Research Article

Gabriel Thiberge*, Heather Burnett and Barbara Hemforth

Developing sociolinguistic evaluative competence: the case of French WH-interrogatives

https://doi.org/10.1515/opli-2025-0072
Received December 31, 2024; accepted September 16, 2025; published online January 9, 2026

Abstract: As children acquire the core rules of their language, they also develop the ability to use it in a sociolinguistically adapted way. Although some works have addressed the production of sociolinguistic variables, few have explored the development of a sociolinguistic evaluative competence in children and their ability to pair linguistic forms and social meanings. With an aged-adapted matched-guise paradigm, we investigate how monolingual children aged between 3 and 11 years old (N = 136) acquire French WH- interrogatives. The acquisition of the variants available in adult French has often been linked to their syntactic complexity, but the quantitative data we present offer new insights. By comparing the adult and child networks of social representations associated with interrogative variants, we show that some linguistic forms carry social meaning very early on. This is evidence for a more complex picture of the factors weighing on how children may acquire competing syntactic variants.

Keywords: experimental sociolinguistics; language acquisition; social persona; French; wh- interrogatives

1 Introduction

Although first language acquisition has developed into a robust area of study in the past 50 years, how children acquire sociolinguistic competence remains understudied. As Chevrot (2024) describes in his recent 'state-of-the-art' chapter, studies such as Roberts and Labov (1995) and Kerswill (1996) investigated sociolinguistic variation in child language in the 1990s, but "it was not until the second decade of the twenty-first century that communities of researchers began to structure themselves around the organization of scientific events or the production of books and journal issues devoted to the acquisition of sociolinguistic variation" (Chevrot 2024: 2). Likewise, Smith and Durham (2019) attempt to bridge the gap between language acquisition researchers and variationist approaches to adult language use, saying "Sociolinguistic research tells us how adults of particular varieties vary with respect to [linguistic] structures. What we know far less about is how these combine in the acquisition of variation at the very earliest stages" (Smith and Durham 2019: 4). According to Chevrot and Foulkes (2013) and Chevrot (2024), research on sociolinguistic acquisition has three objectives: "to describe how children produce, perceive, and evaluate sociolinguistic variables at different ages; to understand how factors such as input, maturation, and socialization influence and drive developmental changes; and to model sociolinguistic knowledge and its interaction with social knowledge at different stages of development" (Chevrot 2024: 6–7). As these authors describe, there has been much progress on parts of this research program in recent years, and there is now a substantial body of work documenting and analyzing in detail whether/how children of different ages use different sociolinguistic variants in different contexts. Although much smaller, there is now also a developing line

***Corresponding author: Gabriel Thiberge**, Université Paris Cité, Laboratoire de Linguistique Formelle, CNRS & Université Toulouse Jean Jaurès, CLLE, CNRS, Toulouse, France, E-mail: gtac@tuta.io. https://orcid.org/0000-0002-9727-7860
Heather Burnett and Barbara Hemforth, Université Paris Cité, Laboratoire de Linguistique Formelle, CNRS, Paris, France. https://orcid.org/0000-0002-7389-0679 (H. Burnett). https://orcid.org/0000-0001-8371-1323 (B. Hemforth)

of research that studies how children of different ages cognitively perceive sociolinguistic variables (see Rosenthal 1974; Buson 2009; Kinzler and DeJesus 2013, among others). This being said, the evaluation aspect of sociolinguistic competence has been largely left aside, with very few works specifically studying the social categories or properties that children associate with (or *index* (Eckert 2008; Ochs 1992)) different sociolinguistic variants. This is unfortunate, since understanding sociolinguistic evaluation has been known to be crucial to understanding sociolinguistic variation, and therefore sociolinguistic competence, since the very beginning of variationist sociolinguistics (Labov 1964, 1966). The interest in the role that the social properties/categories associated with variants (aka their *social meanings*) play in how speakers use them has only grown in recent years (see Eckert 2012; Eckert and Labov 2017, among others); we therefore argue that the paucity of studies of sociolinguistic evaluation by children constitutes a serious gap in the literature.

This paper aims to help fill this gap through presenting a sociolinguistic evaluation experiment of French WH- interrogatives with child participants (3–11 years old) and adult controls. WH- interrogation in French is a complex system of alternative syntactic variants that convey a similar meaning in a classic semantic analysis. For example, the three syntactic structures in (1) are all possible ways of saying "When will you arrive?" in French: (1-a) is the *fronting* + *inversion* variant, in which both the adjunct WH word *quand* 'when' and the finite verb *arriveras* 'will arrive' are fronted. (1-b) is the *fronted* variant, in which only the WH word is fronted, and (1-c) is the *In Situ* variant, in which neither the WH word nor the verb are fronted.

(1) a. *Quand* arriveras-tu ? **Fronting + inversion (FINV)**
b. *Quand* tu arriveras ? **Fronting (F)**
c. Tu arriveras *quand* ? **In situ (IS)**
'When will you arrive?'

Corpus studies such as Quillard (2001), Tailleur (2013), and Thiberge et al. (2021) have shown that these three structures are in sociolinguistic variation, with the FINV form (1-a) being the "standard/prestige" variant, i.e. the variant favoured in formal contexts and by speakers with higher socio-economic status and/or from intellectual professions. French WH- interrogation is therefore an ideal sociolinguistic variable with which to study social meaning and evaluation. By far, the most common way that social meaning has been studied experimentally is through the *Matched Guise Technique* (MGT) (Lambert et al. 1960). In a MGT experiment, participants listen to samples of recorded speech or read short texts (called *guises*) that are designed to match as much as possible, differing only in the linguistic phenomenon studied. Each participant is exposed to only one of the guises, and after hearing it, their beliefs and attitudes towards the speaker are assessed, usually via questionnaire (see Kircher 2015 for overviews of this methodology). The original uses of the MGT were to study language attitudes, i.e. listeners' inferences about the properties of speakers of different languages; however, in the past 15 years, the MGT has become widely used within variationist sociolinguistics to study subtle differences in the evaluation of sociolinguistic variants in adults (see Campbell-Kibler 2006, 2007; Levon 2007, among many others). To our knowledge, very few MGT experiments have been conducted with young children. This is not an accident: the MGT involves being able to associate sociolinguistic perception with particular linguistic predicates on scales which, in many cases, are related to concepts such as *prestige*, *status*, *solidarity* or *wealth* (Kircher 2015). However, as Chevrot (2024) describes in his review, "while the judgments of the youngest children are based on the truth value of utterances or politeness, it is only at the age of twelve that the first justification based on prestige and correctness are produced" (Chevrot 2024: 8).

It is therefore clear that, for the standard psycholinguistic tool for studying sociolinguistic evaluation to be appropriately used with children, some modifications must be made. This is what we do in this paper, with a simpler design presented to children than to adults (less experimental conditions). We also have more appealing materials: scales that are not abstract and numbered, but rather ranging from one character to another, each incarnating a social characteristic, in the same vein as for example the *smiley* scales used by Ambridge (2010). We suggest that such modifications of the MGT paradigm can be helpful to further exploring this highly understudied aspect of the development of sociolinguistic competence. Our results show that in fact some associations between language and social properties emerge very early on (as early as age 4 for some properties), but that these associations are defined within their very individualized perception of the world. We follow Buson and Billiez

(2013) and hypothesize that children build these associations progressively, by salience and stereotypical processes, but that the perspective children have on the world is very different from adults'. For instance, the concept of *richness* is not the same for a child and for an adult, whose experiences of the world are very different and who will associate 'richness' to different subsets of society and/or different social behaviors. As a consequence, it is only natural that children start by having have different associations between sociolinguistic variants and social properties than adults, before becoming more adult-like around age 10–11. In this way, our data allows for a more fine-grained take on how the link between language and social properties evolve in children.

The paper is structured as follows: First, we present a literature review on sociolinguistic evaluation in children but also on the acquisition of French WH- interrogatives. Then, we present data from a matched-guise task where adult L1 speakers of French provided judgments on the stereotypical social properties of people using the three main WH- interrogative variants. It appears users of the fronting + inversion variant (example (1-a) above) are associated with properties such as richness and higher education. These results from adults are a crucial control point to which we can compare the results we obtained with children, which we present in another section. This gives a first window into the building of the network of associations between linguistic forms and social meanings through the years. In a final section we discuss the general findings of our study and their implications for the study of language acquisition in general.

2 Literature review

Although some studies have looked into how children perceive language and how they progressively link sociolinguistic variants to different social traits, they are still fairly scarce. What little work there is mostly focuses on social traits like warmth or likeability, which we argue is a somewhat indirect measure of the precise network of relationships children have to build between linguistic forms and social meanings (i.e. social stereotypes).

2.1 The emergence of a sociolinguistic evaluative competence

Labov (1964) develops a very prominent model of how children acquire and develop their sociolinguistic overall abilities and can be summarized as a 4-stage sequential process. In this model, children under the age of 5 (*basic grammar* stage) first acquire the core grammatical rules and the basic lexicon of their linguistic system, based mainly on the input they receive from their parents. Between age 5–12, children build up their vernacular variety (local dialect) by receiving input from their expanding peer group. In the *early adolescence* stage, children then begin to develop some degree of consciousness of the social significance of their local dialect when they are confronted to other dialects, mostly at school and by crossing paths with more adults. In the *late adolescence* stage (also called *stylistic variation* stage), children progressively learn how to adapt their productions according to the requirements of particular interactions, such as using standard variants in formal situations. This model is inspired by a grammatical framework where general rules precede variable rules, and although it has been criticized and/or amended by decades of following research, this sequential view is still a very prevalent take on how children develop language.

As Chevrot (2024) points out, there is however still a general dearth of empirical investigations into how children perceive sociolinguistically charged language. In an overview of research ranging from 1958 to 2013, Nardy et al. (2013) found 39 studies investigating the production of sociolinguistic variants in children. These studies have almost exclusively focused on the production of phonetic/phonological variants – mostly in English –, probably as a consequence of a long-standing tradition of sociophonetic studies in sociolinguistics. This line of research, however, has provided solid evidence for an earlier ability in children to produce sociolinguistic variants in distinct contexts. Roberts (1994) for example found that children aged between 3;2–4;11 are able to alternate between non-standard/vernacular *-in* and standard *-ing* verbal endings according to their interlocutor (more standard productions when directed to an adult). In French, works such as Chevrot et al. (2011) have linked the production of *optional liaison* by children aged from 2;3 to 6;0 to their socioeconomic status (SES). Older

children produce more of these non-mandatory phonological alternations (e.g. *gros éléphant* 'big elephant' which can be pronounced either [gʁozelefɑ̃] with the liaison or [gʁoelefɑ̃] without it) overall, in a target-like fashion, but children from higher SES backgrounds use them more proficiently, and earlier, than children from lower SES backgrounds.

Such works have brought forth evidence for a more nuanced picture than that proposed in Labov (1964), with heavy influences from usage-based approaches to language use, and in particular Bybee (2006)'s exemplar theory. Influenced by their respective linguistic input (which is highly constrained by socioeconomic factors), children's early linguistic productions are a first window into their sensitivity to sociolinguisic variation. To consistently tease apart contexts where one variant is more suited than another, they have to assess the social properties of said contexts much earlier than what Labov postulated in his earliest proposals.

Some works have studied the "evaluative" component of sociolinguistic competence more directly. Buson (2009) and Buson and Billiez (2013) for example assessed the conscious evaluative capacity of children with regard to socially charged language. By asking children aged 9–11 to comment on answering machine messages with different French registers, these works establish that "metapragmatic comments" are already possible at around that age, with explicit judgments such as "this isn't a proper way to speak French". With a few English sentences that are not sociolinguistically marked, Vaughn and Becker (2024) documented the emergence of the 'social-semiotic landscape' of 94 American English speaking children aged between 5 and 12 y.o. The authors ran questionnaires that explicitly targeted how the children felt about a person they had just heard, but also why they felt that way. They binned the answers according to what type of knowledge or perspective was used: for instance, did the child compare what was what said to what he/she likes to do, did the child place the person in a social type of macro-demographic category, etc. The authors show that the proportion of 'public'- and 'evaluative'- type answers increase with age. These comments either contain explicit reference to an ideological schema, i.e. what some categories of people are supposed to talk like, or at least contain evaluations or descriptions of the speech signal itself, denoting a judgment driven by how the person spoke. Independent of sociolinguistic variation, these results illustrate how children become able to make social assessments over language and speakers in a progressive way.

For a more implicit assessment of younger children's evaluative sociolinguistic ability, Rosenthal (1974) used boxes speaking either African American Vernacular English (AAVE) or Standard English to elicit sociolinguistic attitudes in preschool children (age 3;0–5;11). The results point to a difference in perception with regard to *niceness* and other social traits (e.g. need for a gift) depending on the variety of English the children heard. Comparing Hawaiian Creole English to Standard English, Day (1980) found with the same protocol that these attitudes shifted between kindergarten and first grade, with influences from SES-related factors. A more recent work (Kinzler and DeJesus 2013) made use of a matched-guise-like paradigm where, as inspired by Lambert et al. (1960), children aged 5–6 had to locate different people on a variety of social scales after hearing some statements which were either native-accented (American English) or foreign-accented (French-accented English). A first experiment found that when the statements contained neutral sentences, American-accented speakers were rated higher on some scales (friendliness, niceness, 'living around here') but not on others (smartness, 'in charge'). In a second experiment with socially charged statements (American-accented speakers saying mean things, foreign-accented speakers saying nice things), foreigners became nicer, friendlier and smarter than American English speakers. In yet another matched-guise inspired experiment, Dollinger et al. (2024) investigated the perception of foreign-accented English with Canadian English speaking children aged from 7 to 12. They found that younger children (aged 7) seem more tolerant with regard to accentedness than older children (who rate accented Englishes harder to understand than Standard Canadian English), which the authors take as first evidence for the apparition of some degree of prejudiced attitudes against non-native accented language.

This – very scarce – body of research on child evaluative abilities has paved the way for a more systematic approach to investigating sociolinguistic abilities in children, both in production and evaluation of socially situated language. Our work aims at contributing to this line of research by assessing the latter in French, by focusing on a syntactic point of variation (French WH- interrogatives) and by expanding the social properties that are being assessed. To get a better view of how children develop their understanding of the relationships between language and social characteristics, we argue it is necessary to go beyond the in-group proximity measures that

focus on friendliness (for instance by asking children to also assess prestige-related traits), but also to have a more quantitative approach across age groups (with more children taking the tasks, which requires a rapidly and easily deployable experimental paradigm). To go beyond the traditional focus on sociophonetic points of variation, we argue that the system for WH- interrogation in French is an ideal case study.

2.2 WH- interrogatives in French and their acquisition

As mentioned in the introduction, there are different linguistic variants for French WH- interrogatives (1). In adult French, many different works have analyzed this alternation phenomenon, from different linguistic perspectives. From the syntactic point of view, it is generally established that the so-called *in situ* form ("Tu arriveras quand ?") is the base form from which all other variants derive, with different descriptions and constraints depending on the authors' theoretical frameworks (Abeillé 2007; Ross 1967; Sag 2010). Semantico-pragmatic approaches have documented the influence of information structure on the production of the different variants (Beyssade 2006; Chang 1997; Cheng and Rooryck 2000). Phonotactic studies have presented empirical evidence for the number of syllables of both the interrogative and non-interrogative parts of the question being a predictor of the variant used in modern French (Hamlaoui 2011), while both macro- and micro-diachronic changes have been documented with respect to the preferred variants (Larrivée 2019; Thiberge et al. 2021). Finally, different sociolinguistic approaches have linked the different uses of interrogatives to diatopic (Guryev and Delafontaine 2015; Mathieu 2009) and diastratic (Adli 2006; Coveney 2011; Quillard 2001) variation.

This amount of work illustrates how complex the WH- interrogative system is in French, yet only the syntactic factors seem to have permeated to acquisition research. It has been widely documented that *in situ* variants appear first in L1 French children productions, around age 2 but with considerable individual variation (Hulk 1996; Prévost 2009; Zuckerman and Hulk 2001). The derivation of other variants from the *in situ* one has led researchers to build a "complexity metric" according to which complexity increases gradually across variants so that they become more difficult to process and produce by typically developing children – but also by atypically developing children and L2 learners (Jakubowicz 2011; Scheidnes and Tuller 2010).

Another account (Faure and Palasis 2021; Palasis 2013) proposes a diglossic approach to the acquisition of the French WH- interrogation system. In this proposal, children acquire two different grammars one after the other: the first one is that of a language used with family and peers, while the other one is that which one uses outside these circles. This view is, in a sense, very similar to Labov (1964)'s, albeit with slightly different stages and with a heavier weight of syntactic constraints.

3 The social meaning of interrogative variants in French (adult matched-guise task)

Based on corpus data, Thiberge et al. (2021) give a first view into the complex network of factors at play in the use of the four main interrogative variants in modern French: all three variants from example (1) repeated below + one variant similar to (1-b) where the *est-ce que* idiomatic expression is added between the WH word and the subject (1-d).

(1) a. *Quand* arriveras-tu ? **Fronting + inversion (FINV)**
b. *Quand* tu arriveras ? **Fronting (F)**
c. Tu arriveras *quand* ? **In situ (IS)**
d. *Quand* est-ce que tu arriveras ? **Fronting + *est-ce que* (FESK)**
'When will you arrive?'

More precisely, these authors show that speaker age group and context formality (and the interaction between the two) can reliably predict speakers' preferences for a variant in a given situation. Following work in third-wave

sociolinguistics inspired approach to language use (Eckert 2012), where different social *personae* (Beltrama 2020; Ochs 1992; Podesva 2011) are built by speakers and perceived by listeners depending on what language is used, we hypothesize that this context-sensitive use of the different interrogative variants follows from different social meanings that are attached to them. In order to investigate the social meanings, we follow a long tradition in social psychology and experimental sociolinguistics, and use the matched-guise task (Kircher 2015; Lambert et al. 1960), a well-tested paradigm to assess the social properties that are assigned to speakers depending on some linguistic productions. From a methodological point of view, these results from an experiment with adults are a necessary step before exploring the relationship between linguistic forms and social meanings in children. Adult data is the reference level (or "target behavior" that children are developing over time) to which we can compare results from an experiment with children (see below).

3.1 Design and materials

We devised an auditory matched-guise experiment where participants had to listen to recordings of dialogues between two speakers (A and B) on a computer. Person A sets a small discursive context and person B asks a WH-question. After hearing each recording, participants had to assess six different social characteristics of person B (only) with six corresponding 7-point scales. The scales targeted dimensions such as Age, Richness, Sociability, Geographical Origin, Education, and Hobbies. For a better comparison of both the adult and the child experiments, we used pictographic representations of social properties rather than textual descriptions of the dimensions participants had to assess. The pictures used for this were normed in a separate experiment (see Supplementary Materials in the OSF repository at https://osf.io/7r6kv/). These six scales were selected among the nine that were normed because they yielded the best results on the social dimensions that were targeted, and because they were more reusable with children. Figure 1 illustrates a typical item (The text under the audio players reads: "The last person who speaks is more like…").

As for the experimental design, we manipulated two elements, with a TYPE variable for the syntactic variant of interrogative (3 levels: *in situ*, *fronting with inversion*, *fronting*) and with a FORMALITY variable for the context (2 levels: *informal* or *formal*), resulting in $2 \times 3 = 6$ different experimental conditions. 30 target items were created, with context formality being normed in a separate experiment. 30 filler items not containing any interrogative

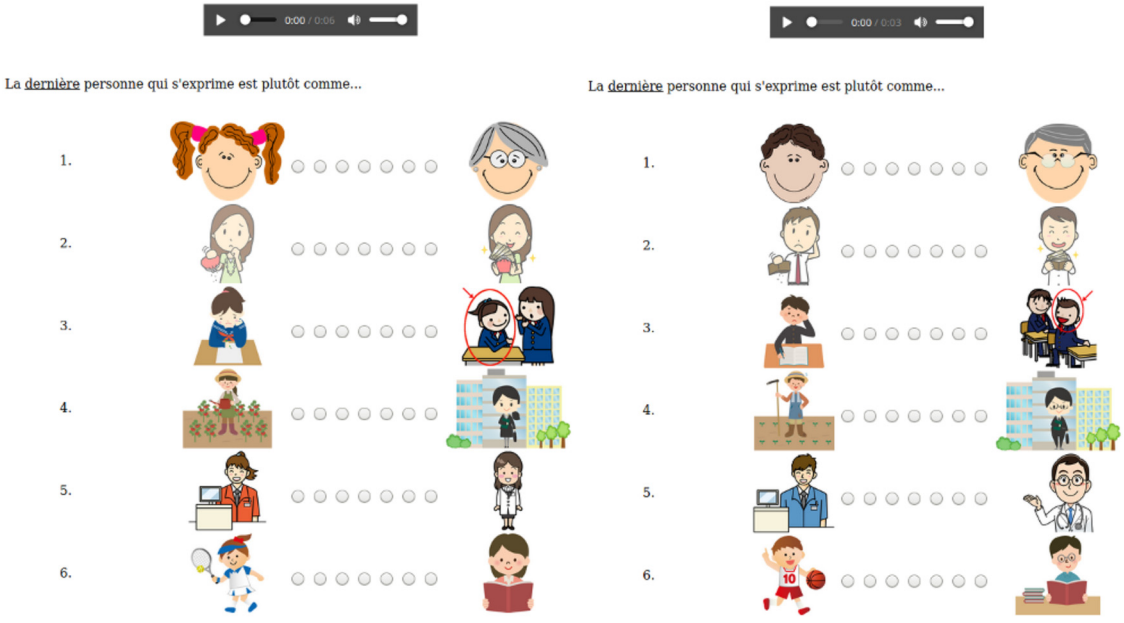

Figure 1: Example item from MGT-A as seen by participants.

Table 1: Example item from MGT-A in all its conditions.

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

were added, as well as 3 practice items. Four L1 adult speakers of French (F1, F2, M1, M2) were recorded in a soundproof room while they read all the items in all the conditions, with a naturalistic tone. The recordings were then rearranged in Audacity (AudacityTeam 1999-2021) into the A/B dialogues so as to get all possible combinations of voices (F1-F2, F1-M1, F1-M2, F2-F1, F2-M1, F2-M2, M1-F1, M1-F2, M1-M2, M2-F1, M2-F2, M2-M1). In another norming study we assessed the *a priori gender* of the voices so the drawings used for the scales would be coherent with them, and we checked that all four voices did not differ too much in terms of how colloquial they sounded to avoid introducing biases with regard to the formality variable. An example item in all its conditions can be found in Table 1, and a complete list of the materials can be found in the Supplementary Materials.

3.2 Analysis and predictions

Previous works (see above) have shown that, for continental French at least, fronting with inversion is associated with a more standard way of speaking French and is more used in formal contexts and by people older than 30 y.o. As a consequence, with participants from a relatively homogeneous group of participants (from France), we expected this specific variant to give rise to higher ratings on age, richness, urban background, longer education and frequent reading scales. On the sociability scale, we expected some kind of interaction between TYPE and FORMALITY, with fronting with inversion making the person sound a bit more 'out-of-place' or weird in more formal contexts.

We analyzed the respective and combined influences of our manipulated variables (TYPE, FORMALITY) on the ratings given by participants on the six different social scales. Ratings were given on 7-point scales, which we coded from -3 (drawing on the left) to +3 (drawing on the right). We used Bayesian cumulative-link models (one per scale) to investigate the relationship between our independent and dependent variables, since they are best suited for ordinal data.[1] Just like classical (frequentist) inferential statistics, these models yield numeric estimated coefficients (so-called *estimates*, $\widehat{\beta}$) that help assessing the influence between variables, but they do not yield binary p-values. Rather, they generate a probability distribution of many simulated estimates that help evaluate how trustworthy the yielded estimate is for a given predictor (or combination of predictors). The probability of the estimated coefficient to be greater, $P(\beta > 0)$, or lesser, $P(\beta < 0)$, than 0 indicates how reliable the evidence from the dataset is regarding the existence of an effect of the predictor under investigation. Given the number of datapoints in this experiment, we will report probabilities higher than 0.90 as "robust evidence" for the existence of an effect, and probabilities where $0.80 < P < 0.90$ as indicative of a trend. We will also report 95 % Credibility Intervals (95 % CrI), which are the numerical values of the posterior distribution between which there is a 95 % chance of finding the true value of the estimated coefficient.

In line with other works on French WH- interrogatives, we hypothesize that age group might also be predictive of some differences between the uses and ratings of different WH- interrogative variants, so we included it in the

1 For an overview of why the Bayesian framework is ideal for psycholinguistic and linguistic datasets, see Sorensen et al. (2016).

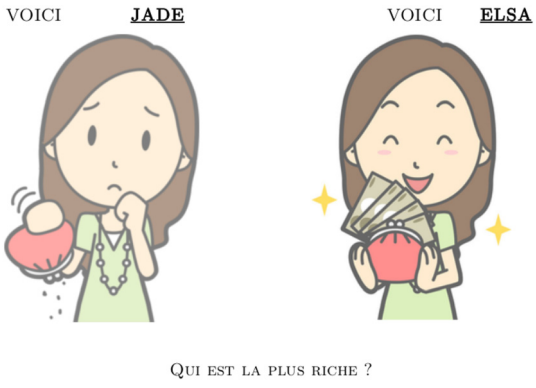

Figure 2: Familiarization screen from MGT-A.

analyses as a binary predictor (+/– 30 years old) although we did not control for it during participants' recruitment. This predictor, as well as the TYPE and FORMALITY ones were mean-center coded to allow for a more direct interpretation of the outputs (Brehm and Alday 2022) (–0.5/+0.5 coding for FORMALITY and age, with +0.5 being *formal* contexts and *more-than-30-year-old* respectively; for the TYPE variable the reference level was the FINV variant). 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). To account for possible variation across items and participants, we included these in the random effects structure of maximally-specified models, including suitable random slopes (Barr et al. 2013). The full technical specifications and outputs of the models can be found in the Supplementary Materials.

3.3 Participants and procedure

52 people from continental France took part in the experiment. They were recruited via the RISC network (CNRS, UMR 3352), social networks and among university students. For this experiment, age was not a controlled variable and the adult population was artificially considered homogeneous in its approach to the linguistic variants, to which the children data could then be contrasted. As such, participant age ranged from 18 to 77 (mean = 26.8, median = 22.5). The experiment was programmed on the IbexFarm platform (Drummond 2016) and hosted on university servers, but experimental sessions took place in the lab. Participants listened to the stimuli in a soundproof room and rated the speakers on the scales on a computer. Before the experiment began, we informed participants of their rights, of the general nature of the task, and we presented them with the drawings they would see for the scales (12 different sets of contrasting pictures that explicitly introduced the targeted social dimension, see Figure 2, where the texts read: "Here is Jade. // Here is Elsa. // Who is richer?").

After, this, participants were also invited to locate themselves on a paper-printed version of the scales, to familiarize them with the materials and to get a better grasp of how they felt about their meanings. Then, the real experiment began. A session was approximately 30 minutes long and participants were compensated accordingly.

With 30 target items, 30 filler items and 3 practice items, we received $63 \times 52 = 3{,}276$ answers per scale. The analyses we present below focus on the 1,560 answers per scale for target items only.

3.4 Main results

3.4.1 Correlations

We first checked for correlations between answers. Figure 3 offers a global view of the network of correlations between scales, and thus between social dimensions, that were calculated.

The numbers on the bottom left part of the graph are the Spearman's ρ rank correlation coefficients for the different scales (a cross means the correlation was not statistically significant), with a more visual representation on the top right part (blue means a positive relation between values, the bigger the circle the higher the coefficient). All in all, and just as in the norming experiment where we assessed the social properties associated with all the pictures

Figure 3: Correlation matrix for the different scales from MGT-A.

we used (see Supplementary Materials), it appears there are some associations between scales. For example, answers on both the age and the richness scales are somewhat correlated ($\rho = 0.42$), but less so than the hobbies and education scales, where high education and reading seem to be quite often associated, $\rho = 0.57$.

3.4.2 General results

Figure 4 compiles all answers from participants. Each point from a given colored polygon represents the overall mean of answers for a given scale in a given condition. More than the precise numeric values for each point, it is the different shapes of the polygons that stand out, with a similar pattern for the *fronting* and *in situ* types and a different one for the *fronting with inversion* one.

An in-depth inferential statistical analysis (see Table 2, and the Supplementary Materials for a more complete view of the models outputs) yields robust evidence for virtually all of these differences, on all scales, that set the *fronting with inversion* type apart from the other two, with:

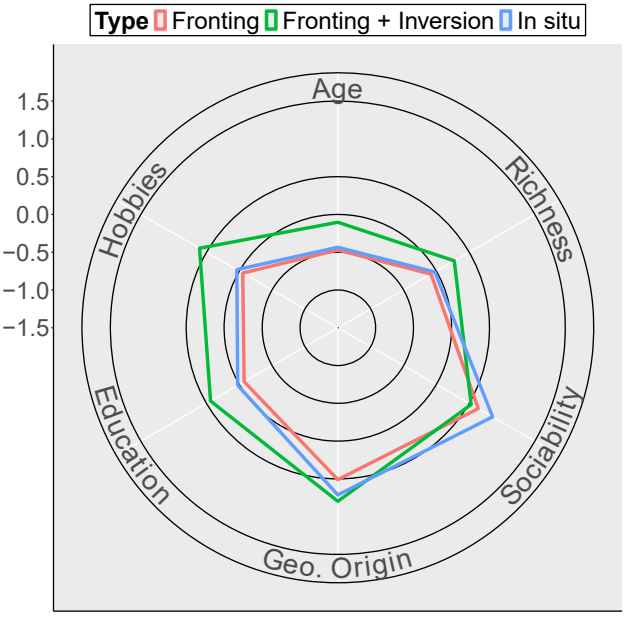

Figure 4: Overall results from MGT-A.

Table 2: Main results from the models for all scales.

Scale	Contrast	$\widehat{\beta}$	$P(\beta > 0)$	$P(\beta < 0)$	95 % CrI
Age	F versus FINV	−0.63		1	[−0.93, −0.32]
	IS versus FINV	−0.54		1	[−0.87, −0.21]
Richness	F versus FINV	−0.70		1	[−1.02, −0.37]
	IS versus FINV	−0.59		1	[−0.97, −0.20]
Sociability	F versus FINV	0.17	0.90		[−0.09, 0.43]
	IS versus FINV	0.56	1		[0.19, 0.94]
Geo. origin	F versus FINV	−0.42		1	[−0.73, −0.11]
	IS versus FINV	−0.15		0.85	[−0.42, 0.13]
Education	F versus FINV	−0.91		1	[−1.29, −0.55]
	IS versus FINV	−0.74		1	[−1.07, −0.40]
Hobbies	F versus FINV	−1.12		1	[−1.50, −0.72]
	IS versus FINV	−1.01		1	[−1.42, −0.61]

- F and IS sentences being associated with lower ratings on the "age" scale than FINV sentences (for both, $P(\beta < 0) = 1$).
- F and IS sentences being associated with lower ratings on the "richness" scale than FINV sentences (for both, $P(\beta < 0) = 1$).
- F and IS sentences being associated with higher ratings on the "sociability" scale than FINV sentences (respectively, $P(\beta > 0) = 0.90$ and $P(\beta > 0) = 1$).
- F and IS sentences being less associated with ratings pointing to an urban background on the "geographical origin" scale than FINV sentences (respectively, $P(\beta < 0) = 1$ and $P(\beta < 0) = 0.85$ – only a trend for IS sentences).
- F and IS sentences being associated with lower ratings on the "education" scale than FINV sentences (for both, $P(\beta < 0) = 1$).
- F and IS sentences being less associated with ratings pointing to reading on the "hobbies" scale than FINV sentences (for both, $P(\beta < 0) = 1$).

The combination of these suggests there is a specific social persona associated with using fronting with inversion: that of an above-average educated, older, richer person who reads books more often. On the other hand, using the fronting with inversion variant is generally perceived as socially weird.

3.4.3 Interactions

Interactions between the TYPE and FORMALITY variables point to some influence of the latter on the perception of the *persona* indexed by the variants, at least on some dimensions (see Table 3). More precisely, compared to FINV sentences:

- F sentences are associated with even lower ratings on the "geographical origin" scale in *formal* contexts than in *informal* ones (trend, $P(\beta < 0) = 0.89$).

Table 3: Selected results from the models for all scales (TYPE*FORMALITY interactions).

Scale	Contrast	$\widehat{\beta}$	$P(\beta > 0)$	$P(\beta < 0)$	95 % CrI
Geo. origin	F versus FINV	−0.31		0.89	[−0.81, 0.19]
Education	F versus FINV	−0.65		0.99	[−1.15, −0.16]
	IS versus FINV	−0.60		0.98	[−1.15, −0.06]
Hobbies	F versus FINV	−0.39		0.94	[−0.89, 0.11]
	IS versus FINV	−0.62		0.98	[−1.21, −0.03]

- Both F and IS sentences are associated with even lower ratings on the "education" scale in *formal* contexts than in *informal* ones (respectively, $P(\beta < 0) = 0.99$ and $P(\beta < 0) = 0.98$).
- Both F and IS sentences are associated with even lower ratings on the "hobbies scales" in *formal* contexts than in *informal* ones (respectively, $P(\beta < 0) = 0.94$ and $P(\beta < 0) = 0.98$).

For the purpose of this experiment, adults were artificially considered as a homogeneous group to which we could compare children, and the age variable was not controlled for and is thus quite unbalanced ($N_{+30\text{Y.O.}} = 10$). Some results, however, are statistically robust and are reminiscent of corpus findings from Thiberge et al. (2021), with people from different age groups exhibiting different linguistic behavior in their productions of interrogative variants. For instance (see Table 4), when compared to the FINV sentences:

- Both F and IS sentences received higher ratings on the "richness" scale from participants older than 30 (respectively, $P(\beta > 0) = 0.96$ and $P(\beta > 0) = 0.99$).
- Both F and IS sentences received higher ratings on the "geographical origin" scale (respectively, $P(\beta > 0) = 0.98$ and $P(\beta > 0) = 0.96$).
- Both F and IS sentences received higher ratings on the "education" scale (respectively, $P(\beta > 0) = 0.98$ and $P(\beta > 0) = 0.99$).

Further still, some statistically robust 3-way interactions between the TYPE*FORMALITY*AGE variables indicate that, on a par with corpus findings, different age groups may assess the use of some variants differently in different contexts, possibly because of different internalized sociolinguistic norms. For instance, the model run for the "education" scale provides robust evidence for these interactions for both F and IS sentences when compared to FINV (respectively, $\widehat{\beta} = -2.11$, $P(\beta < 0) = 0.99$, 95 % CrI = [–3.60, –0.66] and $\widehat{\beta} = -1.64$, $P(\beta < 0) = 0.98$, 95 % CrI = [–3.23,–0.05]). These interactions underline that, in our experiment, participants older than 30 reliably rated F- and IS- speakers as much less educated than FINV speakers in the formal contexts as opposed to informal contexts, in contrast to participants younger than 30. In fact, participants younger than 30 do not exhibit much variation on this scale from one context to the other, while participants older than 30 find that using fronting with inversion in informal contexts makes the speaker sound *less* educated than when they use simple fronting or in situ (and the reverse in formal contexts). While the focus of this specific task was not to allow for fine-grained distinction between adult age groups, a more controlled approach to the age variable in further works would allow for stronger generalizations as to how, even in the adult population, the associations between forms and meanings are still evolving over time.

4 The progressive build-up of social meaning: replicating the matched-guise task with children

The results from the adult experiment show a clear-cut distinction between at least the fronting with inversion variant and the other two with regard to the social meanings they activate in listeners. To assess if, when and how this adult pattern develops in children, we devised an adapted matched-guise task (MGT-C) which we ran with 3- to 11- year-old children.

Table 4: Selected results from the models for all scales (TYPE*AGE interactions).

Scale	Contrast	$\widehat{\beta}$	$P(\beta > 0)$	$P(\beta < 0)$	95 % CrI
Richness	F versus FINV	0.91	0.96		[–0.11, 1.97]
	IS versus FINV	1.59	1		[0.47, 2.72]
Geo. origin	F versus FINV	0.80	0.98		[0.01, 1.64]
	IS versus FINV	0.68	0.96		[–0.07, 1.47]
Education	F versus FINV	1.07	0.98		[0.02, 2.15]
	IS versus FINV	1.13	0.99		[0.23, 2.05]

4.1 Design and materials

Since it was important that even very young children were able to complete the task, we made sure the child's task was simpler, shorter and more accessible than the adult one. Overall the paradigm was still the same (i.e. still a matched-guise task programmed on the IbexFarm platform), in which child participants had to listen to speech productions and then associate them to a "scale" embodying some social characteristics. However:

- The number of items was reduced. From 4 practices + 30 fillers + 30 target items, we selected 4 practices + no filler + 18 target items from the adult experiment which seemed the easiest for children to understand (see complete list in the Supplementary Materials).
- The number of conditions was reduced, and thus item length was reduced. We removed the formality condition and kept only the core aspect of the experiment: the TYPE condition (with the same three levels: *fronting with inversion*, *fronting*, *in situ*).
- The number of response scales per item was reduced. We focused on 3 dimensions we thought would work best with children: *richness*, *education* and *hobbies.*
- Since we wanted the task to be enjoyable and fast-paced enough to maintain a child's attention throughout, we replaced the 7-point scales by binary choices (between property pairs which were the drawings used as scale limits in the adult experiment: poor/rich, cashier/doctor, sports/books). Figure 5 illustrates how it looked (the texts read: "The voice is rather the voice of…").

Overall, the experiment was a simple 1 × 3 design, with children hearing either a F/FINV/IS version of an interrogative sentence (***Où*** *tu vas ?* / ***Où*** *vas-tu ?* / *Tu vas* ***où*** *?* – "**Where** are you going?"). Then they had to choose between two characters, picking the one who they believed fit better with what they heard. Since we wanted a simple task with simple items, only one property pair was shown at a time. For this reason, we built 3 blocks in each experimental list, with each block containing a different variant of each target sentence, which was associated to a different property pair. Each list was thus 18 × 3 = 54 target items, preceded by the 4 training items. To counterbalance everything (all interrogative variants being presented with all three social dimensions for all items, and with at least 2 different voices – 1 male and 1 female), we created 6 lists (see Supplementary Materials). Item presentation inside a block was randomized.

4.2 Variables, analyses, and predictions

Our dependent variable was thus the choice for either the picture on the left (poor/cashier/sports) or the picture on the right (rich/doctor/books), which we recoded 0/1 for the analyses. Accordingly, we ran Bayesian logistic

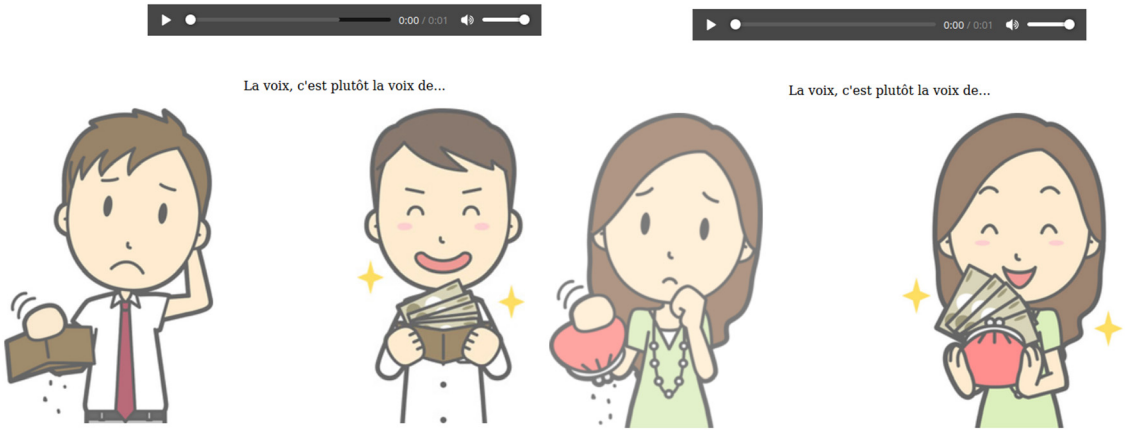

Figure 5: Example items from the child experiment ("richness" scale).

regression using the *brms* package. Bayesian models were particularly adapted here, given the high variability we could expect from an experiment with children. We ran one model for each dimension, and we will report results for each separately. Contrary to the adult experiment, given the exploratory nature of this research and the noisiness of the data we collected, we will report probabilities higher than 0.85 as "robust evidence" for the existence of an effect, and probabilities where $0.70 < P < 0.85$ as indicative of a trend.

Our main independent variable was the syntactic TYPE of the interrogative (reference level: *fronting with inversion*). Since we are interested in the evolution of the associations between interrogative type and social properties over time, we assigned children to different GROUPS that roughly encapsulate both age and actual class group (as given by the school were the experiment took place – see below for details). This variable was recoded numerically, from 1 to 4. We also included random intercepts for participants and items, with random slopes (see Supplementary Materials for full specifications).

Based on current knowledge of how social perception develops in children, we could anticipate different things:

(1) Children seem to be able to make "metapragmatic comments" on sociolinguistic variation, that somewhat resemble adult judgments, at around age 9–11 (Buson 2009). We thus expected children from the corresponding age categories to exhibit a similar pattern as the one exhibited by adults on the scales we selected.
(2) More precisely, we expect a progressive build up of this pattern, with the youngest children making different associations – or making no apparent association – between linguistic forms and social meanings, with the oldest children making associations in an adult-like way. In this view, in-between children would exhibit a different pattern from the youngest group but still not as solidified as their older peers.
(3) Given that children progressively build their ability to explicitly establish how they feel about a person depending on how they talk (Vaughn and Becker 2024) but also that very young children exhibit production patterns similar to that of adults for at least some sociolinguistic variables (Roberts and Labov 1995), we also anticipate potentially different trajectories for the associations between language and social properties, with some associations established very early on while some might take longer to appear.

4.3 Participants and procedure

Participants were 136 children from a mostly monolingual community from a small town in Western France. The study was prepared long beforehand, with a convention signed between the experimenter's university and the school, an ethics committee approval (Uni. Paris Cité, IRB number 00012019-19), and signed consent from all the children's parents. Children who did not want to take part on the days the experiment took place were free to do so (but nearly all the children at the school participated).

The task was presented as a mini-game the children had to play during teaching hours. For the youngest children (age 3–5), the experimenter, who had previously been introduced to the children, played the game together with them on a tablet computer, under supervision by a trusted adult (school staff) in a quiet room. For children aged 6 and above, groups of 5–6 children played the game independently at the same time, on school computers in the same small quiet room with a trusted adult, after a quick introduction of the task by the experimenter. The experimenter then went from child to child to ensure everything was working fine and to answer questions. During the game, children were told they had to help researchers who had mixed audio files, by assigning them to people they thought had uttered them. At the end of each experimental block, a small screen showed that progress had been made and that the scientists were becoming more and more relieved. At the end of the game, a small diploma was given to all the children to thank them, even to those who did not participate. The task took a few minutes for the oldest children, and session times varied greatly for younger children, up to 20 minutes.

To ensure maximum protection of the children, no personal information except their age and their "class-group" was linked to the data. Tables 5 and 6 give an overview of these dimensions (mean age: 7;06, median age: 7;00). Given that the school had multiple overlaps between age and class groups, for the analyses we merged class groups higher than CP in two "supergroups" based on input from school teachers relating to the levels of pupils:

Table 5: Children's class group (MGT-C).

GS	CP	CE1	CE2	CM1	CM2
18	23	29	20	17	29

Table 6: Children's age (MGT-C).

Age 3	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9	Age 10	Age 11
1	3	14	23	28	20	24	22	1

Table 7: Children's final grouping for analyses (MGT-C).

Group-class	GS	CP	CE1-CM1	CM1-CM2
Total children	18	19	59	40
Total children according to official level	GS(18)	CP(19)	CP(4) CE1(29) CE2(20) CM1(6)	CM1(11) CM2(29)
Answers kept	677	741	2,282	1,575

the first group (*CE1-CM1*) mixes children from CE1 and CE2 classes but also a few "advanced" CP children put in a majority-CE1 class and a few CM1 children who were in a majority-CE2 class, the second group regroups the vast majority of CM1s and CM2s (see Table 7 for the detailed grouping).

Out of the $54 \times 136 = 7{,}344$ possible answers, we discarded 1,761 because the experimenter noticed that, when, because of randomization parameters, an item was followed by another one with the same answer scale (i.e. two different sentences but the same pictures to assign to them), children often just re-clicked on the same picture they had just chosen, thinking it was a bug. For similar reasons, we also removed answers when children took less than 2 s to answer ($N = 308$). In total, we kept 5,275 answers (1,786 on the *richness* scale, 1,743 on the *education* scale, 1,746 on the *hobbies* scale). Despite this conservative approach to the data, Bayesian modeling still allows for good generalization over these results.

4.4 Main results

We first present results for each scale individually, before moving to the global picture.

4.4.1 Results for the 'richness' scale

Figure 6 provides a visualisation for the proportions of choices from the children for the 'Rich' picture, depending on the sentence TYPE they heard (color lines) and the class/age GROUP they belong to (x-axis).

A model run with these variables (see full specifications in the Supplementary Materials) yields robust evidence for an overall effect of type for the IS versus FINV contrast ($\widehat{\beta} = 0.58$, $P(\beta > 0) = 0.92$, 95 % CrI = [–0.21, 1.41]), meaning that, independent of the children's age, IS sentences are consistently more associated with the rich person than FINV sentences. The model also yields tentative evidence (trend) for a distinction between F and FINV

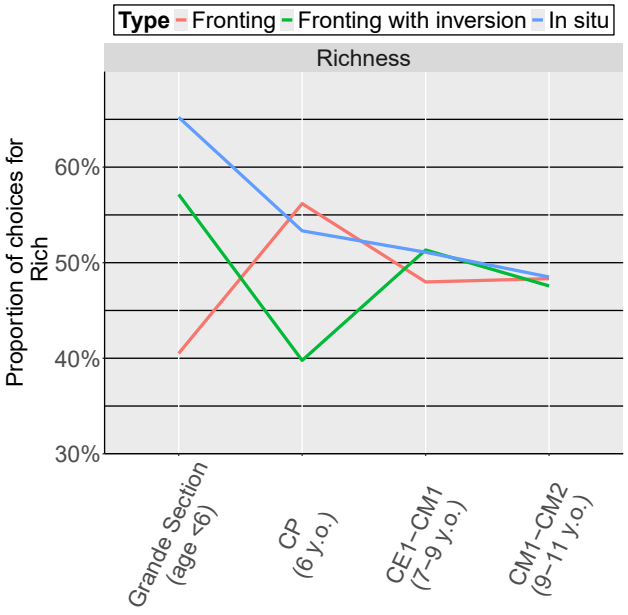

Figure 6: Results on the richness scale depending on age/class group (MGT-C).

sentences that would go in the opposite direction (F sentences less associated with the rich person overall, $\widehat{\beta}$ = –0.25, $P(\beta < 0)$ = 0.73, 95 % CrI = [–1.05, 0.55]).

These associations are not constant over time however, and the model also yields evidence for TYPE*GROUP interactions for both the IS versus FINV (robust, $\widehat{\beta}$ = –0.15, $P(\beta < 0)$ = 0.87, 95 % CrI = [–0.42, 0.11]) and F versus FINV (trend, $\widehat{\beta}$ = 0.08, $P(\beta > 0)$ = 0.73, 95 % CrI = [–0.18, 0.34]) contrasts. Both correspond to the difference between FINV sentences and the other two types reducing over time (IS being less associated to richness and F slightly more as children age).

4.4.2 Results for the 'education' scale

Figure 7 provides a visualisation for the proportion of choices from the children for the 'Educated' picture (doctor, as opposed to the stereotypically less educated cashier), depending on the TYPE and GROUP variables.

Here, the overall picture appears to be different from the previous one. The model run for this scale yields evidence for an effect of TYPE only for the IS sentences compared to the FINV ones ($\widehat{\beta}$ = –0.48, $P(\beta < 0)$ = 0.88, 95 % CrI = [–1.30, 0.33]). This is evidence for a constant distinction over time between the two types, and the absence of interaction with the GROUP variable further highlights that this difference stays consistent from the early stages on. There is no conclusive evidence for a difference between F and FINV sentences, either with or without taking age/class into account.

4.4.3 Results for the 'hobbies' scale

Figure 8 gives the results for the proportions of choices from the children for the 'books' picture, depending on sentence TYPE and children GROUP.

Yet again, the overall picture is different from the other two, with the model run for this scale yielding some evidence (trend, $\widehat{\beta}$ = –0.10, $P(\beta < 0)$ = 0.78, 95 % CrI = [–0.37, 0.17]) for a TYPE*GROUP interaction for the IS versus FINV contrast only. There does not seem to be an overall difference between the three sentence types, but when looking at it more precisely, FINV sentences do in fact become more associated with books and reading than IS sentences when the children become older.

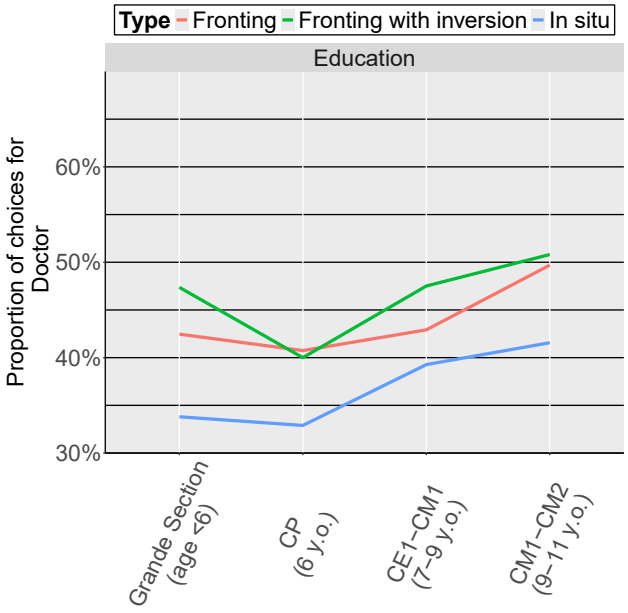

Figure 7: Results on the education scale depending on age/class group (MGT-C).

4.4.4 Overall picture and discussion

Figure 9 compiles the results for all scales in the child experiment (on the left), with a screenshot of all associations between linguistic forms (color lines) and social stereotypes (*x*-axis) over all 4 developmental stages (1 cell per group), and compares it to the same three scales from the experiment with adults (on the right, but it should be noted that the dependent variable and scale is not the same). This allows for a more intuitive comparison of the network of associations between forms and social meanings in children, at four different stages of development, to the same network in adult speakers.

A first observation is that the clear-cut pattern exhibited by adults, with in situ and fronting with inversion being really differentiated, is seen in older children but not in younger ones. This supports our first predictions (1 and 2), with a progressive build-up of the pattern seen in adults as children get older. It should be noted however

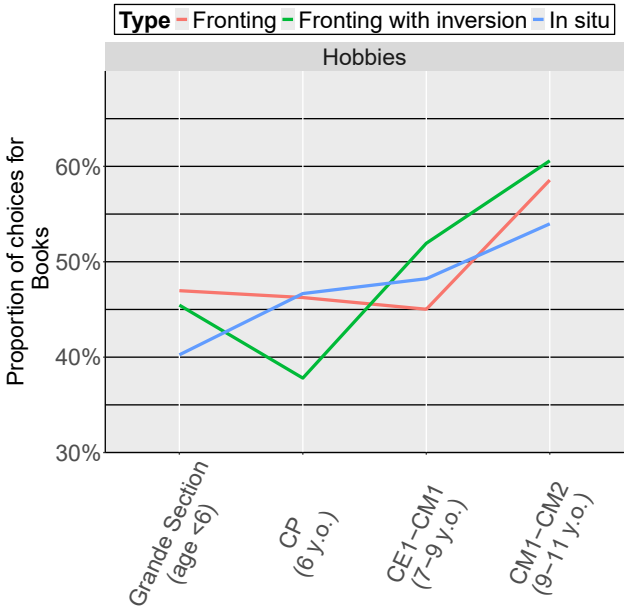

Figure 8: Results on the hobbies scale depending on age/class group (MGT-C).

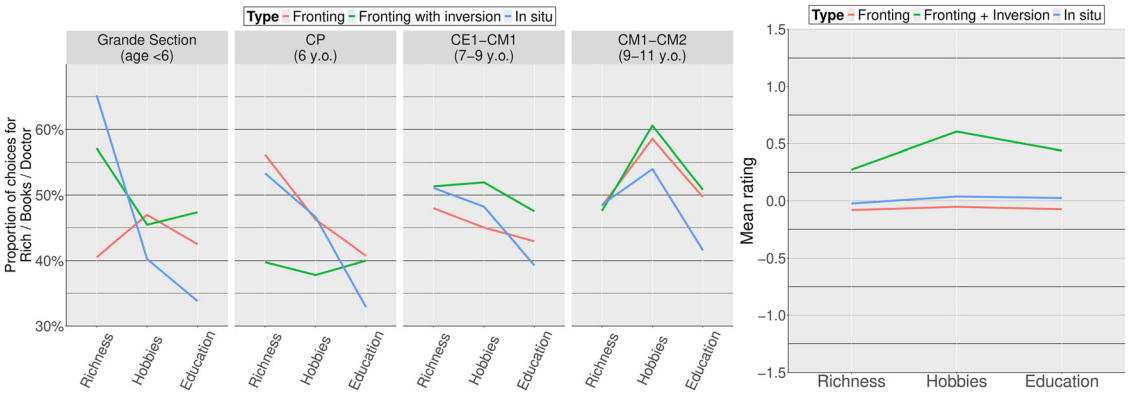

Figure 9: Overall results from the child matched-guise (left) compared to the adult experiment (right).

that, although older children do indeed make a distinction between *fronting with inversion* and *in situ*, and if the associations between these forms and social meanings go in the same directions as in adults (i.e. *fronting with inversion* is more associated with educated persons, doctors, than with cashiers, and more associated with books than with sports, compared to *in situ*), the *simple fronting* case is different. In older children (ages 9–11) it behaves very similarly to *fronting with inversion*, while in younger ones the two variants are substantially different. One explanation might be that both variants are actually not very prevalent in spontaneous adult French, thus making them both marked enough from a sociolinguistic perspective that children perceive them as 'apart' from the *in situ* variant, which is much more frequent (more than 60 % uses on adult corpus data). It is possible that the adult-like pattern solidifies after the age of 12, when children are confronted with more situations which allow them to better distinguish between the situational uses of both marked variants.

A second observation is that all three associations between form and social meaning, i.e. the three scales (richness, hobbies, education), behave in very different ways from a developmental perspective. On the *richness* scale, whatever differences between variants are seen in younger children disappear over time; on the *hobbies* scale the difference observed in adult data (fronting with inversion more associated with books) only appears in older children; and on the *education* scale the degree of association between fronting with inversion and the doctor/educated person (similar to that of the fronting variant, but greater than for the in situ variant) appears very early on and remains constant over time. This supports our third prediction regarding how different form-social meaning associations might follow different developmental paths because of how children may view the world very differently at different stages of their upbringing. Nearly all children go to the doctor from their first years on, which, according to a usage-oriented approach *à la* Buson (2009) and Buson and Billiez (2013), they can make use of very early on to building up a repertoire of many exemplars where stereotypical interactions with the physician involve interrogatives with fronting (and maybe a verb-subject inversion) such as *Comment vas-tu ?* or *Où as-tu mal ?* ('How are you? Where does it hurt?'). For the *hobbies* scales, children learn how to read over the years, and it is through this progressive access to literature that they can access more complex books where they are exposed to sentences with fronting with inversion variants. For the *richness* scale, children get more knowledge of how the world operates as years go by and, from very limited interview data (post-session discussions with the children where the experimenter asked them what kind of person they had in mind when asked about 'rich individuals' they knew), we observed that young children conceive *richness* as mostly related to fame (with answers along the lines of 'that football player/that famous singer I saw on TV the other day'). They do not really relate it to social prestige as an adult internalizes it, hence the disconnect between the two scales in children, which was not seen in adults. Furthermore, for the most part, children did not treat the *education* scale as a measure of what adults would consider education, and the link between years of education and professions like doctor versus cashier is not apparent to them. So the results on that scale should probably be thought of in terms of what situations the two drawings (cashier vs. doctor) embodied: going shopping and talking to someone behind the register on one hand, and going to the doctor on the other hand.

5 Conclusions

Taken together, the results from the two experiments we presented are evidence for a progressive build up of the sociolinguistic evaluative competence in children acquiring their L1. The first experiment brought new elements allowing for a better understanding of the social meaning of WH- interrogative in adult French. In complement to existing corpus data, our results provide new insights as to why different variants may be used differently by different groups of people in different contexts. The social properties that are projected upon a person using one or the other variant may be different from group to group (i.e. age categories), and may slightly differ from context to context (i.e. as a function of formality). All three variants seem to convey quite different social *personae* with *fronting with inversion*, the standard form promoted by grammars for French (see Riegel et al. (2014), for instance), being more associated with higher education, richness, or book-oriented hobbies than *in situ* or *simple fronting* variants. The second experiment focused on child acquisition of these associations on three dimensions ('education', hobbies, richness). The results show that these dimensions of the social meaning of the WH-interrogative system appear progressively in children, but that they all follow a different developmental trajectory.

We take this as further evidence that a more thorough and systematic investigation of what the social world of children look like is needed, and we argue that adapted matched-guise tasks are useful tools for this project, that could be administered, for example, across different schools from different regions with children from the same linguistic community. Furthermore, these tasks can – and should – target many different social properties, just as in adult experiments, and not be limited to in-group similarity measures such as kindness or friendliness. By expanding over these dimensions and by multiplying observations over a large number of children from roughly comparable linguistic backgrounds at the same time, researchers in the acquisition field will be better equipped to integrate the sociolinguistic aspects of language into models of language acquisition. This highlights the need for a multifactorial approach to language acquisition overall, and particularly to the acquisition of syntactic alternation phenomena. Looking only at the intrinsic complexity of competing linguistic forms is very important, but it might not be enough to account for the mechanisms through which children learn how and when to use them. Alternating forms in adult target grammars are often – if not always – carrying subtly different sociolinguistic meanings. Children learn and acquire these subtleties at the same time they are acquiring the underlying syntactic structure of these variants.

Acknowledgments: We thank the children, parents, teachers and in particular the director of the school involved in the experiments for their availability and investment in the experiments.
Informed consent: Informed consent has been obtained from all individuals included in this study.
Ethical approval: Since the main experimental paradigm involved children, it was crucial that we followed the best practices to carry out this research. As such, we submitted our experimental plan to a university ethics committee for consideration and acceptation (Uni. Paris Cité, IRB number: 00012019 – 19).
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.
Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission. GT conducted the experiments described in this article as well as the statistical analyses, and redacted the first draft of this paper. HB and BH advised on the building of the experimental paradigms as well as the statistical analyses, and proposed revisions of the first draft. All authors contributed to the revisions and agreed upon the final version of the manuscript.
Conflict of Interest: Authors state no conflict of interest.
Data Availability Statement: The datasets generated during and/or analysed during the current study are available in the OSF repository, https://osf.io/7r6kv/ (DOI: https://doi.org/10.17605/OSF.IO/7R6KV).

References

Abeillé, Anne. 2007. *Les grammaires d'unification*. Paris: Hermès.

Adli, Aria. 2006. French wh-in-situ questions and syntactic optionality: Evidence from three data types. *Zeitschrift für Sprachwissenschaft* 25(2). 163–203.

Ambridge, Ben. 2010. Children's judgments of regular and irregular novel past-tense forms: New data on the English past-tense debate. *Developmental Psychology* 46(6). 1497–1504.

AudacityTeam. 1999–2021. *Audacity (R): Free audio editor and recorder*. Version 2.4.2. https://audacityteam.org/ (accessed 20 September 2020).

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.

Beltrama, Andrea. 2020. Social meaning in semantics and pragmatics. *Language and Linguistics Compass* 14(9). e12398.

Beyssade, Claire. 2006. La structure de l'information dans les questions: quelques remarques sur la diversité des formes interrogatives en français. *Linx. Revue des linguistes de l'université Paris X Nanterre* 55. 173–193.

Brehm, Laura & Philipp M. Alday. 2022. Contrast coding choices in a decade of mixed models. *Journal of Memory and Language* 125. 104334.

Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80. 1–28.

Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the R package brms. arXiv preprint arXiv:1705.11123. https://doi.org/10.48550/arXiv.1705.11123.

Buson, Laurence. 2009. Variation stylistique entre 5 et 11 ans et réseaux de socialisation scolaire: usages, représentations, acquisition et prise en compte éducative. Doctoral Dissertation. Université Stendhal-Grenoble III.

Buson, Laurence & Jacqueline Billiez. 2013. Representations of stylistic variation in 9-to 11-year-olds: Cognitive processes and salience. *Linguistics* 51(2). 325–354.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82. 711–733.

Campbell-Kibler, Kathryn. 2006. Variation and the listener: The contextual meanings of (ING). *Penn Working Papers in Linguistics: Papers from NWAV* 34. 53–64.

Campbell-Kibler, Kathryn. 2007. Accent, (ING), and the social logic of listener perceptions. *American Speech* 82(1). 32–64.

Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Goodrich Ben, Michael Betancourt, Allen Riddell. 2017. Stan: A probabilistic programming language. *Journal of Statistical Software* 76. 1.

Chang, Lisa. 1997. Wh-in-situ phenomena in French. Doctoral Dissertation. University of British Columbia.

Cheng, Lisa L.-S. & Johan Rooryck. 2000. Licensing wh-in-situ. *Syntax* 3(1). 1–19.

Chevrot, Jean-Pierre. 2024. Acquiring sociolinguistic variation in a first language: Toward a socialization-based framework. In *Routledge handbook of variationist sociolinguistics* (to appear). https://hal.science/hal-04377247v1

Chevrot, Jean-Pierre & Paul Foulkes. 2013. Introduction: Language acquisition and sociolinguistic variation. *Linguistics* 51(2). 251–254.

Chevrot, Jean-Pierre, Aurélie Nardy & Stéphanie Barbu. 2011. Developmental dynamics of SES-related differences in children's production of obligatory and variable phonological alternations. *Language Sciences* 33(1). 180–191.

Coveney, Aidan. 2011. L'interrogation directe. *Travaux de Linguistique* 2. 112–145.

Day, Richard R. 1980. The development of linguistic attitudes and preferences. *TESOL Quarterly* 14. 27–37.

Dollinger, Stefan, Vanessa Chan, Kate Pasula & Anthony Maag. 2024. How linguistically tolerant or insecure are school-aged children? A matched-guise, gamified approach for 6- to 12-year-olds in Canada. In Mirjam Schmalz, Manuela Vida-Mannl, Sarah Buschfeld & Thorsten Brato (eds.), *Acquisition and variation in world Englishes: Bridging paradigms and rethinking approaches*, 307–334. Berlin, Boston: De Gruyter Mouton.

Drummond, Alex. 2016. *IbexFarm* (version 0.3.9). https://adrummond.net/ibexfarm

Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12(4). 453–476.

Eckert, Penelope. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41(1). 87–100.

Eckert, Penelope & William Labov. 2017. Phonetics, phonology and social meaning. *Journal of Sociolinguistics* 21(4). 467–496.

Faure, Richard & Katerina Palasis. 2021. Exclusivity! Wh-fronting is not optional wh-movement in Colloquial French. *Natural Language & Linguistic Theory* 39. 57–95.

Guryev, Alexander & François Delafontaine. 2015. La variabilité formelle des questions dans les écrits SMS. *Travaux Neuchâtelois de Linguistique* 63. 129–152.

Hamlaoui, Fatima. 2011. On the role of phonology and discourse in Francilian French wh-questions. *Journal of Linguistics* 47(1). 129–162.

Hulk, Aafke C. J. 1996. The syntax of wh-questions in child French. In *Amsterdam series in child language development*, 129–172. Available at: https://hdl.handle.net/11245/1.120759.

Jakubowicz, Celia. 2011. Measuring derivational complexity: New evidence from typically developing and SLI learners of L1 French. *Lingua* 121(3). 339–351.

Kerswill, Paul. 1996. Children, adolescents, and language change. *Language Variation & Change* 8(2). 177–202.

Kinzler, Katherine D. & Jasmine M. DeJesus. 2013. Children's sociolinguistic evaluations of nice foreigners and mean Americans. *Developmental Psychology* 49(4). 655–664.

Kircher, Ruth. 2015. The matched-guise technique. In Hua Zhu (ed.), *Research methods in intercultural communication: A practical guide*, 196–211. Hoboken: Wiley-Blackwell.

Labov, William. 1964. Stages in the acquisition of standard English. In R. Shuy, A. Davis & R. Hogan (eds.), *Social dialects and language learning*, 77–104. Champaign, IL: National Council of Teachers of English.

Labov, William. 1966. *The social stratification of English in New York City*. Philadelphia: University of Pennsylvania Press.

Lambert, Wallace E., Richard C. Hodgson, Robert C. Gardner & Samuel Fillenbaum. 1960. Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology* 60(1). 44–51.

Larrivée, Pierre. 2019. Historical pragmatics, explicit activation and wh- in situ in French. In *Romance languages and linguistic theory 15: Selected papers from 'Going Romance' 30*, 113–132. Frankfurt.

Levon, Erez. 2007. Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society* 36(4). 533–554.

Mathieu, Éric. 2009. Les questions en français: micro-et macro-variation. *Le Français d'Ici: Études Linguistiques et Sociolinguistiques de la Variation*, 37–66.

Nardy, Aurélie, Jean-Pierre Chevrot & Stéphanie Barbu. 2013. The acquisition of sociolinguistic variation: Looking back and thinking ahead. *Linguistics* 51(2). 255–284.

Ochs, Elinor. 1992. Indexing gender. In A. Duranti & C. Goodwin (eds.), *Rethinking context: Language as an interactive phenomenon*, 335–358. New York: Cambridge University Press.

Palasis, Katerina. 2013. The case for diglossia: Describing the emergence of two grammars in the early acquisition of metropolitan French. *Journal of French Language Studies* 23(1). 17–35.

Podesva, Robert J. 2011. Salience and the social meaning of declarative contours: Three case studies of gay professionals. *Journal of English Linguistics* 39(3). 233–264.

Posit Team. 2023. *RStudio: Integrated development environment for R*. Boston, MA: Posit Software, PBC. Available at: https://www.posit.co/.

Prévost, Philippe. 2009. *The acquisition of French: The development of inflectional morphology and syntax in L1 acquisition, bilingualism, and L2 acquisition*. Amsterdam, The Netherlands: John Benjamins Publishing.

Quillard, Virginie. 2001. La diversité des formes interrogatives: comment l'interpréter? *Langage & Société* 1. 57–72.

R Core Team. 2023. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Riegel, Martin, Jean-Christophe Pellat & René Rioul. 2014. *Grammaire Méthodique du Français*. Paris: Presses Universitaires de France.

Roberts, Julie. 1994. Acquisition of variable rules: (-t,d) deletion and (ing) production in preschool children. Doctoral Dissertation. University of Pennsylvania.

Roberts, Julie & William Labov. 1995. Learning to talk Philadelphian: Acquisition of short a by preschool children. *Language Variation & Change* 7(1). 101–112.

Rosenthal, Marilyn S. 1974. The magic boxes: Preschool children's attitudes toward black and standard English. *The Florida FL Reporter*, no. 12: 55–62, 92, 93.

Ross, John R. 1967. Constraints on variables in syntax. Doctoral Dissertation. Massachusetts Institute of Technology.

Sag, Ivan A. 2010. English filler-gap constructions. *Language* 86. 486–545.

Scheidnes, Maureen & Laurice Tuller. 2010. Syntactic movement in the production of French wh-questions: The role of computational complexity versus L1 transfer in adult L2 acquisition. in Vicenç Torrens, Linda Escobar, Anna Gavarró & Juncal Gutiérrez (eds.), *Movement and clitics: Adult and child grammar*, 185–217. Cambridge: Cambridge Scholars Publishing.

Smith, Jennifer & Mercedes Durham. 2019. *Sociolinguistic variation in children's language: Acquiring community norms*. Cambridge: Cambridge University Press.

Sorensen, Tanner, Sven Hohenstein & Shravan Vasishth. 2016. Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology* 12(3). 175–200.

Tailleur, Sandrine. 2013. The French wh interrogative system: Est-ce que, clefting? Doctoral Dissertation. University of Toronto. Available at: https://hdl.handle.net/1807/36013.

Thiberge, Gabriel, Flora Badin & Loïc Liégeois. 2021. French partial interrogatives: A microdiachronic corpus study of variation and new perspectives in a refined pragmatics framework. *Faits de Langues* 51(2). 179–202.

Vaughn, Charlotte & Kara Becker. 2024. Documenting the emerging social-semiotic landscape in children ages 5 to 12. *Language & Communication* 95. 16–30.

Zuckerman, Shalom & Aafke Hulk. 2001. Acquiring optionality in French wh-questions: An experimental study. *Revue Quebecoise de Linguistique* 30(2). 71–97.