

A variationist analysis of two French attitude expressions: *je pense* and *je crois*

Delin Deng
University of Florida
ddeng@ufl.edu

Abstract

Previous research has reported that *je pense* and *je crois* are interchangeable when used as attitude expressions (see, for example, Gosselin, 2015, 2018; Rendulić & Kanaan-Caillol, 2016; Angot, 2021). This paper conducts a variationist analysis of a corpus collected in Orléans, France, in 2008 to examine the variable use of *je pense* and *je crois* as attitude expressions among 10 French native speakers. The results of the regression analysis indicate that all three tested linguistic factors are significant, while no extralinguistic factors show significance in relation to this variable use. On one hand, we confirm that the variable use of *je pense* and *je crois* is not a change in progress in the apparent time. On the other hand, our analysis provides further insights into how the use of these variables is not only conditioned by semantic or pragmatic differences but also by the linguistic contexts in which the particles are used.

1 Introduction

There is a substantial body of literature on the French attitude expressions or personal opinion expressions *je pense* (I think) and *je crois* (I believe) (see, for example, Gosselin, 2015, 2018; Rendulić & Kanaan-Caillol, 2016; Angot, 2021). Various terms have been used to refer to them in the past decades.

Simons (2007: 1034) referred to verbs such as *see*, *hear*, *think*, *believe*, *discover*, and *know*, as clause-embedding verbs. According to him, “the

embedded clause carries the main point of the utterance, while the main clause serves some discourse function”. Gachet (2014: 147) posited, “the initial phrase (*je crois* ‘I believe’) is a peripheral clause, and the following clause is the main clause.” Gosselin (2018: 180) referred to them as “verbal expressions of personal opinion”. He argued that, instead of taking the traditional unitary point of view, each of them has “a specific meaning” and “are not always substitutable for one another”. Angot and Hansen (2021) analyzed *je pense*, emphasizing that it is a similar construction to *je crois* as pragmatic markers, which “fulfil both interpersonal, face-related functions and discourse-organizational functions” (Angot & Hansen, 2021: 1).

In this article, we aim to explore, using a quantitative method, the linguistic and extralinguistic factors that influence the choice between *je pense* and *je crois* in French native speech. The current study seeks to address the following questions: If we consider *je pense* and *je crois* to be functionally interchangeable as attitude verbs introducing an embedded clause, what sets them apart in terms of the linguistic environments in which they appear within an utterance? Is the choice between the two forms also conditioned by certain extralinguistic factors in native speech? Is it an ongoing change or a completed change?

Therefore, this article will be structured as follows: first, the long-stand puzzled set by attitude expressions will be reviewed and discussed. Second, methodology of the current work, including information on the corpus used, speakers, tokens, linguistic and extralinguistic factors to be examined

in this article as well as the statistical analysis, will be presented. Following this, the statistical results will be tabulated and discussed in detail. Lastly, a conclusion of the current work, its limitation as well as future implications will be laid out.

2 Puzzle

Attitude expressions, such as *je pense* and *je crois*, attracted so much attention as they challenge our existing understanding of the relations between propositions and their truth-values. Let us consider the following scenario:

I just arrived at the company and settled down at my desk. I have not chatted with anyone yet. At that time, John entered my office and asked me if Marie had already arrived at the office. Usually, Marie arrives at the office earlier than H el ene. So, I responded:

(1) Si H el ene est dans le bureau, je crois que Marie est d ej a arriv ee.

‘If H el ene is in the office, I believe that Marie has already arrived.’

(2) H el ene est dans le bureau.

‘H el ene is in the office.’

(3) Je crois que Marie est d ej a arriv ee.

‘I believe that Marie has already arrived.’

If we define:

p= H el ene est dans le bureau. ‘H el ene is in the office.’

q= Je crois que Marie est d ej a arriv ee. ‘I believe that Marie has already arrived.’

My sentence could be represented as:

$p \rightarrow q$

By modus ponens, if $p \rightarrow q$ is true, and p is true, then q must be true.

However, this poses some problems. ‘A person believes p’ is true only if ‘that person knows p’ is true. Nevertheless, given the scenario we provided, if I had just arrived at my office and had not chatted with anyone else, it would not be clear how I could know whether Marie had arrived. Therefore, it seems that we encounter a situation where modus ponens has led us from true premises to a false conclusion. So, how do we address this puzzle?

Some earlier studies proposed that expressions such as *je crois* play the role of a mitigator, which attenuates the certainty of a statement (See, for example, Benveniste, 1996; Borillo, 1982; Vet, 1994). Following this line of reasoning, our sentence *Je crois que Marie est d ej a arriv ee* ‘I believe that Marie has already arrived’ could be understood as indicating some degree of uncertainty. If this is indeed the case, then consider the following sentences:

(4) Patrick cro it que Marie est d ej a arriv ee.

‘Patrick believes that Marie has already arrived.’

(5) Patrick est convaincu que Marie est d ej a arriv ee.

‘Patrick is convinced that Marie has already arrived.’

Does (4) imply that Patrick is not sure of it? Clearly not. We would not deny that (4) would be compatible with (5). (4) implies that Patrick is convinced that Marie has already arrived. But why (3) implies that I am not sure of *Marie est d ej a arriv ee* ‘Marie has already arrived.’, but (4) implies that he is sure of *Marie est d ej a arriv ee* ‘Marie has already arrived.’? Why do they seem to be in contradiction?

Gosselin (2018: 182) highlighted that *croire* ‘believe’ “is a verb that only indicates that the speaker of the utterance does not presuppose the content of the complement”. He then argued that this could be explained by the *logic of conviction* proposed by Lenzen (2004). When a person believes p (p stands for any proposition), what that person really believes is not p itself but rather knowing p. In other words, this representation should not be expressed as B(a, p) but as B(a, K(a, p)) (a: person; p: proposition; B: believe; K: know). As Lenzen pointed out, “knowledge and conviction are subjectively indiscriminable in the sense that person a cannot tell apart whether she is ‘only’ convinced that p or whether she really knows that p” (Lenzen, 2004: 973). The same would apply to *je pense*. When a person ‘thinks’ p, they think that they know p.

So where do ‘he is sure of p’ and ‘I am not sure of p’ come from, as in (3) and (4)? As reasoned by Gosselin (2018: 183), “if a speaker uses non-factive epistemic expressions, like *je crois/suis certain(e)/persuadé(e)/convaincu(e) que*, ‘I believe/am sure/persuaded/convinced that’, it triggers an implicature from the utterance. The interpreter will think that if the speaker has used not just *p* or *je sais que p*, it is because she does not believe that she knows that *p* and therefore she is not really convinced that *p*, hence the systematic mitigation effect, which may seem contradictory to what the statement says literally.”

Now we have resolved our puzzle. However, another question arises: it appears that in oral French, native speakers use *je pense (que)* and *je crois (que)* in a quasi-interchangeable manner. Numerous previous studies have qualitatively discussed the semantic or pragmatic differences between the two. Is there any quantitative evidence that can shed light on their differences? Do native speakers tend to prefer one over the other in specific circumstances? If so, what are these circumstances? These are the questions that we aim to address with the current work.

3 Methodology

3.1 Corpus and data

The corpus we will use for the current study is ESLO 2 (Enquêtes Sociolinguistiques à Orléans: <http://eslo.huma-num.fr/index.php>, Baude and Dugua, 2011). It is an online corpus comprising sociolinguistic interviews with native speakers of French in Orléans, a city located approximately 120 km south of Paris. The variety of French spoken in Orléans is closer to the central French variety, which is considered to be accentless and closer to standard French. The ESLO 2 corpus was initiated in 2008 and is still under development. It includes various modules, ranging from interviews to questionnaires. For the current study, we will solely use the interview module, which consists of 81 interviews conducted in French. All interviews were

transcribed using Transcriber (Barras et al., 2001) and can be downloaded from the website. We imported all Transcriber files into Elan (2021) to identify the relevant occurrences and their surrounding linguistic environments for our final analysis.

3.2 Speakers

For this study, we randomly selected 10 speakers in ESLO 2 (5 females and 5 males). Table 1 provides detailed information on these 10 speakers, including their assigned ID (represented by two letters followed by one or two digits), gender, age at the time of the interview, and socioeconomic status (SES).

Speakers	Gender	Age	SES
QF28	m	58	high
MC59	m	81	low
GK11	m	31	high
BV1	m	23	low
BT17	m	28	middle
LX10	f	65	low
KC3	f	23	low
HT398	f	33	high
AN43	f	39	high
AJ38	f	21	middle

Table 1: Detailed information on 10 speakers.

From Table 1, we observe that the age range of the selected speakers is relatively representative, encompassing the younger, middle-aged, and older generations. Regarding SES, we categorized the speakers into three main groups based on the information provided in their ESLO profile: low (including blue-collar workers, manual workers, and the unemployed), middle (comprising technicians, supervisors, white-collar and office workers), and high (consisting of businesspeople, educated professionals, and intellectual workers). As indicated in Table 1, the distribution of SES among the 10 speakers is relatively balanced, thereby minimizing the potential for SES bias in our final statistical analysis.

3.3 Tokens

In total, we identified 190 occurrences of *je pense* and *je crois* (114 occurrences of *je pense* and 76 occurrences of *je crois*) in ESLO 2. However, the following cases are excluded from our final analysis:

1) Occurrences that appear in negation:

ex. 1 : donc y a pas je crois pas qu'y a un langage jeune euh orléanais

'so there is not I do not think that there is a youth language uh Orleanese'

ex. 2 : mais je pense pas euh

'but I do not think uh'

2) Occurrences that do not introduce an embedded clause:

ex 3 : euh sur Orléans je crois euh des dans le l'ha- l'habillement dans le dans le vêtement quoi les

'um in Orleans I believe um some in the clo- the clothing in the in the clothes what the'

Therefore, only 164 occurrences are included in our final analysis. Table 2 presents the detailed distribution of *je pense* and *je crois* in this study.

	<i>je pense</i>	<i>je crois</i>	Total
No. included	100	64	164
No. excluded	14	12	26
Total	114	76	190

Table 2: Distribution of *je pense* and *je crois*.

3.4 Linguistic factors

Table 3 presents the linguistic factors that might be relevant to choosing *je pense* and *je crois*. For each factor, we have at least two different levels (groups) to look at.

Factor	Levels
tense of the verb in the embedded clause	present
	future
	imperfect
	perfect
	pluperfect
status of the embedded clause	conditional
	judgment of reality
	judgment of the value

presence of <i>que</i>	present
	absent

Table 3: Linguistic factors to be examined.

As shown in Table 3, firstly, we will examine the tense of the verb in the embedded clause, while considering *je pense* and *je crois* as part of the matrix clause. We aim to determine if either form is more closely associated with a particular tense. Secondly, we will investigate the status of the embedded clause. In this regard, we will adopt the classification proposed by Gosselin (2018), which distinguishes between judgments of reality and judgments of value. As Gosselin (2018: 180) explains, "a judgment of reality states what the case is (it describes a situation), while a value of judgment consists of speaking well or ill of an individual or situation". Thirdly, we will analyze the presence of *que* before the embedded clause. While *que* is obligatory in written French when introducing the embedded clause, it is optional in oral French. With this factor, we aim to determine if either form shows a preference for the omission or retention of the particle *que*.

3.5 Extralinguistic factors

Table 4 presents the two extralinguistic factors to be examined in this study. For SES, we utilized the information provided by the corpus and classified the speakers into three groups: low, mid, and high SES. Regarding the age factor, we considered the age of the speaker at the time of the interview. Given the limited number of speakers (ten), we did not group them into different age categories. Instead, the age factor will be treated as a continuous variable for statistical purposes.

Factor	Levels
SES	low
	mid
	high
Age	continuous

Table 4: Extralinguistic factors to be examined.

3.6 Statistical analysis

For this article, we will use the mixed-effects regression model carried out in the R environment using Rbrul (Johnson, 2009). The model distinguishes the following levels for statistical significance: $p > 0.1$, not significant; $.05 < p < 0.1$, marginally significant; $p < .05$, significant; $p < .01$, very significant; $p < .001$, highly significant. For the results, the model provides one p -value for each predictor (the independent variable) to indicate if this predictor is statistically significant for the dependent variable. Meanwhile, it also provides the factor weight and log odds for each level of the predictor to indicate which level(s) favors/disfavors the chosen variable.

For our analysis, we look at the dependent variable, the attitude expression, at binary classifications of *je pense* vs. *je crois*. The fixed independent variables are both the linguistic and extralinguistic factors presented above. All fixed factors except for age are categorical. As we use the age of the speakers at the time of the interview, the age factor is thus continuous. To include the mixed-effects, we use participants as a random variable. For the modeling, we performed the one-level test.

Since “participants” is treated as a random variable, for the following section, we only provide the results for the fixed variables for further discussion. However, the detailed results for individual participants are provided in the appendix for readers’ reference.

4 Results & discussion

Table 5 presents the regression analysis results of *je pense/je crois*. Our results indicated that all three linguistic factors, the status of the embedded clause ($p=0.000733$; f.w.: value judgement: 0.752; judgement of reality: 0.248), tense of the verb in the embedded clause ($p=0.00231$; f.w.: conditional: >0.999 ; future: 0.792; present: 0.584; perfect: 0.51; imperfect: 0.341; plusperfect: <0.001) and presence of *que* ($p=0.0489$; f.w.: present: 0.598; absent: 0.402), are statistically significant for the choice

between *je pense* and *je crois*, while no social factors have been found to be statistically significant.

Among the three linguistic factors, the status of the embedded clause appears to be the most influential factor contributing to the choice between *je pense* and *je crois*. When the embedded clause represents a judgment of value, French native speakers are more inclined to use the form *je pense*, whereas when the embedded clause represents a judgment of reality, they are more likely to opt for *je crois*.

Regarding the tense of the verbs in the embedded clause, our results indicated that the conditional, future, and present tenses tend to favor the use of *je pense*, while the perfect, imperfect, and pluperfect tenses tend to favor the use of *je crois*. In other words, *je pense* is more commonly associated with present and future tenses, while *je crois* is more commonly associated with perfect tenses. *Je pense* is more likely to be used when referring to ongoing or future events, whereas *je crois* is more likely to be used when referring to past events. It is also noteworthy that the conditional tense is exclusively used with *je pense*, while the pluperfect tense is never used with this form.

Finally, regarding the presence of the particle *que* following these two attitude expressions, our results indicate that *que* is more likely to be present when the variant *je pense* is used, and more likely to be omitted when the form *je crois* is employed. In other words, native speakers of French tend to prefer using *je pense que* over *je crois que*.

<i>Je pense/Je crois</i>				
Input prob.	0.61			
Total no.	164			
Log. likelihood	-84.865			
	logodds	tokens	%	f.w.
Status	$p=0.000733$			
value	1.11	28	92.9	0.752
reality	-1.11	136	54.4	0.248
Tense	$p=0.00231$			
conditional	16.908	11	100.0	>0.999

future	1.339	15	86.7	0.792
present	0.341	96	61.5	0.584
perfect	0.038	20	50.0	0.51
imperfect	-0.661	19	36.8	0.341
pluperfect	-17.966	3	0.0	<0.001
Que	p=0.0489			
present	0.397	104	70.2	0.598
absent	-0.397	60	45.0	0.402
SES	Not significant			
High	0.328	65	69.2	[0.581]
Low	0.251	68	63.2	[0.562]
Middle	-0.579	31	38.7	[0.359]
Age	Not significant			
continuous				
+1	-0.005			
Speakers	Random			

Table 5: Regression analysis results of *je pense/je crois*.

While linguistic environment appears to be particularly influential in the choice between *je pense* and *je crois* for native speakers, none of the tested extralinguistic factors tested has been found to be significant for this choice. The use of either form is associated with specific age groups or SES groups, indicating that this variation between the two variants is not an ongoing change in contemporary French. Instead, it represents a completed change over time. The use of both forms has become widespread across all social classes in French.

5 Conclusion

In this study, in contrast to earlier qualitative studies, we conducted a mixed-effects regression analysis on data obtained from a corpus of oral French speech by native speakers. Our aim was to quantitatively examine two attitude expressions, *je pense* and *je crois*. We considered both linguistic and extralinguistic factors in our analysis. The results revealed that all three linguistic factors, namely the tense of the verb in the embedded clause, the status of the embedded clause, and the presence of the particle *que*, were found to be statistically significant in relation to the choice between the two

variants. These findings provide further evidence that, in addition to semantic and pragmatic differences, the linguistic context also plays a role in determining which variant speakers will choose. On the other hand, none of the extralinguistic factors were found to be significant in this choice, suggesting that the variable use of these expressions is not an ongoing change but rather a relatively stable feature of native speech.

However, it should be noted that the current study has limitations due to the small number of tokens analyzed, which means that the findings are not conclusive. Instead, this work can be seen as a pilot study, providing a starting point for further investigation.

In future studies, expanding the dataset by adding more data would be beneficial to conduct a more comprehensive analysis of the real-time use of *je pense* and *je crois* in French native speech. The ESLO corpus consists of two parts, ESLO 1 (1968-1974) and ESLO 2 (2008-), and the time interval between these two collections allows for a comparison of the changes in the use of these variables over a forty-year period. This comparative analysis would help determine if the observed variation is indeed a completed change in contemporary French, as well as shed light on any potential changes in the linguistic factors influencing the choice between the two variables over time.

Second, in our study, we only examined two social factors. With a larger dataset, it would be possible to incorporate additional factors, such as educational background or social network, to explore potential intergroup differences in the use of *je pense* and *je crois*.

Third, since our analysis was based on the interview module of the ESLO corpus, which predominantly represents informal speech, it would be valuable to investigate the use of *je pense* and *je crois* in other contexts, such as lectures, conferences, or casual conversations among family members. This would allow us to explore potential

variations in usage across different communicative settings.

Lastly, it would be beneficial to examine the use of *je pense* and *je crois* in the speech of individuals from Francophone countries other than France. Given that particles may undergo different stages of pragmaticalization in different regions, it is likely that usage patterns vary. Comparing the ESLO corpus with other corpora from diverse Francophone contexts would provide us with a broader understanding of this phenomenon.

References

- Angot, J. Epistemic and subjective expressions in French: the case of *je pense*, *je crois* and *je trouve*. Doctoral dissertation (The University of Manchester, 2021).
- Angot, J. and M. M. Hansen. “The meaning and functions of French *je pense* (que): a Constructionalist and interactional account.” In *Studies at the grammar-discourse interface*. (John Benjamins Publishing Company, 2021).
- Barras, C., E. Geoffrois, Z. Wu and M. Liberman. “Transcriber: development and use of a tool for assisting speech corpora production.” *Speech Communication*, 33(1-2) (2001), 5-22.
- Baude, O. and C. Dugua. “(Re) faire le corpus d’Orléans quarante ans après: quoi de neuf, linguiste?”, *Corpus*, (10) (2011), 99-118.
- Benveniste, É. *Problèmes de linguistiques générales I* (Paris: Gallimard, 1996).
- Borillo, A. “Deux aspects de la modalisation assertive: *croire* et *savoir*.” *Cahiers de grammaire*, 4 (1982), 5-38.
- ELAN (Version 6.2) [Computer software]. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. (2021). Retrieved from: <https://archive.mpi.nl/tla/elan>
- Gachet, F. “Syntactic hypotheses about so-called ‘que-deletion’ in French.” In *Parenthesis and Ellipsis* (De Gruyter Mouton, 2014), 147-172.
- Gosselin, L. “L’expression de l’opinion personnelle: *Je crois/pense/trouve/considère/estime que p.*” *L’information grammaticale*, 144 (2015), 34-40.
- Gosselin, L. “French expressions of personal opinion: *je crois/ pense/ trouve/ estime/ considère que p.*” In *Epistemic Modalities and Evidentiality in Cross-Linguistic Perspective* (De Gruyter Mouton, 2018), 179-195.
- Hay, J. “Statistical analysis.” In Marianna Di Paolo/Malcah Yaeger-Dror (eds.), *Sociophonetics: a student’s guide* (Abingdon: Routledge, 2011), 198-214.
- Johnson, D. E. “Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis.” *Language and linguistics compass*, 3(1) (2009), 359-383.
- Lenzen, W. “Epistemic logic.” In *Handbook of Epistemology*. (Springer, Dordrecht., 2004), 963-983.
- Mooney, D. “Quantitative approaches for modelling variation and change: a case study of sociophonetic data from Occitan.” In *Manual of Romance Sociolinguistics*. (De Gruyter, 2018).
- Rendulić, N. and L. Kanaan-Caillol. “*Je crois que, je pense que*: valeurs et variation dans un corpus oral diachronique.” In *SHS Web of Conferences* (EDP Sciences, 2016), 27, 02014.
- Simons, M. “Observations on embedding verbs, evidentiality, and presupposition.” *Lingua*, 117(6) (2007), 1034-1056.
- Vet, C. “Savoir et Croire.” *Langue Française*, 102 (1994), 56-68.

A Appendix. Individual difference.

Participants (random)				
	intercept	tokens	p/p+c	f.w.
std dev	0.276	164	0.61	...
LX10	0.156	22	0.727	0.539
HT398	0.143	14	0.786	0.536
GK11	0.109	28	0.821	0.527
BT17	0.106	5	0.6	0.527
BV1	-0.003	18	0.667	0.499
MC59	-0.035	9	0.444	0.491
AN43	-0.072	8	0.5	0.482
AJ38	-0.104	26	0.346	0.474
KC3	-0.12	19	0.579	0.47
QF28	-0.186	15	0.467	0.454