In *Actes du xe congrès international de linguistique et philologie romanes*. Klincksieck.
- Posit team. (2023). *Rstudio: Integrated development environment for r*. Posit Software, PBC. Boston, MA. <http://www.posit.co/>
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Reinhardt, J. (2019). 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>
- 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. Unpublished manuscript. 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. <https://files.eric.ed.gov/fulltext/ED016965.pdf>
- Sag, I. A. (2010). English filler-gap constructions. *Language*, 486–545. <https://www.jstor.org/stable/40961690?seq=1>
- Schütze, C. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.
- Söll, L. (1983). L'interrogation directe dans un corpus de langage enfantin. *Études de grammaire française descriptive*, 45–54.
- Sorace, A., & Keller, F. (2005). Gradience in linguistic data. *Lingua*, 115(11), 1497–1524.
- 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: a journal of general linguistics*, 2(1).

- 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.
- 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.
- Wagner, S. E. (2012). Age grading in sociolinguistic theory. *Language and Linguistics Compass*, 6(6), 371–382. <https://scinapse.io/papers/1664970667>