‘Nous fêterons’ or ‘On va fêter’? Mimicking Age-Sensitive Variation with ChatGPT

Valerie Hekkel, Friederike Schulz (University of Potsdam)*, Marta Lupica Spagnolo* (University of Potsdam)

vhekkel(at)posteo.de, frieschulz(at)uni-potsdam.de, lupica(at)uni-potsdam.de

Abstract

This study explores ChatGPT’s capability to mimic age-sensitive linguistic variation in contemporary French, particularly focusing on older adult speech. Our investigation aimed to assess whether ChatGPT could (1) align its naive responses with age-related language use, (2) demonstrate explicit knowledge of age-related linguistic variation, and (3) modify responses based on such knowledge. Using contexts from the LangAge corpus, ChatGPT was prompted to answer questions from the perspective of speakers of different ages (30– 90) in different interview years (1980–2020), with a specific focus on the use of first-person plural subject clitics (*nous*/*on*) and future tenses (*futur simple/proche*). The results revealed that ChatGPT’s responses predominantly favored formal linguistic variants across all ages. While expert-knowledge injection significantly increased the usage of formal variants, there was no systematic influence of age, birth year, or interview year on variant selection. A partial exception is represented by speakers aged 70 for whom ChatGPT displayed heightened linguistic uncertainty in the naive answer. By contrast, the variant distribution in (3) is mainly motivated by ChatGPT’s expert knowledge generated in (2). These findings highlight the potential and limitations of current LLMs in capturing age-specific variation while encouraging further integration of sociolinguistic methods into LLM research.

Keywords

sociolinguistics, AI, LLM, GPT4, age, generation, Apparent-Time, age-grading, variation, French, first-person plural, clitics, future tenses

1 Introduction[1]

In the rapidly evolving field of Large Language Models (LLMs), the integration of sociolinguistic perspectives remains a largely untapped area. This paper aims to bridge this gap by exploring the capabilities of ChatGPT in mimicking age-sensitive linguistic variants, focusing specifically on the use of first-person plural subject clitics and future tense variants in French. We investigate the variation between the usage of *nous* and *on* (first-person plural clitics), as well as *futur simple* and *futur proche* (inflected future and periphrastic future), as these elements have been found to be age-sensitive indicators.

[1] The paper has been discussed and approved by all authors. The writing has occurred as follows: Valerie Hekkel is responsible for Sections 1, 2, 2.1, 3, 3.3, as well as the Python coding and experiment design; Marta Lupica Spagnolo is responsible for Sections 3.4, 4.1, 4.3, and the statistical analysis. Section 4.2 was jointly written by Valerie Hekkel and Marta Lupica Spagnolo, and sections 2.2.1, 2.2.2, 3.1, 3.2 and 5 were jointly written by Friederike Schulz and Valerie Hekkel. Section 6 emerged from a collaborative writing by all three authors.

AILing
AI-Linguistica.
Linguistic Studies on AI-Generated Texts and Discourses

Hekkel, Valerie & Schulz, Friederike & Lupica Spagnolo, Marta
Nous fêterons’ or ‘On va fêter’?
Mimicking Age-Sensitive Variation with ChatGPT
AI-Linguistica 2024. Vol. 1 No. 1
DOI: 10.62408/ai-ling.v1i1.11
ISSN: 2943-0070

The motivation for this study arises from the observation that while LLMs have made significant strides in various applications, their proficiency in reflecting the linguistic nuances of different age groups, especially older adults, has not been thoroughly investigated. Existing research on LLMs in combination with sociolinguistics has primarily focused on areas such as privacy violations, where linguistic clues are analyzed for insights into a speaker's metadata (Staab, Vero, Balunović and Vechev 2023), and the biases inherent in these models (Markl 2022; Ostapenko, Wintner, Fricke and Tsvetkov 2022). Additionally, some studies have utilized sociolinguistics for polling latent opinions using earlier versions of language models (Feldman, Dant, Foulds and Pan 2022). A study by Salewski, Alaniz, Rio-Torto et al. (2023) analyzed the impact of the personas LLMs are prompted to impersonate on their performance in different tasks. *Age* (2-60) was among the variables that were used to define the personas. While this study explores the effect of social variables on an LLM output, it does not define linguistic variation as target variable. Altogether, these approaches do not fully exploit the interplay between current LLMs, such as ChatGPT, and sociolinguistics.

Our research seeks to establish a new entry point into the study of sociolinguistics and LLMs by examining how well ChatGPT can adapt its output to age-specific language use, particularly focusing on the nuances of first-person plural clitics and French future tenses, as their more formal variants (*nous* and *futur simple*) have been shown to be more frequent for older adults (Coveney 2000; Roberts 2012; Abouda and Skrovec 2015; Sankoff and Wagner 2020). The central question of our investigation is: "Can ChatGPT effectively mimic age-sensitive linguistic variants, particularly those associated with the language of older adults?". Through this research, we aim to shed light on the model's linguistic versatility and its potential to represent the diversity of human language across different demographic segments.

In Section 2, we provide an overview over sociolinguistic theories concerning the relation between age/ birth year and linguistic variation (2.1) variation patterns observed for the two linguistic phenomena under scrutiny (2.2). The methodological framework of our study is delineated in Section 3, containing an outline of our input-data (3.1), a detailed account of our experiment design (3.2), the annotation guidelines (3.3), and the formulation of our research questions (3.4). The results in Section 4 comprises a comparison of the distribution of linguistic variants before and after expert prompting (4.1), an analysis of their interplay with the speakers' metadata, such as age, birth and interview year (4.2), and an exploration of the relative importance of those and other extralinguistic factors, such as 'expert-knowledge injection', on the linguistic outcomes (4.3). The paper culminates in a discussion (5), synthesizing findings and implications of the study, followed by a conclusive summary that highlights key insights and reflections (6).

2 Background

The underlying reasons for anticipating an age-sensitive and thus diastratic variation in the usage of the two analyzed linguistic variables (first-person plural subject clitics and future tenses in French) are primarily rooted in two other variational phenomena: diachronic variation and diaphasic variation.

Diachronic Variation: it refers to the changes in language use over time. This aspect of linguistic evolution is particularly relevant to our study as it encapsulates how the use of first-person plural clitics and future tenses has evolved in a time frame of about one century. Such linguistic changes are expected to be reflected in the age-related variation of these linguistic variables.

Diaphasic Variation: it refers to the variation related to communicative situations and their degree of formality. Given that both the first-person plural clitics and future tenses in French are also considered to be markers of formality (see Section 2.2 for further details), their use is expected to vary not only with time but also with the degree adherence to a written norm, which, as will be shown, is more prominent for persons of advanced age.

After a brief overview of the relationship between diachronic language variation and age (2.1), Section 2.2 explores the state-of-the-art regarding the interactions between age and the two analyzed linguistic variables.

2.1 Linguistic Variation Across Age

The expectation of age-sensitive variation in the linguistic variables under study is premised on various synchronic and diachronic variation patterns. These patterns detail the ways in which language use can vary over time within individual life spans and for linguistic communities, as well as how variants are distributed over different age groups at a given point in time.

One of these patterns is described by the “Apparent-Time Hypothesis”. The term was coined by Labov (1966 [2006]; 1978) to refer to the method of using “the present to explain the past” (Labov 1978). This approach uses the synchrony of linguistic variation to make inferences about its diachronic evolution. It claims that certain linguistic patterns, once established during the early years, tend to remain relatively stable throughout an individual’s life (see Sankoff 2005: 1003), so that a cross-sectional sample based on age can mirror language change.

Distinct from the stable usage patterns suggested by the Apparent-Time Hypothesis is the concept of “age-grading” (Hockett 1950; Labov 1963), which draws inferences about diachronic developments from synchronic linguistic variation. These variations may align with language change but are not necessarily indicative of it (Ashby 1991; Wagner 2012).

In addition to age-grading, other diachronic phenomena go along with language variation, including generational change, as outlined in the summary provided by Wagner (2012). These phenomena are captured in Table 1, which classifies various diachronic patterns based on the stability of linguistic variants at both individual and community levels.

Table 1: Classification of diachronic phenomena based on the stability of linguistic variants in individual and community usage over time, along with associated synchronic patterns. “Stable” marks to the absence of change, while “unstable” indicates a diachronic variation. An instability on the individual level refers to the variability throughout an individual’s life span. This would mean that individuals adopt certain linguistic variants as they age. An instability on the community level describes a linguistic change, which does not necessarily have to be accompanied by individual diachronic variation. While these two variation levels represent the longitudinal view, the synchronic pattern is a cross-sectional one, looking at whether variation occurs in dependence of age. Extracted from Wagner (2012: 373).

		Individual	**Community**	**Synchronic Pattern**
1	Stability	Stable	Stable	flat
2	Age-grading	Unstable	Stable	monotonic slope with age
3	Generational change	Stable	Unstable	monotonic slope with age
4	Communal change	Unstable	Unstable	flat
5	*Lifespan change*	*Unstable*	*Unstable*	*monotonic slope with age*

As will be described in Section 3, our study design allows for both, a longitudinal and cross-sectional analysis of the variation in ChatGPT-generated texts. This approach enables our analysis to disambiguate the different levels at which the described diachronic phenomena occur.

2.2 Age-Sensitive Linguistic Dynamics in French: Clitics and Future Tenses

In the study of language variation and change, certain linguistic features exhibit marked sensitivity to the age of speakers. This section presents the State of the Art regarding the variation of the two variables that we have chosen for our study: the use of first-person plural subject clitics and future tenses in French. These variables were chosen for their demonstrated variability in relation to speaker age, thus offering a fertile ground for exploring ChatGPT’s capability to mimic age-sensitive linguistic variation.

First-Person Plural Clitics: The first-person plural clitics (see 2.2.1), present an example of age-sensitive linguistic variation in French. Research indicates that the choice between the forms *nous* and *on* is not merely a matter of syntactic preference but also reflects sociolinguistic factors, including the age of the speaker. Furthermore, as to be shown, the variants (or their usage frequency) might be subject to an ongoing linguistic change.

Future Tenses: Similarly, the choice between different future tense constructions in French (the inflected future vs. the periphrastic future) is not only motivated by linguistic factors, e.g. grammatical person, temporal-aspectual properties, but also offers insights into age-related language dynamics (see 2.2.2).

These variations not only mirror language change but also align with age-related linguistic preferences. The tendency to use one form over the other varies across different age cohorts with *futur simple* being more commonly associated with older than younger adults (Sankoff and Wagner 2020).

2.2.1 First-Person Plural Clitics

The alternation between the clitic pronouns *nous* and *on*, both with the meaning of ‘we’, e.g. *nous fêtons* for ‘we celebrate-1pl.’ or *on fête* for ‘we celebrate-3sg.’ is a well-known morphosyntactic feature of present-day spoken French.[2] Their variation, which has been observed since the early 20th century, has been related to an entanglement of diaphasic and diachronic factors (see for example Söll 1974 [1980]: 137; Weinrich 1989).

Numerous studies (Coveney 2000; Gerstenberg 2011; Söll 1969) show a pronounced preference for *on* over *nous* in contemporary spoken language. However, a complete displacement, as assumed by Bally (1952: Chapter I, 8), has not yet taken place. Indeed, some scholars argue for a stability of these variants (for example Blanche-Benveniste 1997).

Coveney (2000) shows that *on* is clearly preferred in spoken language in corpus data from northern France. He attributes this preference to stylistic and pragmatic aspects. The use of *nous* is largely restricted to formal contexts or is used to emphasize or contrastively highlight the pronoun (Coveney 2000: 456). In Montreal French, the use of *nous* is also considered marginal, as it occurs mainly in formal contexts. Speakers using *nous* usually have a higher socio-economic status, tend to be older and have a high level of education (Laberge 1977: 141), which might be traced back to a greater need to speak in a norm-oriented way. As to the age variable, Laberge points out that, in her study, 40.5% of the speakers older than 50 use the clitic *nous*, while only 8.2% of the younger speakers do. She attributes this phenomenon to either *nous* being typical for the older age-group, or it being an older variant that is about to disappear, the latter being indicated to be more probable (Laberge 1977: 137).

Gerstenberg (2011: 237) suggests that the norm-orientation conveyed during school years also influences the avoidance of *on* in the LangAge corpus, which is composed of interviews carried out in the Orléans area in 2005.[3] The participants’ data show a *nous*-*on* ratio of 0.09, which is very similar to the ratio in the ESLO1 corpus (Serpollet, Bergounioux, Chesneau and Walter 2007), which contains interview data from Orléans collected between 1968 and 1971 (Gerstenberg 2011: 238). Gerstenberg (2011: 239) highlights that the diachronic stability in the *nous*-*on* ratio over the past 40 years doesn’t provide a stable ground

[2] *Nous* is used with the first-person plural of the verb, while *on* is accompanied by a third person singular verb.

[3] The LangAge corpus has since been expanded longitudinally a.o. with interviews conducted with the same speakers in 2010, 2015, and 2023; see El Sherbiny Ismail, Gerstenberg, Lupica Spagnolo et al. (2022) for more information.

for assuming an ongoing language change. However, her study shows that the speakers between 71 and 82 use *nous* more frequently than younger ones. The socio-professional status is identified to be a better predictor for a higher *nous* frequency in the first set of LangAge interviews (Gerstenberg 2011: 244).[4]

This overview shows that, while the presence of an ongoing change is controversial, a relation between age and the choice of the clitic variant has been detected repeatedly. The elevated frequency of *nous* in the language use of older individuals has mainly been linked to socio-economic factors or normative pressure that the speakers underwent during their school years. Following the Apparent-Time Hypothesis described in Section 2.1, a possible linguistic change in progress could also manifest itself in a higher frequency of *nous* for older individuals. Results from previous research show the potential relevance of the age variable for the use of the first-person plural clitics and raise the question of whether the text generations by ChatGPT show a similar distribution.

2.2.2 Future Tenses

In modern spoken French, the inflective future tense (*fs*, fr. *futur simple*; e.g. *nous fêterons* 'we will celebrate') co-exists with the periphrastic future tense (*fp*) formed by means of *aller* 'to go' and the infinitive of the verb (fr. *futur proche*; *nous allons fêter* 'we are going to celebrate').

Paoli and Wolfe (2022: 131–132) provide a summary of the findings pertaining to the differences between the periphrastic and synthetic future forms in French. They highlight that the periphrastic form expresses a stronger temporal proximity to the future event (a.o. Blanche-Benveniste 1990: 188), as well as more certainty that this event will occur (a.o. Rebotier 2015: 3), compared to the synthetically formed *futur simple*. Furthermore, the two variants tend to be dependent on grammatical person, linguistic register, and sentence polarity (a.o. Poplack and Turpin 1999). As Paoli and Wolfe (2022: 131–132) document, this complementarity exhibits diatopic variability (see also, as indicated in Paoli and Wolfe 2022; Poplack and Turpin 1999; Wagner and Sankoff 2011; King and Nadasdi 2003) and is subject to inter-speaker variation. Without elaborating further, Paoli and Wolfe (2022: 131) also assume an ongoing change in the distribution of the two future variants.

A notable increase in the frequencies of *fp* in contrast to a decrease of *fs* has been found by Abouda and Scrovec (2015: 9), comparing the corpora ESLO1 and ESLO2 (Serpollet, Bergounioux, Chesneau and Walter 2007 for more details on the corpora). In particular, the increase is mainly due to constructions with *dire* 'to say' introducing a direct speech, e.g., *on va dire* 'we would say', which however do not express futurity.

[4] Gerstenberg (2011) classifies professions into four categories (*ouvriers* 'workers', *employées* 'employees', *cadre moyens* 'middle managers', *cadre supérieurs* 'senior executives'. The higher the occupational group, the higher the proportion of *nous*.

Wagner and Sankoff (2011: 298) detect an increase in the use of *fs* in Montreal French from 1971 to 1984. Younger speakers who in 1971 were identified as categorical users of *fp* in affirmative contexts, increased the use of *fs* and started using it in affirmative contexts in the 1984 data. Wagner and Sankoff attribute this result to an age-grading phenomenon: as they age, speakers adhere to a more conservative linguistic norm. Furthermore, Wagner and Sankoff (2011: 300) observe the difference in increasing frequency of the *fs* to be smaller for participants (+3%) who were 45 or older in 1971 than for younger speakers (+6%). Thus, according to their interpretation, the shift towards the more formal variant *fs* occurs at the transition to full adulthood rather than being a development in later life (Wagner and Sankoff 2011: 304). An affiliation with the middle or upper class is a further predictor for an increase of the formal variant (see also Sankoff and Wagner 2020). The authors see a possible explanation for the discrepancy of an ongoing change in favor of *fp* and the observed longitudinal and cross-sectional tendencies in the brevity of the analyzed time-period. They caution against overestimating the rate at which such a linguistic change might take place (Wagner and Sankoff 2011: 305).

Blondeau (2006) examines the use of the French future variants in a longitudinal study involving 12 participants that were interviewed in 1971, 1984 and additionally to Wagner and Sankoff (2011) also in 1995. She observes an increase of *fs* from 14% to 23% by 1984, followed by a marginal decrease to 22% (categorizable as *stability*) by 1995 (Blondeau 2006: 83). This observation is consistent with the analyses conducted by Wagner and Sankoff (2011). The participants included in Blondeau’s study represent a rather young cohort, their average age being 23 in 1971 (Blondeau 2006: 83–84). She proposes two diachronic trends: the first is a linguistic change marked by an increased frequency of *fp*, which she links to the elevated frequencies of *fp* observed for young participants interviewed in 1984 for the first time (as shown by Zimmer 1994). The other possible trend is an age-grading phenomenon, where the increase in *fs* could be linked to a life stage in which the participants enter the job market (Blondeau 2006: 85).

In summary, similarly to the first-person plural clitics, age has been found to be a predictor for the use of the more formal variant, the *futur simple*, in different varieties of contemporary spoken French. For the future tenses, this observation is mainly attributed to an age-grading phenomenon for speakers passing from young age to adulthood. While there are indicators of an ongoing change in favor of *futur proche*, the possibly complementary use of the variants due to their different temporal-aspectual properties makes it difficult to determine the details of such a change. Micro-diachronic studies revealing an increasing use of *futur simple*, are based on short time frames disclosing rather age-grading patterns than language change.

3 Methodology

This section outlines the methodology employed in our study, which explores the capabilities of ChatGPT, more precisely the GPT-4-turbo 128k model (gpt-4-1106-preview, OpenAI 2023), to generate linguistic patterns reflective of age-specific variance.[5]

3.1 Input Data

Our methodological framework is based on the LangAge Corpus (El Sherbiny Ismail, Gerstenberg, Lupica Spagnolo et al. 2022), which encompasses a collection of biographical interviews conducted with elderly French-speaking individuals. To facilitate an environment beneficial for eliciting the usage of the two linguistic variables of interest, we selected two suitable and recurring themes from the corpus – *festivities* and *future*.

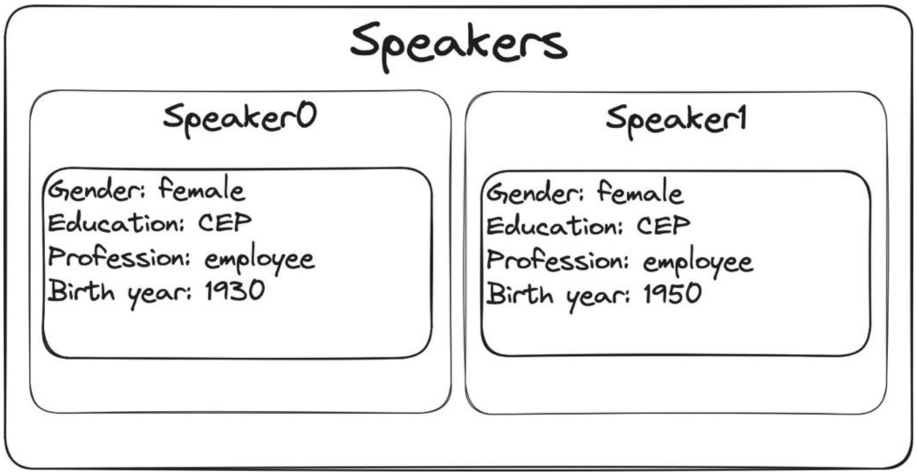
Festivities: The first topic addresses how the interviewees spent Christmas and other celebrations during their childhood. A high occurrence of first-person plural subjects is expected in this context, as the speakers narrate from the perspective of their families.

Future: In the second topic, the interviewees respond to how they imagine their future to be. These responses may include descriptions of future plans. Thus, a relatively high occurrence of future tenses is expected.

As described in Section 2.2, we are aware of the variants of the respective linguistic variables not being perfectly interchangeable. This is particularly true for the future forms for which complex task-sharing dynamics have been described. The choice of the *future* topic aims at reducing the complexity by providing a theme favoring the futurity use of the variants. In order to control the extent of further variability, we used the same answer-framing context across all conversations that belong to the respective theme. It allows us to make more controlled comparisons across different ages and topics, albeit with an understanding that real-world linguistic variability might exceed these experimental conditions. Our study incorporated content from two speakers within the LangAge Corpus, i.e., speakers Mrs. Roger (id 046) and Mrs. Bernard (id 050). English summaries of their interview responses to the two themes were utilized as contextual frames for the answer generation by the LLM. Drawing upon the data from these LangAge speakers, we constructed profiles for two hypothetical speakers to facilitate a controlled generation of responses to the thematic questions. The virtual speakers' metadata (see Figure 1) specified the gender (female), professional history ((former) employee), and educational attainment (*Certificat d'Études Primaires* 'Certificate of Primary Studies', CEP). The distinctive variable was their year of birth, set at 1930 and 1950. Various interview years – 1980, 2000, and 2020 – were

[5] Data, python codes and R-script can be accessed on the GitHub repository https://github.com/Vokksy/ChatAge. Access on request.

also integrated into our prompts. This inclusion allowed the study to address not only age-specific but also generation-specific linguistic variation.



Speakers

Speaker0

Gender: female
Education: CEP
Profession: employee
Birth year: 1930

Speaker1

Gender: female
Education: CEP
Profession: employee
Birth year: 1950

Figure 1: Speaker metadata used for the two virtual speakers delineated in our experiment design. The distinctive variable is *birth year*, which is 1930 for speaker0 and 1950 for speaker1.

By mimicking LangAge Corpus data, we aim to discern the extent to which ChatGPT can replicate the age-specific linguistic variations described in previous studies. Contrary the to the complex picture drawn in the Background section, the described approach limits the range of predictors to the age variable. Other factors (linguistic, geopolitical etc.) that could affect the linguistic variation are intentionally factored out by being controlled in the experiment setting.

The current study thus offers insights into the model’s linguistic adaptability in emulating age-related language preferences.

3.2 Prompting and Conversation Flow

Our methodology involved generating responses from ChatGPT for each speaker across various interview years. For this purpose, three distinct conversations were conceptualized for each combination of speaker, interview year and thematic question (see Figure 2 for the conversational architecture). We ran 30 iterations for each of these combinations. Every conversation contained a single response from the LLM, the first two being independent, the third one depending on the previous two answers.

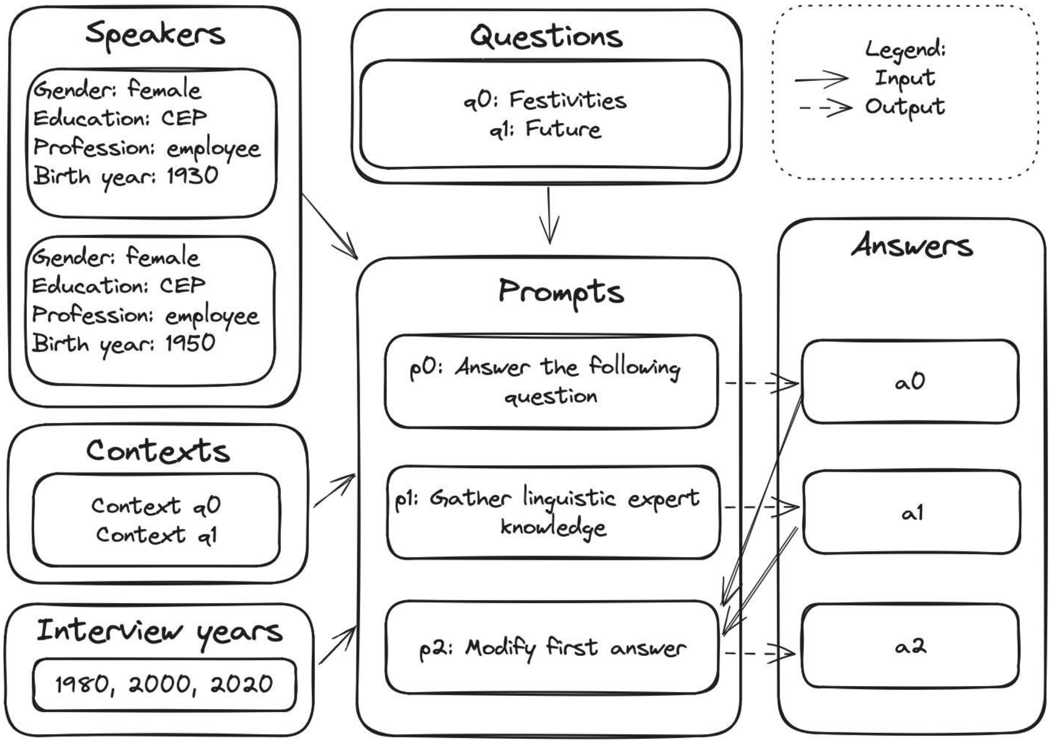


Figure 2: Conversation Architecture: *q0* (*festivities*) and *q1* (*future*) are the two thematic questions being asked in the simulated interview. *p0, p1* and *p2* denote the used prompts, while *a0*, *a1* and *a2* represent the respective ChatGPT's answers. The answers were generated for the three interview years: 1980, 2000 and 2020. In order to answer the two questions, ChatGPT was additionally fed with speaker metadata as well as the question-dependent context (*Context q0*, *Context q1*), an English summary of the answers given to the respective question over all interview years. For every answer, a new conversation was started, so that the current generation wasn't influenced by any previous one. This is also true for *a2*, which did include *a0* and *a1* in its prompt, though.

Naive Answer (answer0, a0): The initial conversation in each set aimed to elicit what we refer to as the 'naive answer' to the questions pertinent to the selected topics. These questions, *les fêtes vous en aviez?* 'Did you celebrate any holidays?' and *le futur c'est quoi?*[6] 'What is the future?', are those found in the LangAge Corpus. The system prompt sent to ChatGPT is the following:

> You are a {age} year old French speaking {gender} living in Orléans. Your education degree is '{education}' and you work(ed) as {profession}. You are participating in a biographic interview in {interview year}. You and the interviewer speak French. {instruction} Your answer should be based on and limited to the following information: {context}

The *instruction* component of the prompt varied depending on the question, either 'The interviewer asks about what you imagine the future to be. Answer to the following question.' or 'The interviewer asks about how you and your family used to celebrate certain festivities. Answer to the following question.'. The *context* was an English summary of responses previously given by the interviewees in the

[6] The wording was adopted from the LangAge corpus.

LangAge Corpus across different interview years (speaker 050 for *festivities* and speaker 046 for *future*). This approach enabled ChatGPT to assume the persona of the described individual and generate responses to the interview questions contextually. We used English as the eliciting language[7] because we didn’t want to influence ChatGPT’s responses with the choice of a specific French variant in the questions.

Expert Knowledge (answer1, a1): The second conversation aimed at ChatGPT generating expert knowledge about the linguistic variable in question (either first-person plural clitic or future tense). We employed the expert-prompting technique as described by Xu, Yang, Lin et al. (2023) instructing ChatGPT to be an expert on the specific linguistic phenomenon and its age-sensitive variation. For this conversation, the following system prompt was sent to ChatGPT:

> The user is interested in language use of elderly speakers. You are a bot, giving answers to the user’s questions by reporting insights from linguistic research. Give only truthful answers. If there is a lot of information, give a summary, if you have only little information, give a list of details that you can find.

The user prompt concludes with the following question specific instructions:

> What does research say about the use of the first-person pronouns ‘on’ and ‘nous’ in the language use of French speaking elderly people compared to the language use of younger people?
>
> What does research say about the use of the future alternatives ‘futur simple’ and ‘futur proche’ in the language use of French speaking elderly people compared to the language use of younger people?

Expert-Knowledge Injected Answer (answer2, a12): The task in the third conversation was to regenerate the naive answer accessing the previously generated expert knowledge. The system prompt was injected with the expert-knowledge, aiming at providing a more grounded modification in the context of age-specific linguistic variation. A comparable undertaking of generating expert knowledge in order to improve an output can be found in Adolphs, Shuster, Urbanek et al. (2021) who used an intermediate seq2seq model for his “Knowledge to Response” approach. Other approaches rely on retrievers for reliable documents to inject expert knowledge and reduce hallucinations (Wang, Wang, Tan et al. 2023). Our decision to leave the gathering of expert knowledge to ChatGPT is based on our intent to enquire how much knowledge about age-related variance it possesses already. The system prompt can here be abstracted to:

[7] We are aware of the possibility that the English variants might influence the choice of the French variants. While this seems less plausible for the first-person plural clitics, it is thinkable for the future tenses. Although we cannot exclude this possible impact from our experiment setting, it is not of immediate relevance for the outcome since our focus lies on the differences between the different answer generations. All answers (answer0 and answer2) share the same English context per topic.

> You have the following interview recorded in {interview year} between an interviewer and the interviewee who is a(n) {age} year old French speaking {gender} with a '{education}' degree and a (former) profession as {profession}, living in Orléans and participating in a biographic interview.
>
> Interviewer: {question}
>
> Interviewee: {naive answer}
>
> And you have the following linguistic knowledge about language use of elderly speakers: {expert knowledge}.
>
> Give only the answer and don't make additional comments.

This conversation is the only one that does include previous LLM outputs. The user prompt following the system prompt requests a modification of the naive answer:

> Taking into consideration the information about the use of the first-person pronouns 'on' and 'nous' in the language use of elderly people, how would you adjust the answer given by the interviewee to the initially asked question accordingly? Answer only by giving the modified answer or the same answer if, after the linguistic consideration, it doesn't need to be modified. Pay attention to the interviewee's age, education degree and (former) profession![8]
>
> Taking into consideration the information about the use of the future alternatives 'futur simple' and 'futur proche' in the language use of elderly people, how would you adjust the answer given by the interviewee to the initially asked question accordingly? Answer only by giving the modified answer or the same answer if, after the linguistic consideration, it doesn't need to be modified. Pay attention to the interviewee's age, education degree and (former) profession!

This approach is meant to facilitate the generation of age-sensitive language variants via inserting a "reasoning step" and adapting the naive answer according to expert knowledge.

3.3 Generated Data and Annotation

Overall, 1080 answers have been generated for this study. Excluding the generation of expert knowledge (answer1), 720 responses (answer0 and answer2) to the two given questions (*festivities* and *future*) were subjected to annotation: two responses (naive and expert) for 30 generations per two speakers and three interview years.

The annotation process is conducted via the same ChatGPT-model (gpt-4-1106-preview, see appendix for prompts) and subsequently corrected manually by the authors of this paper. Variants of future tenses were labeled as _fs_ for *futur simple* and _fp_ for *futur proche*. Concerning the first-person plural, the following variants were annotated: _nous_ for single *nous,* and _on1_ for *on* used as first-

[8] The explicit mentioning of education degree and (former) profession leads back to earlier versions of the experiment, in which additional LangAge (female) speakers' metadata was used to define personas.

person plural subject clitic. Moreover, _on1_ was also used in case *on* was preceded by a stressed *nous*, as in *nous on fêtait*.[9] The final interpretation of *on* was determined by the researchers based on contextual cues. This analysis seeks to determine potential disparities in the utilization of first-person plural subject clitics and future tenses between answer0 and answer2. Furthermore, it aims to investigate whether ChatGPT’s responses exhibit variations based on the age, birth date, and interview year of the individuals being simulated, as well to explore the interplay of these extra-linguistic factors and expert-knowledge on the distribution of age-sensitive variants in answer2.

As for answer1, generations varied in term content. We labeled the information given in the generations of answer1 by asking ChatGPT to classify its own expert knowledge as (i) *nous*/ *fs*/ *on*/ *fp* when a preference of older speakers for a specific variant was expressed, or as (ii) *depends* (= not clear) when no specific variant was indicated as typical for older speakers. This classification will be used in Section 4.3 to examine the effect of answer1 on the distribution of the linguistic variants in answer2.

3.4 Research Questions and Statistics

As indicated in Section 1, the main objective of this study is to investigate the impact of social factors, particularly age and generation, on the utilization of specific linguistic variants in French by ChatGPT, as well as to assess the effect of expert-knowledge injection on ChatGPT’s linguistic choices. This inquiry is articulated through three guiding research questions:

- RQ1: Are there differences in the use of first-person plural subject clitics and future tenses between answer0 and answer2?
- RQ2: If variations are found, does the amount of difference between answer0 and answer2 depend on the birth year, interview year, and/or age of the speakers simulated by ChatGPT?
- RQ3: What is the relative importance of the different sociolinguistic predictors in influencing the preference for one variant over the other, specifically in answer2?

To answer RQ1 (Section 4.1), we look at the proportions, i.e. the relative frequencies of the more informal variants *on* and *futur proche* in relation to the total occurrences of the variable in answer0 and answer2 respectively.[10] Importantly, our null and alternative hypotheses are nondirectional, signifying that we do not anticipate a change in a specific direction (either decrease or increase).

[9] Additionally, we instructed ChatGPT to annotate _on3_ for the pronoun being used as third-person clitic in order to prevent ChatGPT from annotating it as _on1_. _on3_ was not analyzed further.

[10] The results would remain consistent even if we were to use the relative frequency of the other variant.

$$Proportion\ of\ var1 = \frac{Freq_{answer}(var1)}{Freq_{answer}(var1) + Freq_{answer}(var2)}$$

As for RQ2 (Section 4.2), we measure the amount of difference between answer2 and answer0 using the changes in the proportions of a variant in the two responses. The proportion change is calculated by dividing the proportions of a variant in answer2 by the proportions of the same variant in answer0. For this measure, we use the proportion of the more formal, age-marked variants, i.e., *nous* and *futur simple*. This decision is driven by the larger number of data points available for the formal variants compared to the informal ones, thereby ensuring greater statistical reliability. Again, our null and alternative hypotheses are non-directional.

$$Proportion\ change\ = \frac{Proportion_{answer2}(var)}{Proportion_{answer0}(var)}$$

Finally, to evaluate which factors exert a significant influence on the distribution of our variables in the expert-knowledge injected responses (RQ3), we conduct an analysis based on conditional inference trees and random forests (Section 4.3). Conditional inference trees, as described by Levshina (2015: 291), are non-parametric methods employed for regression and classification, relying on binary recursive partitioning. They repeatedly split the data into two subsets, continuing as long as there is a significant association between the target and predictor variables. Conditional random forests are constructed by a large number of conditional inference trees and allow to evaluate more precisely the relative importance of single predictors.

Like generalized linear mixed-effects models, conditional inference trees and random forests enable the inclusion of both random factors (such as *speaker*) and fixed factors (such as *interview year*), avoiding bias towards variables with many levels or continuous predictors, and are also particularly robust to the presence of outliers (see Tagliamonte and Baayen 2012 for a comparison between these statistical approaches). Moreover, since conditional inference trees and random forests are non-parametric methods for algorithmic modeling, they do not assume a specific stochastic distribution of the data and are especially useful in scenarios like our research, where the number of observations is relatively low, but there are numerous potential predictor variables to consider (see also Levshina 2021). In addition to their applicability with limited sample sizes, conditional inference trees and random forests have proven particularly valuable when confronted with a high number of collinear predictor variables, that is variables that are correlated and cannot independently predict the value of the dependent variable (Tagliamonte and Baayen 2012). This flexibility aligns well with our study, where predictors such as *age* and *birth year* (the latter broadly representing *generation*) exhibit collinearity.[11]

[11] On the contrary, interpreting the results of other models, such as generalized linear mixed-effects models, is more challenging in the presence of intercorrelations among independent variables in the dataset (Tagliamonte and Baayen 2012: 22).

We conducted our statistical analysis in R (R Core Team 2021), utilizing the packages *ggplot2* (Wickham 2016), *dplyr* (Wickham, François, Henry et al. 2023), *caret* (Kuhn 2008), *Hmisc* (Harrell 2023) for data visualization and manipulation, *party* (Hothorn, Hornik and Zeileis 2006), and *randomForest* (Liaw and Wiener 2002) for conditional inference trees and random forests. In addition, we used the phyton libraries *seaborn* (Waskom 2021), *matplotlib* (Hunter 2007) and *pandas* (The pandas development team 2020; McKinney 2010) for data visualization, transformation, and aggregation.

4 Results

4.1 Naive and Expert-Knowledge Injected Answers

The absolute and relative frequencies of *nous* and *on* across answer0 and answer2 are displayed in Table 2 (subject reduplications, such as *nous […] on*, were counted as instances of *on*). Additionally, Figure 3 visually illustrates the proportions of *on* in answer0 and answer2.

Table 2: Absolute and relative frequencies of *nous* and *on* in a0 and a2.

	nous	***on***	**total**
Answer0	749 (63%)	444 (37%)	1193 (100%)
Answer2	1091 (87%)	160 (13%)	1251 (100%)

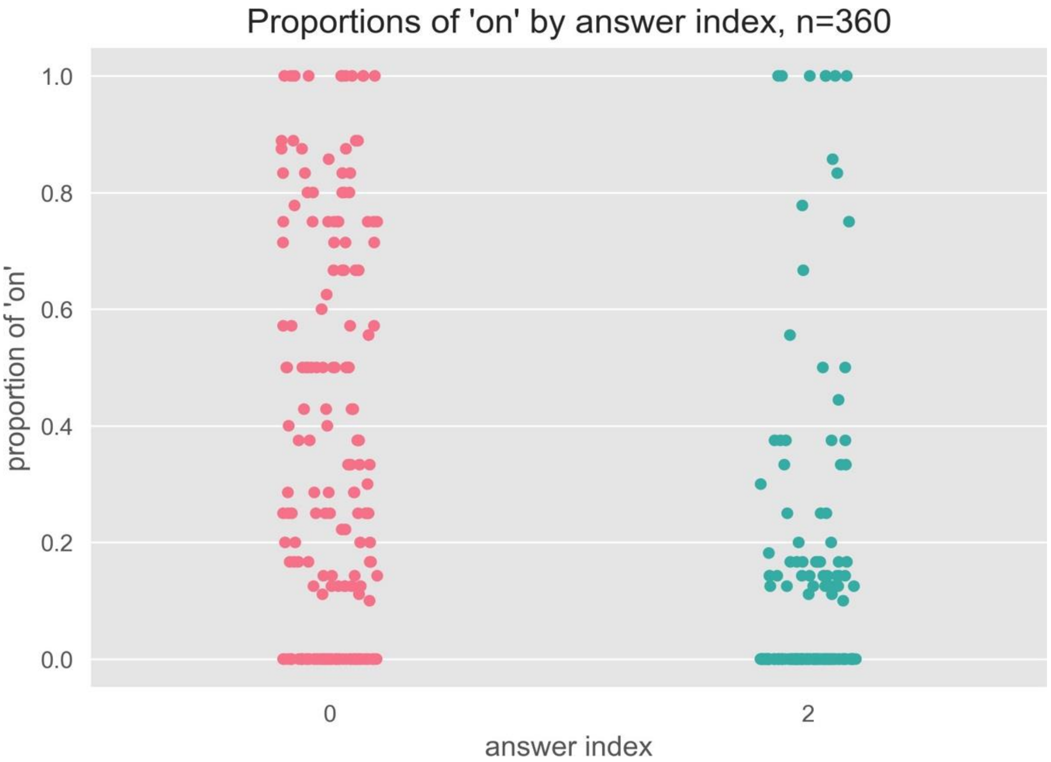


Figure 3: Proportions of *on* in answer0 and answer2 over all conversations. Each dot represents the proportion of *on* over the total in a specific iteration by answer.

As Figure 3 shows, the proportions of *on* are often equal to 0, which is the case for 54 of the 180 instances (30%) of answer0 and 119 of the 180 instances (66%) of answer2. This indicates a tendency of ChatGPT to avoid *on* and use *nous* when expressing first-person plural subject clitics in our data (see also Table 2). Nevertheless, the mean and the standard deviation of *on* proportions in naive responses (mean = 0.36; sd = 0.334) are higher than those in expert-knowledge injected responses (mean = 0.12; sd = 0.242), pointing to greater variability in pronominal choices in naive answers compared to expert-knowledge injected ones. In the latter, the more formal variant *nous* clearly emerges as the preferred choice (1091 occ., 87%). According to a paired two-tailed Wilcoxon test, the difference in the estimated median proportions of *on* between answer0 and answer2 is statistically significant (paired two-tailed Wilcoxon test: V = 5892, p-value < 0.001***).[12]

The distribution of future tenses exhibits both similarities and differences in comparison to the distribution of first-person plural subject clitics, as illustrated in Table 3 and Figure 4.

Table 3: Absolute and relative frequencies of *futur simple* and *futur proche* in a0 and a2.

	futur simple	***futur proche***	**Total**
Answer0	436 (80%)	106 (20%)	569 (100%)
Answer2	1210 (95%)	66 (5%)	1276 (100%)

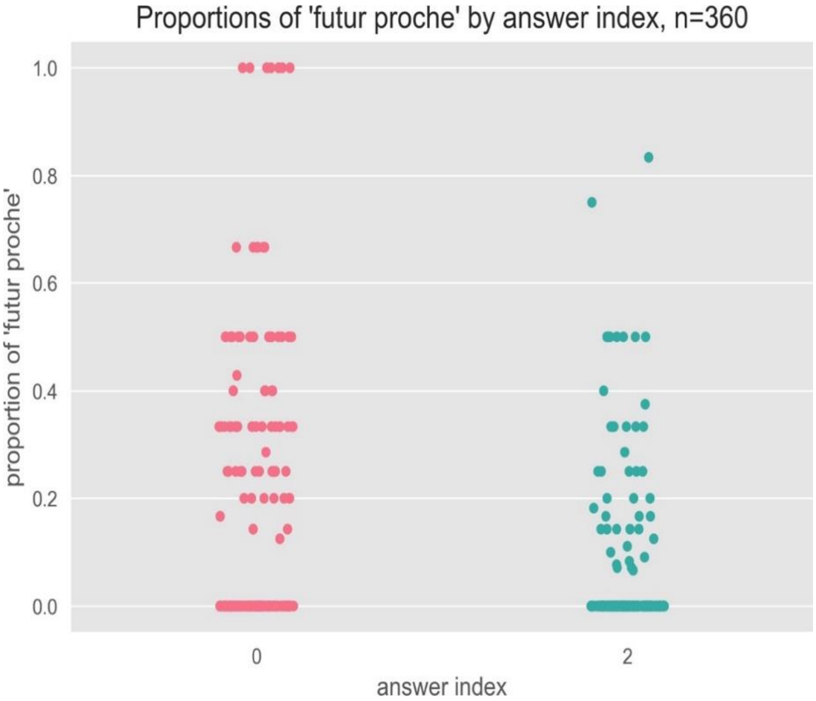


Figure 4: Proportions of *futur proche* in answer0 and answer2. Each dot represents the proportion of *on* over the total in a specific iteration by answer.

[12] We conducted a paired, two-tailed Wilcoxon test for the following reasons: (i) the data points in answer0 and answer1 are dependent, (ii) their differences do not follow a normal distribution as confirmed by a Shapiro–Wilk test, and (iii) our null hypothesis is not directional.

In terms of differences, the means and standard deviations of periphrastic future proportions are comparatively lower when contrasted with those corresponding to the clitic *on* in both naive and expert-knowledge injected responses (answer0: mean_*fp* = 0.19, sd_*fp* = 0.26; answer2: mean_*fp* = 0.07, sd_*fp* = 0.149). This suggests reduced variability in the expression of future tenses by ChatGPT compared to the formulation of first-person plural subject clitics. Thus, for instance, in 98 of the 180 naive responses (54%), ChatGPT never employs the *futur proche.* Even less variation is found in answer2, where the periphrastic future is absent in 137 out of 180 responses (76%). Overall, we observe a consistent decrease in the total number of periphrastic futures from answer0 to answer2 (refer to Table 3).

As for the similarities, we note, also in the case of future tenses, a statistically significant reduction in the proportions of the less formal variant, i.e., *futur proche*, in expert-knowledge injected responses compared to the naive ones (paired two-tailed Wilcoxon test: V = 335.5, p-value < 0.001***). These results substantiate the influence of expert-knowledge injection on the proportions of the linguistic variants. Across the entire dataset, encompassing all ages, birth years, and interview years considered in the study, there is a discernible increase of the proportions of *nous* and *futur simple* from answer0 to answer2.

4.2 Comparison by Birth Year, Interview Year, and Age

The boxplots presented in Figure 5 illustrate the changes in the proportions of *nous* from answer0 to answer2, organized by speakers’ birth year and interview year.

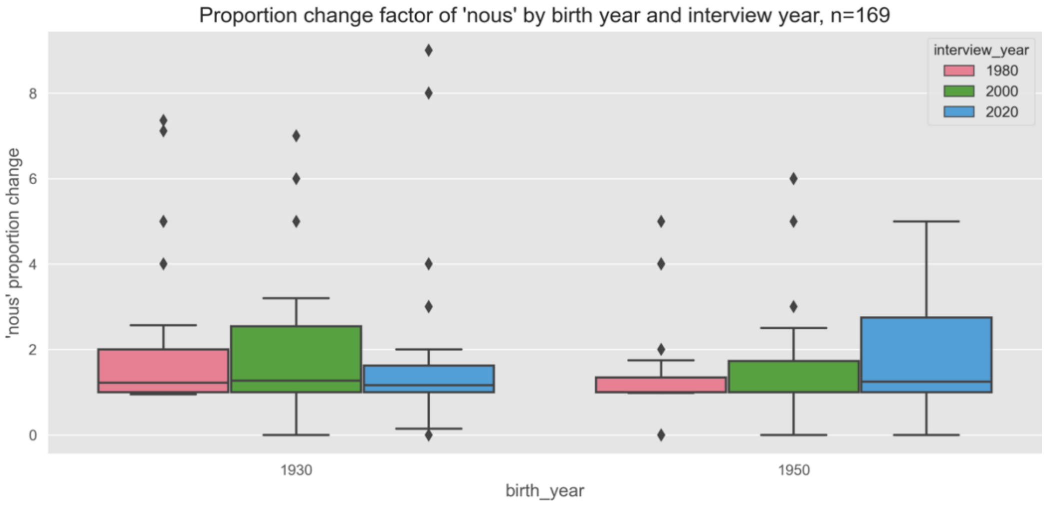


Figure 5: Proportion change of *nous* from answer0 to answer2.

According to a series of paired two-tailed Wilcoxon tests, the proportion change of *nous* between answer0 and answer2 does not show significant differences across birth years and interview years. Nevertheless, it is worth noting that the proportion change exhibits more variation for speaker0 (born in 1930) in 2000 and speaker1 (born in 1950) in 2020 (compare the interquartile ranges of the boxplots in Figure

5), that is, for both speakers, at the speaker age of 70. Notably, the interquartile range of *nous* proportion change is 1.625 for 70-year-olds versus 1.0 for 50-year-olds, 0.625 for 90-year-olds, and 0.342 for 30-year-olds.

The reason for this phenomenon can be found by examining in more detail the answers generated for 70-year-old speakers (see the descriptive statistics in Table 4 and Figure 6).

Table 4: *nous* proportions statistics at speaker age 70.

answer	**birth year**	**count**	**mean**	**std**	**min**	**25%**	**50%**	**75%**	**max**
0	1930	30.0	0.570	0.331	0.000	0.298	0.500	0.865	1.0
0	1950	30.0	0.641	0.302	0.111	0.321	0.732	0.875	1.0
2	1930	30.0	0.894	0.240	0.000	0.875	1.000	1.000	1.0
2	1950	30.0	0.881	0.199	0.000	0.857	0.938	1.000	1.0

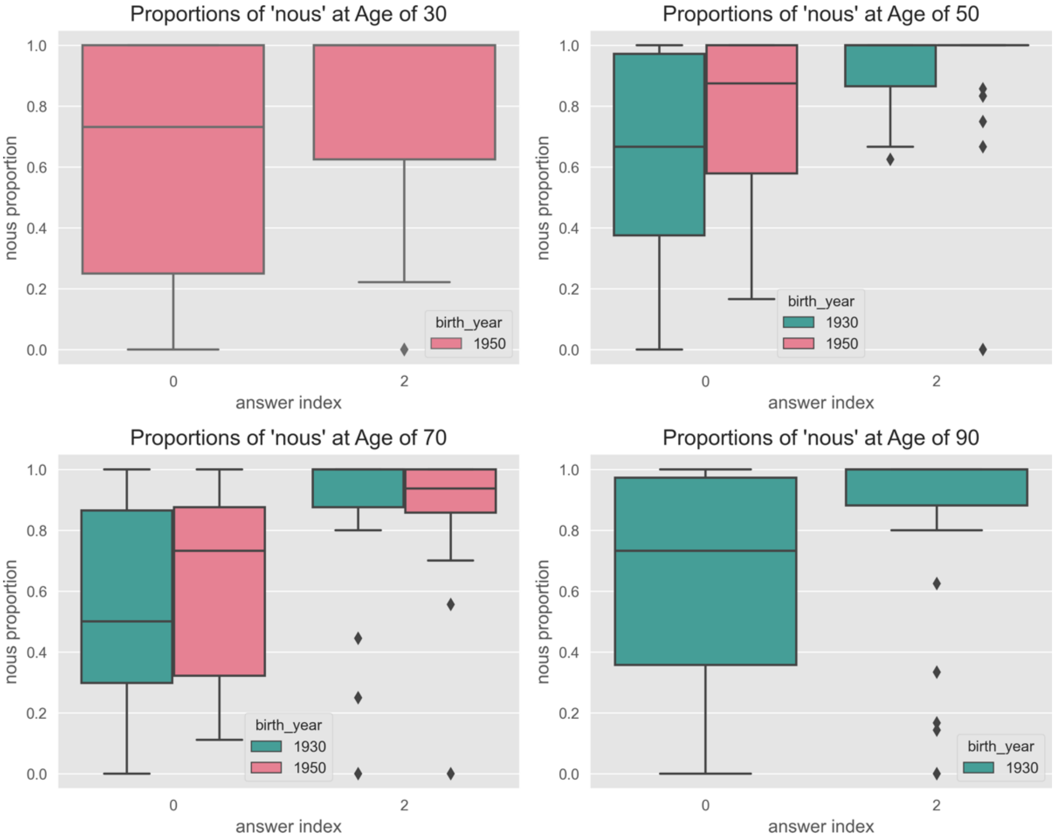


Figure 6: Proportion of *nous* in answer0 and answer2 depending on the speaker's year of birth at different speaker ages.

In answer0, the values of the middle interquartile range (IQR) fall between 0.298 and 0.865 for speaker0 and 0.321 and 0.875 for speaker1. In contrast, in answer2, the second and third quartiles are notably higher, ranging between 0.875 and 1 for speaker0 and 0.857 and 1 for speaker1 (see Table 4). This can be interpreted as a homogenization of *nous* proportions towards values close to 1 in the expert-knowledge injected answer (a2).

Furthermore, the medians of the *nous* proportions in answer0 diverge remarkably between the two speakers in 2000 (see Figure 7). For speaker1 (birth year 1950) the medians remain relatively stable around 0.75 while speaker0’s medians range from 0.5 to 0.88. The generally lower medians for speaker0 are not consistent with the Apparent-Time Hypothesis (see Section 2). Assuming a language change in favor of *on*, the Apparent-Time Hypothesis would predict a more conservative language use for speaker0 (a higher proportion of *nous*), than for speaker1. The data also lacks evidence for the existence of age-grading phenomena since no common age-specific patterns are to be found for the two speakers.

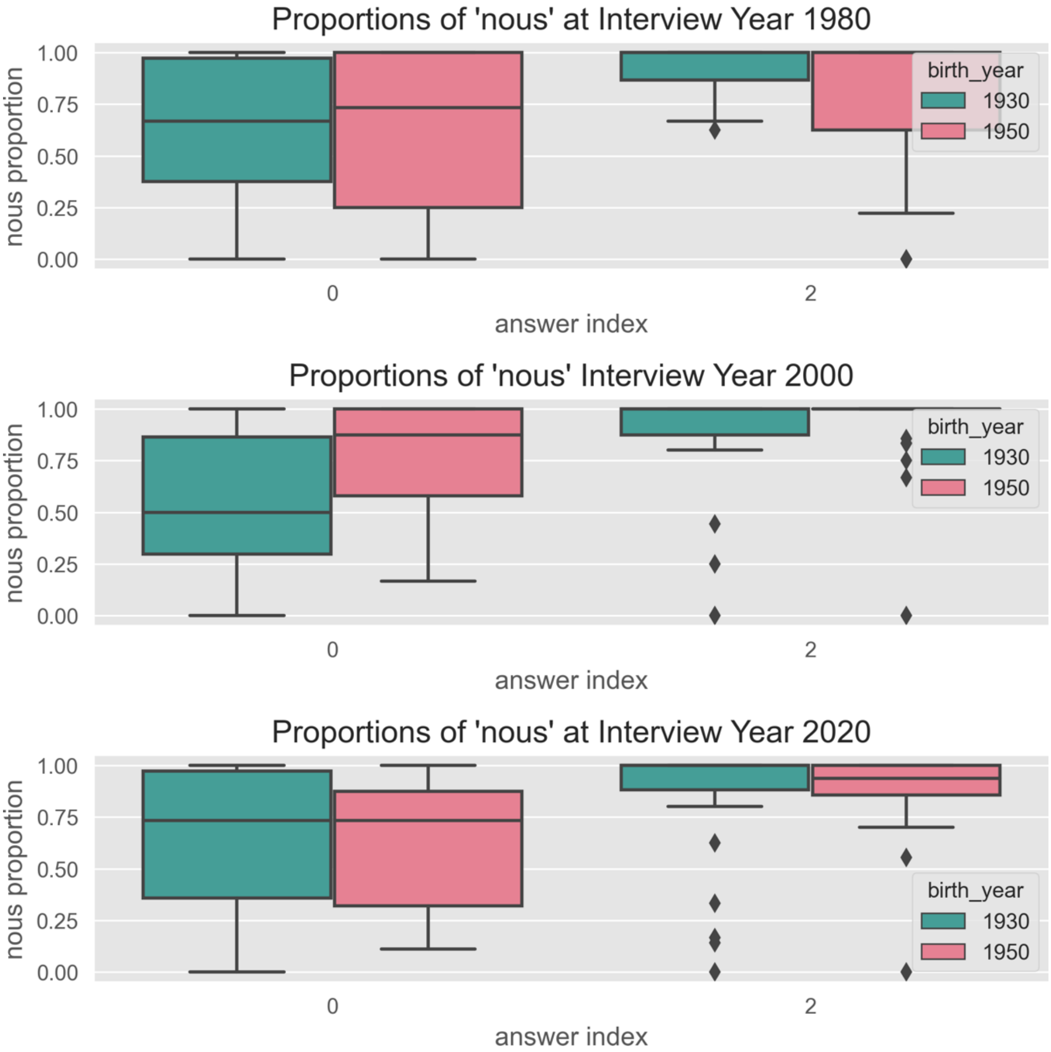


Figure 7: Proportion of *nous* in answer0 and answer2 depending on the speaker's year of birth at different interview years.

These observations are in line with the distribution of the proportion change of *futur simple*, as illustrated in Figure 8.

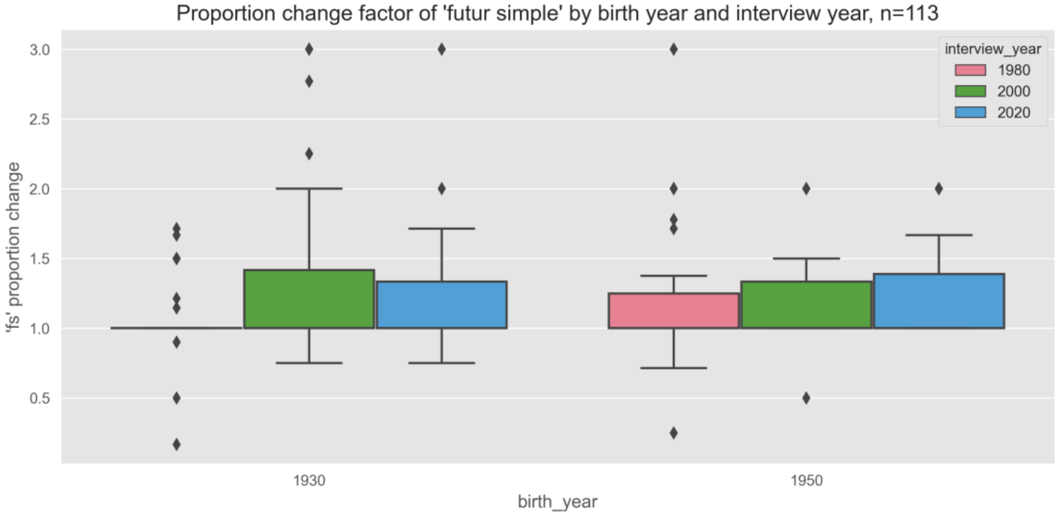


Figure 8: Proportion change of *futur simple* from answer0 to answer2.

Similar to the first-person plural subject clitics, the proportion change of *futur simple* from answer0 to answer2 exhibits greater variability for speakers at the age of 70 (IQR = 0.389 at age 70, compared to IQR = 0.24, 0.33, and 0.25 at the ages of 50, 90, and 30, respectively).[13] Specifically, passing from answer0 to answer2, the proportions of *futur simple* get lifted to 1.0 for almost every answer2 for speakers aged 70, as shown in Table 5 and Figure 9.

Table 5: *Futur simple* proportions statistics at speaker age 70.

answer	**birth year**	**count**	**mean**	**std**	**min**	**25%**	**50%**	**75%**	**max**
0	1930	30.0	0.755	0.322	0.00	0.518	1.0	1.0	1.0
0	1950	30.0	0.809	0.245	0.00	0.667	1.0	1.0	1.0
2	1930	30.0	0.972	0.077	0.75	1.000	1.0	1.0	1.0
2	1950	30.0	0.959	0.113	0.50	1.000	1.0	1.0	1.0

[13] Notably, the difference in the *futur simple* proportion change is statistically significant for speaker0 between 1980 and 2000 (paired two-tailed Wilcoxon test: V = 46, p-value = 0.04847*).

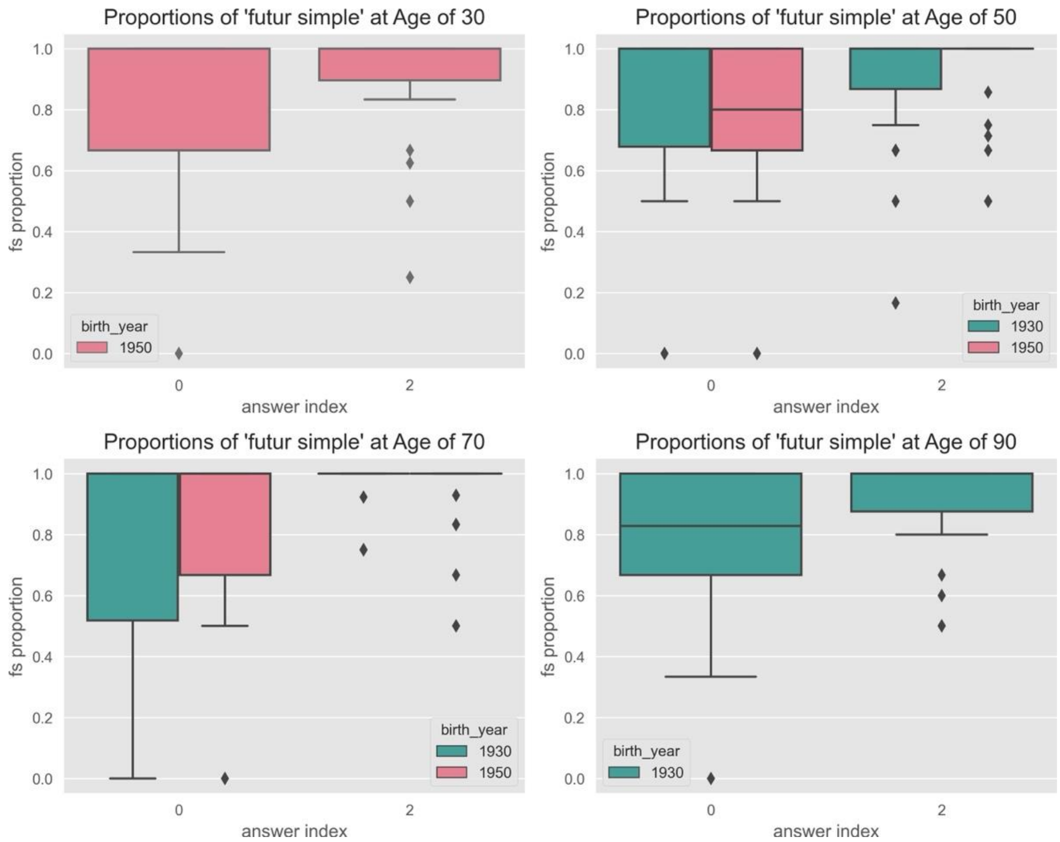


Figure 9: Proportion of *futur simple* in answer0 and answer2 depending on the speaker's year of birth at different speaker ages.

By contrast, the medians of the proportions of *futur simple* diverge considerably less compared to *nous* (compare Figure 7 and Figure 10). This might be due to the overall low frequencies of *futur proche*.

It must be noted though, that, on average, for all combinations of birth year and interview year, the average proportion-change of both variables is ≥ 1.0, as can be seen in the Table 6 and Table 7. This reveals that, on average, the modification of answer0 entails an augmentation of the proportions of the formal variants, irrespective of birth year and interview year (see also Section 4.1).

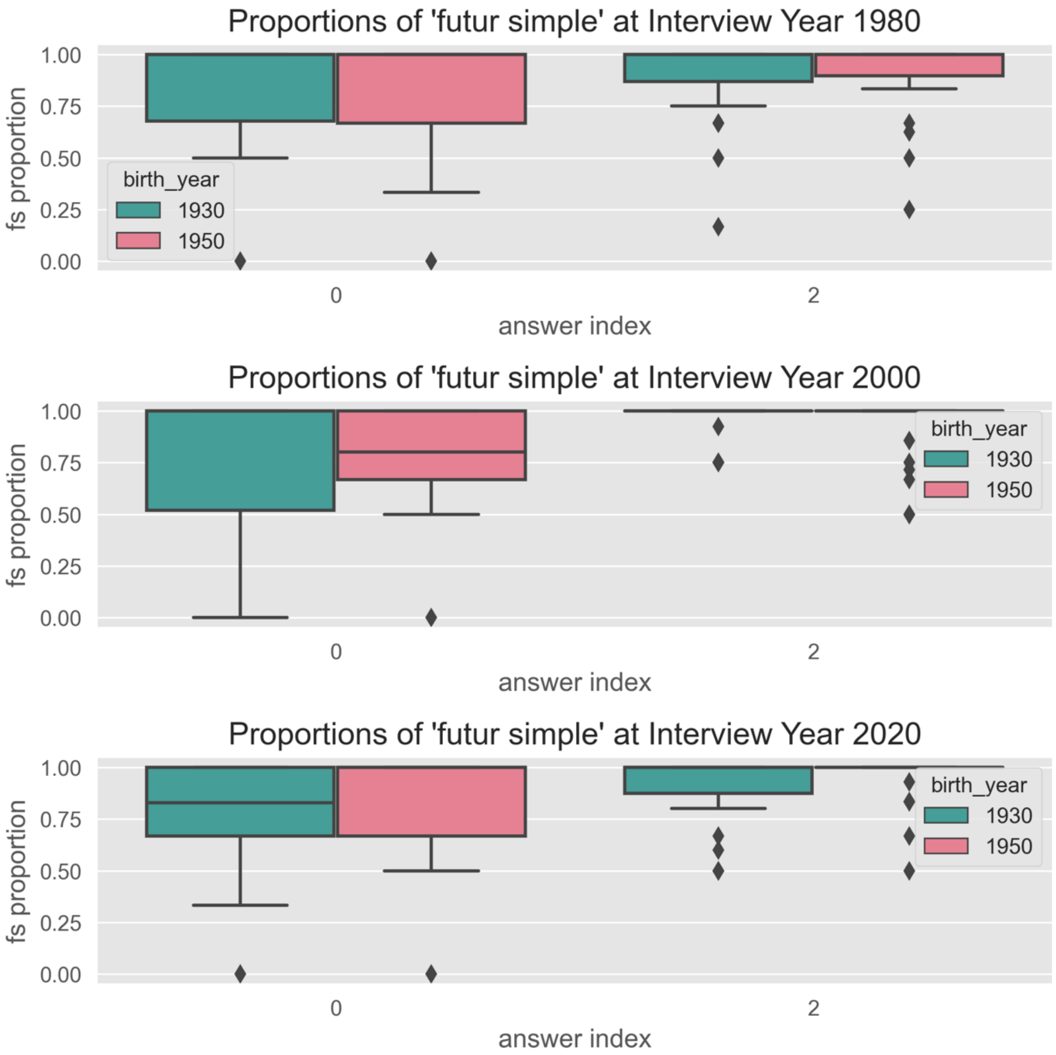


Figure 10: Proportion of *futur simple* in answer0 and answer2 depending on the speaker's year of birth at different interview years.

Table 6: Mean ($\bar{x}$) and median ($\tilde{x}$) of proportion changes for *nous* by birth year and interview year.

Birth Year	**Interview Year**	$\bar{x}$	$\tilde{x}$
1930	1980	1.23	2.09
1930	2000	1.27	2.04
1930	2020	1.17	1.86
1950	1980	1.00	1.52
1950	2000	1.00	1.69
1950	2020	1.25	1.90

Table 7: Mean ($\bar{x}$) and median ($\tilde{x}$) proportion changes for *futur simple* by birth year and interview year.

Birth Year	**Interview Year**	$\bar{x}$	$\tilde{x}$
1930	1980	1.06	1.00
1930	2000	1.39	1.00
1930	2020	1.21	1.00
1950	1980	1.19	1.00
1950	2000	1.16	1.00
1950	2020	1.20	1.00

In summary, the data discussed in this section indicate a higher variation in the proportion change of both variables when mimicking 70-years-old speakers. This higher amount of variation in proportion change is specifically determined by the notable increase in the proportions of the formal, age-marked variants in answer2 compared to answer0. In contrast, the variation in the proportion changes from answer0 to answer2 for the same speakers when they are younger (30, 50 years old) or older (90 years old) is notably lower. The reasons for this distribution may range from ChatGPT’s greater insecurity in interpreting the expert knowledge generated in answer1 for 70-year-old speakers because they are nearing the onset of old age, to possible bias in the distribution of training data for different age groups.[14] However, since it is not possible to access ChatGPT’s training data, we cannot assess which factors not controlled for by our experiment design are responsible for this result.

Thus, even if neither the clitics nor the future tenses indicate a clear diachronic development, the descriptive statistics discussed in this section suggest some slight longitudinal differences in the impact of answer1 in our data, which will be further explored in the next section. In this regard, it is worth noting that the occurrences of informal variants, particularly *futur proche*, are sparse compared to the overall dataset. Due to these sampling imbalances and the collinearity among some examined variables (see *birth year* and *age*), we opt to perform further analysis using random forest and conditional tree methods. These methods enable us to evaluate the relative importance of different factors on the observed distributions (see Levshina 2015: 291–300; Tagliamonte and Baayen 2012 for an overview).

4.3 Exploring Factors Influencing Variant Distribution in Answer2

In the third explorative phase of our analysis, we exclusively focus on the distribution of our variables in answer2. Our research question aims to uncover the relative importance of different factors influencing the preference for one variant over the other in the expert-knowledge injected answers (a2). Potentially relevant

[14] We would like to thank an anonymous reviewer for suggesting the latter reason.

predictive factors include speaker metadata, such as birth year and age, the interview year, conversational idiosyncrasies,[15] and ChatGPT' s expert estimation. The latter variable is a classification of the expert knowledge elicited in answer1 performed by ChatGPT itself. It includes the categories (i) *nous*/ *fs*/ *on*/ *fp* or (ii) *depends* (= not clear), indicating the preference of older speakers for a specific variant.

The distribution of the variants of our two variables is highly uneven across the entire dataset. Even if we concentrate solely on answer2, we find 1091 instances of *nous* against 160 instances of *on* as well as 1210 instances of inflected futures as opposed to 66 instances of periphrastic futures (see Table 2 and Table 3 in Section 4.1). To prevent the random forest from achieving high classification accuracy by simply learning to classify all occurrences as *nous* or as *futur simple*, we trained our model using a subset of our data ("undersampling"). Specifically, we randomly selected a sample of 160 instances of *nous* and 66 instances of *futur simple*, balanced for our independent variables, to be compared with all occurrences of *on* and *futur proche* in answer2.[16] By reducing the sample size of the majority class to match the sample size of the minority class more closely, we can more accurately assess the performance of our model.

Figure 11 shows the relative importance of the four investigated predictors in our random forest for the first-person plural subject clitics.

[15] Idiosyncrasies in our study cannot be attributed to a particular speaker, as no speaker is consistent across iterations and interview years. In fact, each speaker's answer-set iteration is independent of the previous one. Characteristics that might seem to be attributable to a particular speaker are instead related to a characteristic of that speaker, such as defined in the experimental design, i.e. her year of birth and/or her age. Idiosyncrasies only manifest themselves at the conversation or answer level. Therefore, we did not include *speaker* as a random factor in our models.

[16] To create a balanced subset, we used the package 'caret' in R (Kuhn 2008). We thank Marc Schalberger from the statistical consulting team (*fu:stat*) at Free University of Berlin for this suggestion.

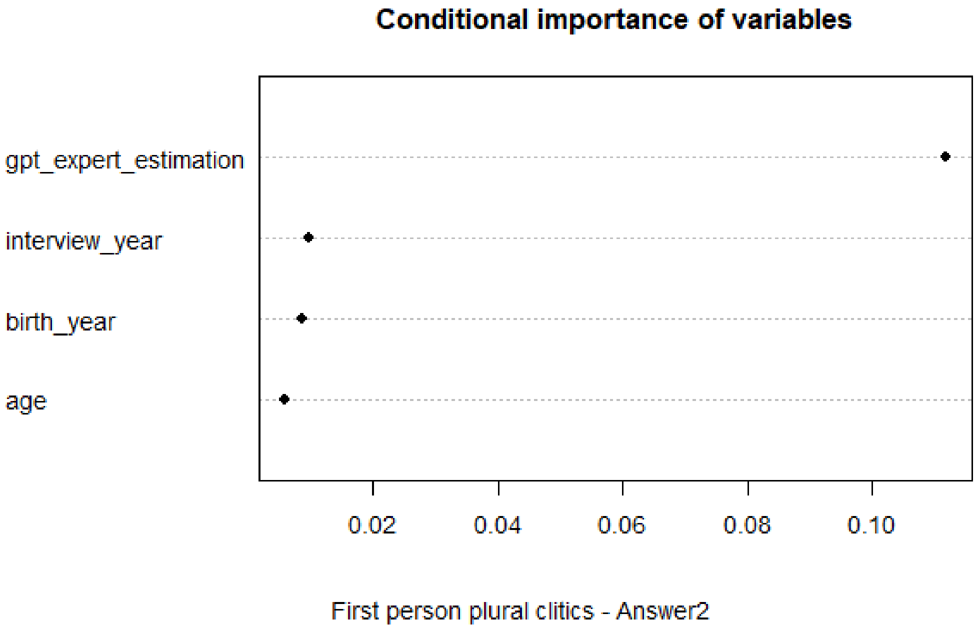


Figure 11: Dot chart depicting the conditional importance of predicting factors in the choice of first-person plural subject clitics in answer2.

As depicted in Figure 11, the significance of the majority of examined predictors hovers around zero, indicating their lack of relevance in explaining the variation between *nous* and *on* in answer2. The notable exception is the factor *ChatGPT expert estimation* (relative importance = 0.11), which stands out as the overwhelmingly most important predictor influencing the choice of first-person plural subject clitics in answer2. Specifically, as it emerges from the confusion matrix of this random forest, the model performs well in identifying *nous*, as evidenced by a higher number of true positives (80%). By contrast, it shows some limitations in distinguishing *on*, as indicated by the higher presence of wrong classifications (42%).

To evaluate the performance of our random forest, we used two cross-validation measures: classification accuracy and the concordance index (also known as the *C*-index). Classification accuracy is calculated by dividing the number of correct predictions by the total number of observations, while the *C*-index is a metric used to evaluate the predictive accuracy of a model and is particularly useful for binary response variables, as in our case (see Levshina 2021 for more details on these cross-validation measures). The *C*-index can be calculated using all samples, including those used for training, or just the out-of-bag (OOB) samples (Levshina 2021: 636–637). The classification accuracy (0.69) and the *C*-index (0.73) of the random forest in Figure 11 indicate that this model discriminates between *nous* and *on* acceptably well.[17] However, if we only use the OOB samples, the *C*-index, which reflects the predictive power of our random forest, falls to 0.62. As suggested

[17] According to Levshina (2021: 633), a *C*-index ranging from 0.7 to 0.8 suggests acceptable discrimination, while a range between 0.8 and 0.9 indicates good discrimination, and a *C*-index above 0.9 means excellent discrimination.

by Levshina (2015: 299), we reran our random forest several times using different random number seeds and various preselected predictors at each split, and the results remained consistent.

The conditional inference tree for the same predictors used in the random forest in Figure 11 is showcased in Figure 12. The single tree exhibits a lower classification accuracy (0.65) and a lower concordance index (*C*-index = 0.65) compared to the random forest, indicating reduced reliability in terms of predictive capability. Nevertheless, the conditional inference tree provides a visual representation of how response variants are distributed depending on predictors.

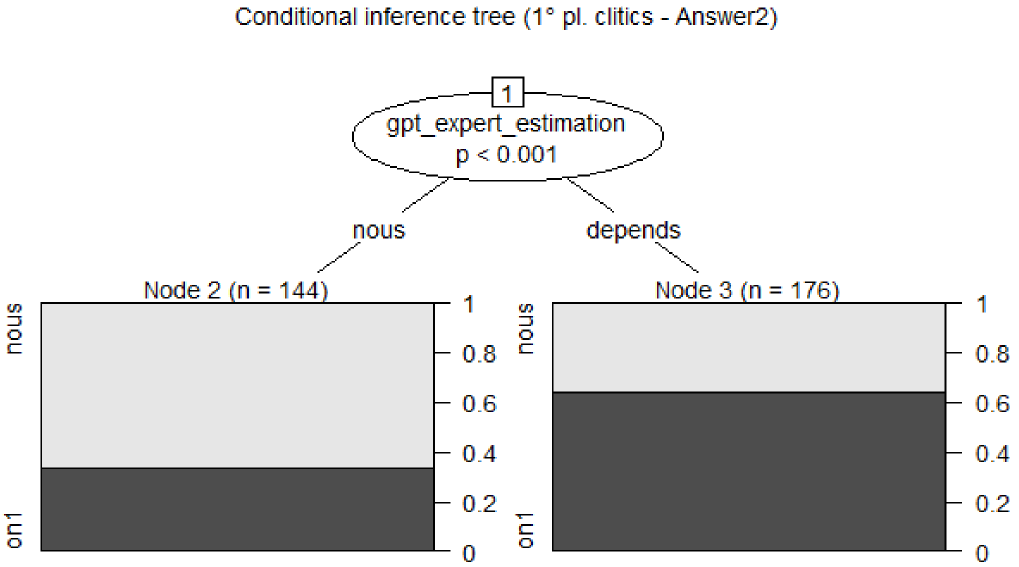


Figure 12: Conditional tree of predicting factors for first-person plural subject clitics in answer2.

In line with the results of the random forest analysis, the conditional inference tree in Figure 12 confirms that ChatGPT's choice of pronouns in answer2 is primarily influenced by the expert knowledge in answer1 (p < 0.001***). For example, if ChatGPT states in answer1 that older people tend to use *nous* in their speech, it will use this pronoun in the subsequent answer2 about 65% of the time. On the other hand, *on* is preferred in about 65% of cases when the expert prompt in answer1 does not give a clear preference for the more appropriate pronoun for an elderly speaker (ChatGPT's expert estimation of answer1 = *depends*).

In contrast to first-person plural subject clitics, all examined factors (i.e., ChatGPT's expert estimation, age, interview year, birth year) for future tenses lie around zero according to our random forest in Figure 13 (accuracy = 0.64; *C*-index = 0.68; out-of-bag *C*-index = 0.54). Therefore, none appears to contribute significantly to shaping the distribution of *futur simple* vs. *futur proche* in ChatGPT's answer2. One possible explanation for the lack of significance of the examined extralinguistic factors could be the still higher relative importance of linguistic factors in shaping the distribution of future tenses (see Section 2.2). However, further research is needed to investigate this hypothesis.

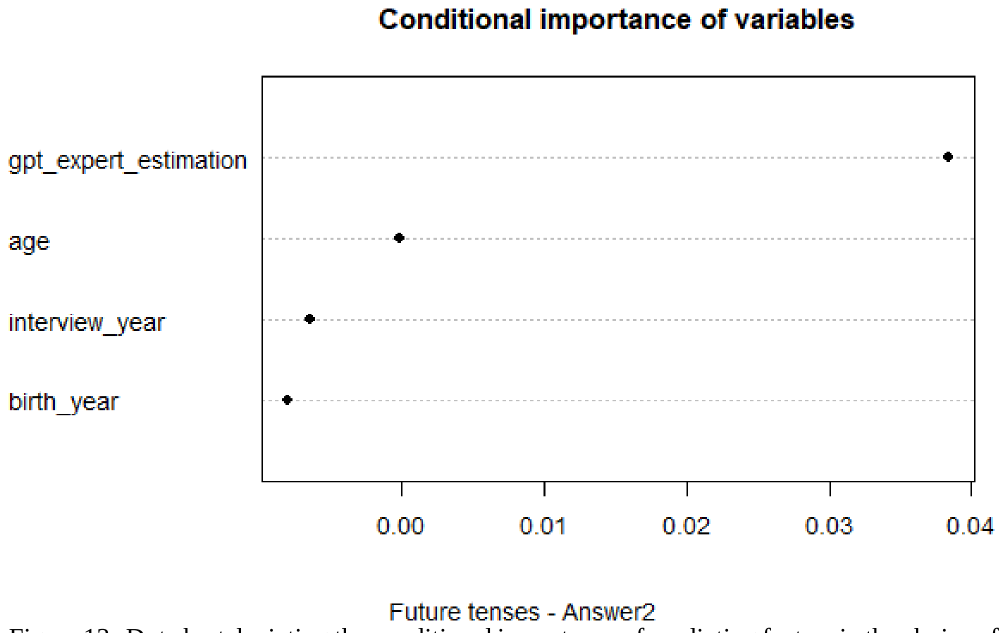


Figure 13: Dot chart depicting the conditional importance of predicting factors in the choice of future tenses in answer2.

In summary, the results derived from this analysis present an intriguing paradox. On the one hand, in answer1, ChatGPT predominantly identifies the formal variants – *futur simple* (110 out of 180 cases) and *nous* (109 out of 180 cases) – as being preferred by older speakers, leading to modifications of answer0 in favor of the age-marked variants in answer2, at least with respect of first-person plural clitics. On the other hand, the modification in answer2 does not seems to significantly reflect age as a determining factor.

To address the risk of skewed model accuracy due to overrepresentation of the most formal variants, we applied an “undersampling” technique in our analysis. This ensures that the model doesn’t merely reflect the existing biased probability distribution of the linguistic variables but can also be used to classify unseen data in the test samples. However, this method considerably reduced the size of our dataset and possibly the predictive power of our random forests and conditional trees (as indicated by their relatively low classification accuracy and *C*-index measures). To gain a deeper understanding of age-related linguistic variation in texts automatically generated by Large Language Models (LLMs) such as ChatGPT, further research with larger, more balanced samples is therefore needed.

5 Discussion

Our study revealed the expert-knowledge in answer1 to be the most predictive factor when it comes to variant proportions in answer2. Neither birth year nor age exerted a systematic influence on the use of age-sensitive variables. Consequently, the most salient dynamic discerned in our analysis is ChatGPT’s tendency to increase the usage of formal variants from answer0 to answer2, a pattern that appears to be independent of the speaker's age or birth year.

In this section,[18] we aim to elucidate and critically examine some of the nuances and constraints inherent in our dataset and findings. This exploration is intended to shed light on the peculiarities and the potential limitations that emerged from our analysis.

Hypercorrective Tendencies: In the majority of instances of expert-knowledge generation (answer1), ChatGPT identified the formal variants as those to be preferred by older adults (109 out of 180 for first-person plural subject clitics and 110 out of 180 for future tenses). This alignment led to an increased frequency of formal variants in the subsequent expert-knowledge injected answers (answer2), observed in 112 out of 180 instances for clitics and, and even more prominently, 141 out of 180 instances for future tenses.

Notably, some instances of the language model's output exhibit hypercorrective tendencies. Such hypercorrections manifest in alterations from impersonal to personal forms, exemplified by the transformation of *on* to *nous* in phrases like "À cette époque, *on* n'avait pas un accès facile aux bibliothèques" ('At that time, people didn't have easy access to libraries') to "À cette époque, *nous* n'avions pas un accès facile aux bibliothèques" ('At that time, we didn't have easy access to libraries') (q0_s0_y1980_i001).

Furthermore, hypercorrections involve an increased overall frequency of first-person plural clitics and future tenses. Specifically, the proportion of clitics shifted from the initial naive answer (answer0) to the expert-informed answer (answer2) without an overall frequency increase in only 56% of examined cases, i.e., 101 of the 180 instances. For future tenses, this phenomenon was even more infrequent, occurring in merely 3 out of 180 cases (1.7%). In terms of future tenses, hypercorrections result in the insertion of future tense constructions where none previously existed, as seen in the transformation from "Je *pense* passer du temps dans mon jardin" ('I am thinking about spending time in my garden') to "Je *penserai* passer du temps dans mon Jardin" ('I will be thinking about spending time in my garden') (q1_s1_y1980_i007). Similarly, the frequency of first-person plural pronouns escalates when they replace other pronouns, as in the alteration from "Même avec notre petit groupe, *je* me sentais très heureuse et gâtée" ('Even with our little group, I was feeling very happy and spoiled') to "Même avec notre petit groupe, *on* se sentait très heureux et gâtés" ('Even with our little group, we were feeling very happy and spoiled') (q0_s1_y1980_i016).

Contextual Impact: We intended to control the semantic and grammatical impact that a context might have on the variation of our variables (see Section 2.2) by providing the same question-dependent context to each conversation. Nonetheless, the divergence in the responses exceeds the choice of the variants. A

[18] Due to the heterogeneous nature of the data within the LangAge corpus and the contrast between our systematic approach to text generation and the more unstructured interview format used in LangAge, we have decided not to compare our results with the respective distributions in the LangAge data described in Section 2.

further detailed examination, particularly a qualitative analysis of answer1, could hence provide deeper insights into the factors influencing the selection of first-person plural clitics and future tense forms. In this regard, a preliminary analysis of answer1 reveals a range of considerations, potentially affecting the linguistic choices observed in answer2.

For instance, in the context of choosing between *nous* and *on,* responses in answer1 indicate that this selection is guided by a combination of personal linguistic habits, the conversational context, and the level of formality of the communicative situation (q0_s0_y2000_i007, q0_s0_y2020_i018, q0_s0_y2020_i024). In terms of the future tense variants, the responses in answer1 suggest that the choice is influenced not only by conceptual and individual preferences (q1_s0_y1980_i002, q1_s0_y1980_i022) but also by subtle semantic distinctions between the variants (q1_s0_y1980_i002). Additionally, it’s crucial to recognize the methodological constraints inherent in this research domain. As one ChatGPT response aptly points out:

> [...] it’s essential to acknowledge the methodological limitations of research in this area: a lot of it depends on the corpus (data set) and the specific group of elderly individuals studied. (q1_s1_y2020_i009)

An exploration of the factors influencing the choice of variants that fall outside our sociolinguistically motivated approach extends beyond the scope of the current study. However, an examination of these factors, such as temporal proximity, grammatical person, or sentence polarity for the future tenses (see Section 2), may provide additional insights into the underlying mechanisms driving the modifications observed from answer0 to answer2.

Interview Context: A likely factor contributing to the high proportion of the formal variants is the interview setting in which we embedded the LLM speakers. We cannot exclude the possibility that the situational context, in this case, the interview format, could significantly influence the selection of more formal language structures. This problem goes beyond LLM prompting and is also occasionally mentioned in diachronic studies. Wagner and Sankoff (2011: 305), for example, raised the question whether shifts towards formal variants might be attributed to stylistic changes of interview situations. Similarly, Hekkel (2021: 83) suggests that an interview genre itself might be evolving. Furthermore, issues of comparability possibly extend to other variables, such as topics and interview partners (Hekkel 2021: 81–83). This highlights another consideration for follow-up studies. Interview situations could be described more thoroughly and contrasted with one another.

A Focus on Age: An additional aspect meriting further examination pertains to the prompt used for generating answer2. Our findings indicate that the injection of expert knowledge typically leads to an elevation in the proportions of the formal linguistic variants. This trend is consistent across all examined variables, including age, birth year, and interview year. While instructions were given to ChatGPT to consider the speaker's age in generating answer2, further prompt engineering to refine the incorporation of expert knowledge, particularly in relation to age, presents a promising avenue for enhancing the model's ability to emulate age-sensitive linguistic variation more accurately.

Implications for Sociolinguistics and LLM Research: The insights gained from our study hold considerable implications for the intersection of sociolinguistics and Large Language Model (LLM) research. The nuanced understanding of how ChatGPT responds to expert-knowledge injections and adapts its language use according to formal variants offers a critical perspective on the capabilities and limitations of current LLMs in replicating human-like linguistic variation. This research not only underlines the potential of LLMs in sociolinguistic studies but also highlights the necessity for more refined methodological approaches in future research. The tendency of ChatGPT towards hypercorrection and its varying response to age-related prompts leaves room for further exploration, particularly in how LLMs can be more effectively prompted to be more aware of age-sensitive variation. These findings pave the way for future studies to explore the complex dynamics of language generation in LLMs and their applications in understanding sociolinguistic phenomena.

6 Conclusion

The aim of this paper was to contribute to a roadmap for the integration of sociolinguistic inquiry into the domain of Large Language Model (LLM) research (see also Staab, Vero, Balunović and Vechev 2023; Feldman, Dant, Foulds and Pan 2022). Central to this exploration was the question of whether ChatGPT is capable of employing age-sensitive variants of first-person plural clitics and future tenses, particularly when prompted to produce responses from the perspectives of two speakers differentiated by their birth and interview years. As discussed in Section 2, both variables show sensitivity to sociological, grammatical, and register-related factors. In this paper, we have focused on age-related variation, while other potentially relevant sociolinguistic variables, such as education level or gender, are controlled in our experiment design.

On a general level, our analysis identified a prevailing inclination towards formal linguistic variants, which may be attributed to an inherent perception of contextual formality associated with interview settings. We observed rising proportions of the formal variants from answer0 to answer2 for both speakers. The observed shift in proportional use of the two linguistic variants from the initial naive answers (answer0) to the expert-knowledge informed answers (answer2) is, to some extent, attributable to an overall increase in the frequency of the targeted linguistic variable. This increase often involved hypercorrective tendencies, transforming first-person singular pronouns into their plural counterparts, or present tense to future tense.

In contrast to the influence exerted by the injection of the expert knowledge, our analysis did not reveal any systematic effect of factors such as age, birth year, or interview year on the proportion of the variants. Furthermore, the observed variation could not be conclusively linked to diachronic phenomena, such as age-grading or the Apparent-Time Hypothesis.

However, a notable anomaly was detected for speakers at the age of 70. In these instances, ChatGPT demonstrated a pronounced level of linguistic uncertainty in answer0 for both speakers and across both linguistic variables. This uncertainty converged towards values close to 1.0 in answer2, more so than for any other age group examined. While age 70 may be deemed close to the onset of old age, the specific reasons for this outcome warrant further investigation.

This study represents a foray into the complex interplay between sociolinguistic variables and LLMs, particularly in the context of age-sensitive language use. Future research could explore more nuanced approaches to integrating sociolinguistic variables, such as educational degree, into LLMs, perhaps through more advanced prompting techniques. Additionally, extending this research to include a broader range of sociolinguistic variables could yield deeper insights into the capabilities and adaptability of LLMs in mimicking human language across age groups and generations.

Acknowledgments

We acknowledge the use of OpenAI’s GPT-4-Turbo in assisting with the editing and refinement of this paper, enhancing both its clarity and presentation in English. However, we ensured that all contributions adhered strictly to the standards and ethical guidelines of academic writing.

Appendix

Prompts for automated annotation: Here you will find the two prompts for annotating our data. This corresponds to our guideline when reviewing the annotations manually after automatic annotation by ChatGPT.

Prompt for annotation of future tenses:
You are a student-assistant bot tasked with annotation language data for a research project in linguistics.
The research project is about the future tenses in French: ‘futur simple’ and ‘futur proche’.
Ignore all other tenses and modes.
You will get as input a text which is language data from a real speaker. In this text, you add the respective annotations after the occurrences (independently of whether they are uppercase or lowercase):
_fs_ for ‘futur simple’, the synthetic future, formed by an infinitive like form and the endings -ai, -as, -a, -ons, -ez, -ont, such as in ‘on fêtera’;
_fp_ for ‘futur proche’, the analytic future tense formed by the auxiliary ‘aller’ plus infinitive, such as in ‘on va fêter’ or ‘nous allons sûrement fêter’;

Ignore all other tenses and moods! Don’t annotate neither present tense (even if it is used to express a future action, such as ‘demain je vais chez vous’) nor the conditional (such as ‘je vivrais’)!
Examples:
on fêtera ton anniversaireà → on fêtera _fs_ ton anniversaire
il va être heureux → il va être _fp_ heureux
je fais → je fais
je pourrais faire → je porrais faire
je pourrai faire → je pourrai _fs_ faire
Output only the annotated text without further comments!

Prompt for annotation of first-person plural clitics:
You are a student-assistant bot tasked with annotation language data for a research project in linguistics.
The research project is about the pronouns that are used to express the first-person plural in French: ‘on’ and ‘nous’ as personal subject pronouns.
AVOID annotating object pronouns (‘il NOUS disait’), or reflexive pronouns (‘nous NOUS lavons’) or ANY stressed pronouns (‘chez nous’, ‘NOUS, nous’). IGNORE everything that is not a first-person plural subject clitic.
You will get as input a text which is language data from a real speaker. In this text, you must add right of the occurrences (independently of whether they are uppercase or lowercase) by the respective annotations:
_on1_ for ‘on’ ONLY IF used for on as subject **first-person plural clitic**, such as in ‘on a fêté’. For ‘l’on’ use ‘l’_on1_’;
nous for ‘nous’ ONLY IF used as subject **first-person plural citic**, such as ‘nous fêtons’;
_on3_ for cases in which the context suggests that the speaker uses an impersonal ‘on’ that does not carry the meaning of first-person plural. Use only if the 1st person plural interpretation can be excluded, for example ‘comme on pourrait l’imaginer’ or ‘on ne peut pas dire que’.
If the clitic appears in combination with a stressed pronoun, only annotate the clitic!
Ignore all other pronouns!
Examples:
On se trouvait → _on1_ se trouvait
On pourrait dire → _on3_pourrait dire
Nous disons → _nous_ disons
nous, on avait → nous, _on1_ avait
nous, nous avions → nous, _nous_ avions
nous nous entendions bien → _nous_ nous entendions bien
pour nous à pour nous
on faisait chez nous → _on1_ faisait chez nous
nous, les enfants, on fait → nous, les enfants, _on1_ fait
Output only the annotated text without further comments!

References

Abouda, Lotfi & Skrovec, Marie. 2015. Du rapport entre formes synthétique et analytique du futur. Étude de la variable modale dans un corpus oral micro-diachronique. *Revue de sémantique et pragmatique* 38. 35–57.

Adolphs, Leonard & Shuster, Kurt & Urbanek, Jack & Szlam, Arthur & Weston, Jason. 2021. *Reason first, then respond: Modular Generation for Knowledge-infused Dialogue*. arXiv: 2111.05204 [cs.CL].

Ashby, William J. 1991. When does variation indicate linguistic change in progress? *Journal of French Language Studies* 1. 1–19.

Bally, Charles. 1952. *Le langage et la vie*. 23rd ed. Kindle-Edition. Geneva: Droz.

Blanche-Benveniste, Claire. 1990. *Le français parlé: Études grammaticales*. Collection Sciences du langage. Editions du Centre national de la recherche scientifique. Paris: Presses du CNRS.

Blanche-Benveniste, Claire. 1997. *Approches de la langue parlée*. Paris: Ophrys.

Blondeau, Hélène. 2006. La trajectoire de l’emploi du futur chez une cohorte de Montréalais francophones entre 1971 et 1995. *Revue canadienne de Linguistique Appliquée* 9. 73–95.

Coveney, Aidan. 2000. Vestiges of ‘nous’ and the 1st Person Plural verb in informal Spoken French. *Language Sciences* 22(4). 447–481.

El Sherbiny Ismail, Eman & Gerstenberg, Annette & Lupica Spagnolo, Marta & Schulz, Friederike & Vandenbroucke, Anne. 2022. L’âge avancé en perspective longitudinale et ses outils: LangAge, un corpus au pluriel. *SHS Web Conf. (SHS Web of Conferences) - 8e Congrès Mondial de Linguistique Française* 138, 10003. 1–14.

Feldman, Philip & Dant, Aaron & Foulds, James R. & Pan, Shemei. 2022. *Polling Latent Opinions: A Method for Computational Sociolinguistics Using Transformer Language Models*. arXiv: 2204.07483 [cs.CL].

Gerstenberg, Annette. 2011. *Generation und Sprachprofile im höheren Lebensalter. Untersuchungen zum Französischen auf der Basis eines Korpus biographischer Interviews*. Analecta Romanica, vol. 76. Frankfurt am Main: Vittorio Klostermann.

Harrell, Frank E. Jr. 2023. *Hmisc: Harrell Miscellaneous*. R package version 5.1-2, https://hbiostat.org/R/Hmisc/.

Hekkel, Valerie. 2021. *Eine soziolinguistische Betrachtung von* parce que-*Strukturen in Synchronie und Diachronie*. PhD thesis. University of Potsdam.

Hockett, Charles F. 1950. Age-grading and linguistic contiguity. *Language* 26. 449–459.

Hothorn, Torsten & Hornik, Kurt & Zeileis, Achim. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3). 651–674.

Hunter, John D. 2007. Matplotlib: A 2D Graphics environment. *Computing in Science & Engineering* 9(3). 90–95.

King, Ruth & Nadasdi, Terry. 2003. Back to the future in Acadian French. *Journal* *of French Language Studies* 13. 323–337.

Kuhn, Max. 2008. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* 28(5). 1–26.

Laberge, Suzanne. 1977. *Étude de la variation des pronoms sujets définis et indéfinis dans le français parlé à Montréal*. PhD thesis. University of Montréal.

Labov, William. 1963. The social motivation of a sound change. *Word* 19. 273–309.

Labov, William. 1966 [2006]. *The Social Stratification of English in New York City*. Cambridge: Cambridge University Press.

Labov, William. 1978. On the use of the present to explain the past. In Baldi, Philip & Werth, Ronald (eds), *Readings in Historical Phonology*, 275–312. Pennsylvania: State University Press.

Levshina, Natalia. 2015. *How to do Linguistics with R: Data Exploration and Statistical Analysis.* Amsterdam: John Benjamins.

Levshina, Natalia. 2021. Conditional Inference Trees and Random Forests. In Paquot, Magali & Gries, Stefan Th. (eds), *A Practical Handbook of Corpus Linguistics*, 611–643. Cham: Springer.

Liaw, Andy & Wiener, Matthew. 2002. Classification and Regression by randomForest. *R News* 2(3). 18–22.

Markl, Nina. 2022. Language variation and algorithmic bias: Understanding algorithmic bias in British English Automatic Speech Recognition. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '22. Seoul, Republic of Korea: Association for Computing Machinery. 521–534.

McKinney, Wes. 2010. Data structures for statistical computing in Python. In van der Walt, Stefan & Millman, Jarrod (eds), *Proceedings of the 9th Python in Science Conference*, 56–61.

OpenAI. 2023. *GPT-4 Turbo Model* (gpt-4-1106-preview). https://www.openai.com (last accessed 7 May 2024).

Ostapenko, Alissa & Wintner, Shuly & Fricke, Melinda & Tsvetkov, Yulia. 2022. *Speaker information can guide models to better inductive biases: A case study on predicting Code-Switching*. arXiv.2203.08979 [cs.CL].

Paoli, Sandra & Wolfe, Sam. 2022. The GO-future and GO-past periphrases in Gallo-Romance: A comparative investigation. In Ledgeway, Adam & Smith, John Charles & Vincent, Nigel (eds), *Periphrasis and Inflexion in Diachrony: A View from Romance*, 123–144. Oxford: Oxford University Press.

Poplack, Shana & Turpin, Danielle. 1999. Does the FUTUR have a future in (Canadian) French? *Probus* 11(1). 133–164.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Rebotier, Aude. 2015. Le futur périphrastique français avec *aller*: un renvoi spécifique l'avenir ou un temps en voie de grammaticalisation ? Une approche contrastive. *Revue sémantique et Pragmatique* 38. 11–34.

Roberts, Nicholas S. 2012. Future temporal reference in Hexagonal French.

University of Pennsylvania Working Papers in Linguistics. Selected Papers from NWAV 40 18(2). 97–106.

Salewski, Leonard & Alaniz, Stephan & Rio-Torto, Isabel & Schulz, Eric & Akata, Zeynep. 2023. *In-Context Impersonation Reveals Large Language Models’ Strengths and Biases*. [Preprint] arXiv:2305.14930 [cs.AI].

Sankoff, Gillian. 2005. Cross-sectional and longitudinal studies. In Ammon, Ulrich & Dittmar, Norbert & Mattheier, Klaus J. & Trudgill, Peter (eds), *An International Handbook of the Science of Language and Society*, 1003–1013. Berlin & New York: De Gruyter Mouton.

Sankoff, Gillian & Wagner, Suzanne Evans. 2020. The long tail of language change: A trend and panel study of Québécois French futures. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 65. 246–275.

Serpollet, Noëlle & Bergounioux, Gabriel & Chesneau, Annie & Walter, Richard. 2007. A Large Reference Corpus for Spoken French: ESLO1 and 2 and its Variations. *Proceedings from Corpus Linguistics Conference Series*. University of Birmingham.

Staab, Robin & Vero, Mark & Balunović, Mislav & Vechev, Martin. 2023. Beyond memorization: Violating privacy via inference with Large Language Models. arXiv: .2310.07298 [cs.AI].

Söll, Ludwig. 1969: Zur Situierung von *on* ‘nous’ im neuen Französisch. *Romanische Forschungen* 81. 535–549.

Söll, Ludwig. 1974 [1980]. *Gesprochenes und geschriebenes Französisch*. Berlin: Schmidt.

Tagliamonte, Sali A. & Baayen, R. Harald. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change*, 24(2). 135–178.

The pandas development team. 2020. *pandas-dev/pandas: Pandas*. Zenodo. Available at: https://doi.org/10.5281/zenodo.8092754

Wagner, Suzanne Evans. 2012. Age Grading in Sociolinguistic Theory. *Language and Linguistics Compass* 6(6). 371–382.

Wagner, Suzanne Evans & Sankoff, Gillian. 2011. Age grading in the Montréal French inflected future. *Language Variation and Change* 23(3). 275–313. https://compass.onlinelibrary.wiley.com/doi/abs/10.1002/lnc3.343

Wang, Jianing & Wang, Chengyu & Tan, Chuanqi & Huang, Jun & Gao, Ming. 2023. *Boosting In-Context Learning with Factual Knowledge*. arXiv: 2309.14771 [cs.CL].

Waskom, Michael L. 2021. seaborn: statistical data visualization. *Journal of Open Source Software* 6(60). 3021.

Weinrich, Harald.1989. *Grammaire textuelle du français*. Paris: Didier.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.

Wickham, Hadley & François, Romain & Henry, Lionel & Müller, Kirill & Vaughan, Davis. 2023. *dplyr: A Grammar of Data Manipulation. R package version 1.1.4*, https://github.com/tidyverse/dplyr, https://dplyr.tidyverse.org.

Xu, Benfeng & Yang, An & Lin, Junyang & Wang, Quan & Zhang, Yongdong & Mao, Zhendong. 2023. *ExpertPrompting: Instructing Large Language Models to be Distinguished Experts.* arXiv: 2305.14688 [cs.CL].

Zimmer, Dagmar. 1994. ‘Ça va tu marcher, ça marchera tu pas, je le sais pas.’ Le futur simple et le futur périphrastique dans le français parlé à Montréal. *Langues et linguistique* 20. 213–226.