Simplification Strategies in French Spontaneous Speech

Ormaechea Grijalba, Lucía; Tsourakis, Nikolaos; Schwab, Didier; Bouillon, Pierrette; Lecouteux, Benjamin

Simplification Strategies in French Spontaneous Speech

Lucía Ormaechea[1,2], Nikos Tsourakis[1], Didier Schwab[2], Pierrette Bouillon[1] and Benjamin Lecouteux[2]

[1] TIM/FTI, University of Geneva, 40 Boulevard du Pont-d'Arve – Geneva, Switzerland

{firstName.lastName}@unige.ch

[2] Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG – Grenoble, France

{firstName.lastName}@univ-grenoble-alpes.fr

Abstract

Automatic Text Simplification (ATS) aims at rewriting texts into simpler variants while preserving their original meaning, so they can be more easily understood by different audiences. While ATS has been widely used for written texts, its application to spoken language remains unexplored, even if it is not exempt from difficulty. This study aims to characterize the edit operations performed in order to simplify French transcripts for non-native speakers. To do so, we relied on a data sample randomly extracted from the ORFÉO-CEFC French spontaneous speech dataset. In the absence of guidelines to direct this process, we adopted an intuitive simplification approach, so as to investigate the crafted simplifications based on expert linguists' criteria, and to compare them with those produced by a generative AI (namely, ChatGPT). The results, analyzed quantitatively and qualitatively, reveal that the most common edits are deletions, and affect oral production aspects, like restarts or hesitations. Consequently, candidate simplifications are typically register-standardized sentences that solely include the propositional content of the input. The study also examines the alignment between human- and machine-based simplifications, revealing a moderate level of agreement, and highlighting the subjective nature of the task. The findings contribute to understanding the intricacies of simplifying spontaneous spoken language. In addition, the provision of a small-scale parallel dataset derived from such expert simplifications, PROPICTO-ORFÉO-SIMPLE, can facilitate the evaluation of speech simplification solutions.

Keywords: simplification, spontaneous speech, French language, expert annotation, ChatGPT

1. Introduction

Automatic Text Simplification (ATS) aims at rewriting texts into simpler variants, by reducing their linguistic complexity, albeit preserving their original meaning (Candido et al., 2009; Horn et al., 2014). ATS has received increased attention in the past few years, in view of its significance from both societal and computational perspectives: it can assist in creating adapted texts for diverse target audiences (De Belder and Moens, 2010; Rello et al., 2013) or serve as a pre-processing step for other NLP tasks such as MT (Stajner and Popovic, 2016).

Providing a simplified version of a given text has typically been applied for newswire content (Xu et al., 2015; Saggion, 2017), healthcare-related documents (Shardlow and Nawaz, 2019; Van den Bercken et al., 2019) and Wiki-based articles (Hwang et al., 2015; Zhang and Lapata, 2017; Ormaechea and Tsourakis, 2023). Hence, ATS demonstrates its predominant application in written-based texts, while its implementation over a spoken modality remains unexplored. Yet, spoken-based texts are not exempt from difficulty.

Traditionally, features associated with complexity are strongly linked to the typical attributes found in formal written-based texts, like high lexical density or the propensity towards long subordination (Brunato et al., 2022). However, complexity also exists in spoken language, but is reflected differently from its written counterpart, mainly because the information structure is also dissimilar. Written text is the *result* of a planned language production, whereas speech, especially the spontaneous kind, is a real-time *process* (Carter and McCarthy, 2017), thus retaining traces of its on-the-fly construction like revisions, false starts, reformulations or self-corrections. Due to these phenomena, spoken language is typically disfluent, which makes speech transcripts particularly challenging to understand. Decomplexifying speech may be of particular interest when transcriptions are further used for:

- Accessibility purposes. A simplified transcript can help clarify the conveyed message and reduce its ambiguity, making it more accessible to several target audiences (*e.g.*, individuals with cognitive disabilities, foreign language learners, non-native speakers, etc).
- Ancillary purposes. Raw transcripts are often difficult to process by NLP pipelines. Providing a meaning-preserving simpler transcript may be helpful as an intermediate representation for other NLP tasks like subtitle translation (Mehta et al., 2020) or speech-to-pictograph cross-modal conversion (Ormaechea et al., 2023a).

With this article, we aim to investigate the strategies followed by experts as to simplify spontaneous French transcripts for a non-native speaking audience, and to compare the resulting simplification

Proceedings of the Workshop on DeTermIt! Evaluating Text Difficulty in a Multilingual Context @ LREC-COLING 2024, pages 90–102

operations with those produced by a generative AI, namely ChatGPT. More precisely, we aim to address the following research questions:

- What are the edit operations performed to obtain a simplified version of a French spontaneous speech transcript?
- How do human simplification strategies align with those adopted by ChatGPT and how suitable are they for a non-native audience?

In this way, we intend to provide an *a posteriori* characterization of the simplification strategies operated on the basis of a spoken spontaneous input. To the best of our knowledge, no such study has been conducted to date. In the absence of guidelines to direct this process, we decided to adopt an *intuitive* simplification approach (Allen, 2009). In this way, we investigate the simplifications produced based on expert linguists' criteria, and then compare them with those generated by ChatGPT.

The structure of this paper is as follows: Section 2 delves into the notion of spontaneity and describes the existing tasks that closely resemble spontaneous speech simplification. In Section 3, we discuss the input data sample, along with the survey design and ChatGPT prompts employed to collect simplifications. The analysis of these outputs, both quantitative and qualitative, is detailed in Section 4. Lastly, Section 5 provides concluding remarks, addresses limitations, and outlines potential pathways for future research.

2. Background and Related Work

2.1. Spontaneity in Speech

Spoken language exhibits differences with respect to written language which go beyond the mode of transmission used by each modality[1], and affect morphology, syntax and vocabulary (Caines et al., 2017). Among other aspects, a defining morpho-syntactic trait of speech is the lack of sentence boundaries, which are conventionally delimited in writing. From a grammatical perspective, spoken language is often characterized by the presence of disfluencies, which emerge as a result of the speaker's real-time processing and notably impact spontaneous speech. Due precisely to this *online* process, the *information packaging* (Halliday, 1985) also differs with respect to the *offline* (namely, written) one. This leads to the selection of different grammatical forms and changes in word order, which, in the case of French, are evidenced by the presence of cleft constructions (*i.e.*, *c'est lui qui a fermé la porte*) or the use of dislocated subjects (*i.e.*, *les enfants, ils arrivent*).

[1] That is, phonetics and prosody of speech versus graphemics and orthography of written language.

Unspontaneous texts constitute a revised and finalized version of a language production. Spontaneous speech, on the other side, is by nature an unfinished product. Due to the absence of prior planning, discourse unfolds in real time, consequently shedding all the traces of its elaboration, such as hesitations, reformulations, repetitions, and false starts (Blanche-Benveniste, 1997). These, unlike their written equivalent, are indelible in an oral modality, and can only lead to an elongation of the utterance (Bazillon et al., 2008). Consequently, the presence of such performance phenomena can produce concatenations of elements having a paradigmatic relation along the syntagmatic axis (Luzzati, 1998). This is evident in the spontaneous utterance illustrated in Figure 1, where disfluent features potentially hinder the correct understanding of the transcript. A simple *despontaneification* operation (see Figure 2) would result in an utterance holding an identical propositional content, and would clarify the conveyed message by eliminating paradigmatic supplements (*i.e.*, *[on va juste] euh [je vais juste]*) that stem from hesitations during the act of speaking.

2.2. Simplification and Compression in Speech

From an automated perspective, implementing simplification operations over speech appears to be an unexplored area. The existing task bearing the closest resemblance is *sentence compression* from speech transcripts (Angerbauer et al., 2019; Buet and Yvon, 2021). The aim of this process is to automatically reduce its length, generally in response to technical imperatives (Daelemans et al., 2004). This explains its relevance for subtitle generation (Luotolahti and Ginter, 2015), where technical restrictions drive the need of shrinking the text displayed on the screen. This is also triggered by the significantly faster pace of speech compared to reading, often motivating the suppression of phatic and deictic elements, as well as the condensation of information (Becquemont, 1996). Yet, the notion of *compression* must be distinguished from that of *simplification*. While the former aims at content reduction and merely preserves the most salient information, the latter seeks to generate a simpler variant without compromising the meaning.

In addition, an analogous task to speech simplification is Easy-to-Understand subtitling, in which an intralingual adaptation of subtitles is crafted to make them more accessible for viewers (Matamala, 2022). Guidelines have been proposed for this goal. While they include grammar- and style-based recommendations for simplification (Bernabé and Cavallo, 2021), they are primarily driven by the inherent spatial and temporal constraints of subtitling.

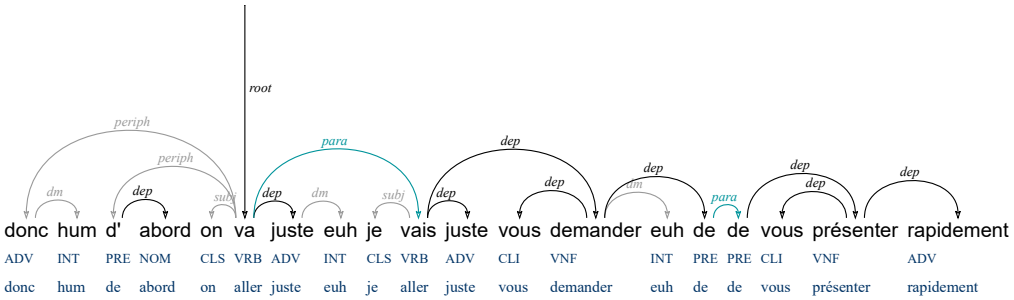


Figure 1: A spontaneous utterance extracted from the French corpus CFPP (Benzitoun et al., 2016). It displays the transcript along with the corresponding lemmas, part-of-speech tags and dependency tree.

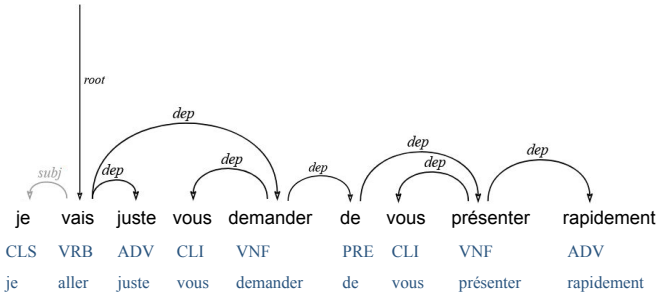


Figure 2: The despontaneified version of the sentence shown in Figure 1.

3. Methodology

As previously noted, we opted to follow an *intuitive* strategy so as to produce simplified versions of spontaneous utterances (Allen, 2009). The absence of preexisting guidelines or syllabi for this task precluded a *structural* approach. Therefore, we decided to rely on the intuition of expert linguists to obtain manual simplifications targeted to non-native speakers, enabling us to empirically investigate the mechanisms involved in simplifying spontaneous utterances in French. We also compared such outputs with those generated through ChatGPT, as we will see below.

Specifically, we decided to focus on French language given: *i*) its rich body of literature describing spontaneity phenomena (Blanche-Benveniste (1997); Bazillon et al. (2008); Luzzati (2013); Evain et al. (2022) to name just a few), and *ii*) the existence of French written-oriented simplification guidelines (Gala et al., 2020). With this work, we intend to address the still unexplored connection between the two areas.

Moreover, we decided to purposely target simplification of spontaneous utterances for a non-native speaking audience, that is, individuals that speak a given language (in this case, French), but have acquired a different first language. Although the scope of this group may be broad, due to the variety of possible cultural backgrounds, language proficiency levels or underlying mother tongues, we specifically focused on non-native speakers given the potential interest that the creation of simpler equivalents may have for this audience. Spontaneous utterances often contain slang expressions and informal register traits, which, along with a dissimilar information structure, can seem unfamiliar to a foreign speaker.

3.1. Source Dataset

In order to analyze the simplification strategies of spontaneous French speech, we resorted to ORFÉO, a well-known platform designed for the study of European French, both in its written and spoken forms (Benzitoun et al., 2016). For our study, we decided to use the latter, known as ORFÉO-CEFC (*Corpus d'Étude du Français Contemporain*), which comprises 12 existing corpora of spoken French. It features segments aligned with audio files at the sentence level, and is enriched with morphosyntactic annotations (such as POS tags or parse trees).

The subcorpora constituting ORFÉO-CEFC cover a wide range of communicative situations (*i.e.*, interviews, tales, phone calls, etc), environments (friendly, academic or familiar) and degrees of spontaneity (Blanche-Benveniste, 1997), ranging from professional situations (like those in `Reunions`) to everyday life ones (as those portrayed in `Cfpp`).

3.2. Sampling

Creating such a resource may be of interest for its eventual reuse as a test set to evaluate spontaneous speech simplification systems. For this reason, we relied on the ORFÉO-CEFC partitioning created by Pupier et al. (2022) and use their evaluation set as the population for our study (ORFÉO-TEST), which amounts to 21,459 segments.

Since the process of manual simplification is a time-consuming task, we opted to extract a subset of the previous distribution. Determining the sample size was key, insofar as we intended to: *i*) analyze a sufficiently representative subset of the original dataset examined, and *ii*) maintain a reasonable workload for annotators, so as to not compromise the stability and consistency of the task.

To ensure the reliability and knowledge extrapolation from the data sample, we performed stratified sampling, dividing the population into 12 distinct strata, each representing a subcorpus of ORFÉO-TEST (see Table 1). Sampling was proportionate, based on the number of segments in each subcorpus, and then randomly collected.

subcorpus	# utt.	%	# sampl.
Cfpb	362	1.69	2
Cfpp	3,232	15.06	15
Clapi	967	4.51	5
Coralrom	1,376	6.41	6
Crfp	2,259	10.53	10
Fleuron	217	1.01	1
Oral-Narr.	1,050	4.89	5
Ofrom	1,476	6.88	7
Reunions	1,245	5.80	6
Tcof	1,997	9.31	9
Tufs	4,525	21.09	21
Valibel	2,753	12.83	13
Total	**21,459**	**100**	**100**

Table 1: Proportional size of each stratum conforming the population. Calculation of the corresponding number of examples for a sample size of 100.

3.3. Human-Based Simplification: Survey Design

As we indicated at the beginning of Section 3, to the best of our knowledge, there are currently no guidelines on how to simplify spontaneous speech transcriptions. This renders a *structural* approach, and thus guideline-adherent, impossible as a simplification strategy. Instead, we opted for an *intuitive approach*, through which to capture the insights of professionals, and on this basis identify the sentence transformations needed to simplify spoken-based French data.

To do this, we decided to set up a manual simplification task based on the stratified sample obtained from ORFÉO-TEST. For this purpose, we enlisted 2 experts, both of them with a solid background in linguistics and a current dedication to research in this field. They both had French (in its European variety) as their native language. The task was hosted on the LimeSurvey platform, and was made accessible from February 1st until February 12th.

As for the survey structure, the sentences derived from our previous sampling were displayed on the LimeSurvey online platform successively. For each spontaneous input sentence, we asked respondents to propose a simplified version (as shown in Appendix A). As part of the instructions, we specified that the goal was to provide simpler equivalents for a French non-native speaking audience. We also asked them to list and explain their chain of thought to transform the input sentence into a simplified equivalent, thus enhancing the explicability of their decision-making. Both fields were mandatory, and we allowed back-and-forth navigation for respondents to revisit their answers.

These instructions, paired with more detailed information, were provided at the beginning of the survey. We kept them visible throughout the execution of the survey, so as to ensure the clarity of the task at hand. To combat potential fatigue during the simplification process, we provided participants with the option to interrupt the task and resume it at their convenience, but always before the due date.

On another note, after a long internal discussion, we opted to merely provide the spontaneous transcript without its corresponding audio file. This decision could have potentially facilitated the task by aiding in disambiguating certain utterances through the inclusion of paralinguistic information (*i.e.*, rhythm, tone, prosody, etc). However, we deliberately chose to challenge participants to simplify based solely on linguistic information. This decision underscores a well-known paradox: spoken language can only be studied on the basis of its written representation (Blanche-Benveniste, 1997).

3.4. Machine-Based Simplification: ChatGPT Prompting

ChatGPT has emerged as an attractive alternative for annotation and typical NLP tasks. Due to its ability to process and generate natural language text, it can assist in various tasks, such as part-of-speech tagging, identifying named entities, or even providing detailed annotations on complex datasets (Gilardi et al., 2023).

We leveraged the OpenAI API and its latest model (`gpt-4-0125-preview`) to simplify spontaneous sentences automatically[2]. The model received the original sentences as input, and produced simpler, more accessible versions of the

[2] Training data: up to December 2023.

same text. Specifically, we ensured that the model was prompted with separate messages to avoid any influence from the dialogue history. Moreover, we used the `temperature=0` setting in every API call to ensure consistency in the model's responses. The prompt included the necessary instructions and was deliberately chosen to be the same as the one given to the human experts (see Figure 7 in Appendix A). We deemed it safer to use the same prompt compared to using distinct ones. This ensured consistency in the information presented to both humans and ChatGPT, enabling a more accurate comparability between responses. The total cost for generating the simplifications of the 100 sentences was approximately 1 USD[3].

4. Results

4.1. Quantitative Evaluation

4.1.1. Taxonomization and Analysis

After the human-based completion of the survey and the machine-based generation of simplified outputs, we taxonomized the different transformations performed to convert spontaneous utterances into the proposed candidate simplifications.

To create the taxonomy, we first analyzed the chain of thought provided by the 3 respondents. On that basis, we derived a macro-categorization using the main edit-based operations: *deletion*, *replacement*, *addition*, *restructuration*, and *copy* (when no alteration to the input was made), that we later subdivided according to the observed linguistic transformations (as shown in Table 2). We then annotated the simplified sentences based on such taxonomy and computed the frequencies for each phenomenon. It should be noted that this stage proved to be more challenging than anticipated: the identification of each operation may not be easily distinguishable, as edits often ensue from jointly applying various transformations (Saggion, 2017). As a result, the computation of occurrences for each phenomenon may have been affected.

As can be seen in Figure 3, it is evident that deletions are the most prevalent among all edit operations. This is hardly surprising in the context of spoken language simplification where hesitations and errors happening during spontaneous speech delivery cannot be undone. While these aspects might be interesting from a pragmatic perspective, they do not provide any propositional content nor relevant semantic information to the sentence, and are thus erased in a simplification context.

Taking a closer look at the distribution of the different suppressed linguistic units (as seen in Plot (a) within Figure 4), it is important to note the dropping of redundant elements such as repetitions or restarts, as well as the suppression of elements related to the enunciation, such as affirmative and negative adverbs (*non, voilà, ouais*), statement verbs (*tu sais*, *je tiens à dire*) or discourse markers (*en fait*). Deletion operations also affect adjectives and adverbs that add little information to the input sentence (*toutes nos traditions* → *nos traditions*).

As for the coherence between the candidate simplifications, the three participants seem to use a similar reasoning to transform the provided inputs, prioritizing deletion operations to achieve simplification. We note, however, that ChatGPT makes more conservative decisions when generating outputs and performs fewer deletions than both humans (see Example I in Table 4 in Appendix B). Between the two linguists, there is an overall symmetry in the number and type of triggered phenomena, although Expert 1 tends to drop more items than Expert 2, especially in terms of restarts and reformulations (see again Plot (a) in Figure 4).

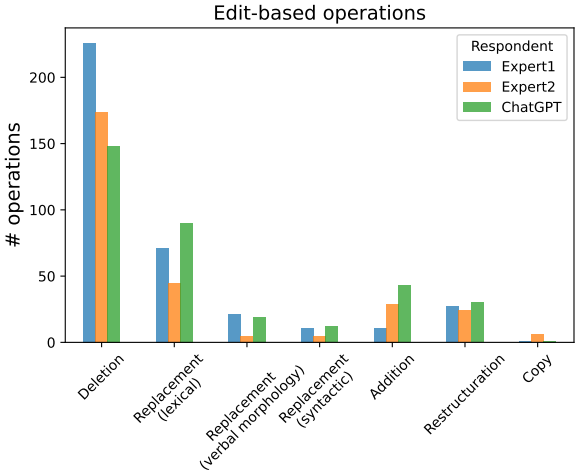


Figure 3: Overview of the distribution of edit-based operations in the analyzed data sample.

As shown in Plot (b), restructuration appeared as a much less frequent edit. All three respondents seldom performed any sentence splitting or merging modifications, very prototypical in written-based simplification. This can plausibly be explained by the typically shorter length of spoken sentences[4]. The prevailing change observed is the reordering

[3] Note that for the used model, the cost for input is 10 USD per 1 million tokens, while for output, it is 30 USD per 1 million tokens (as of April 2024).

[4] It should be noted that the notion of sentence in speech is not straightforward. The absence of sentence boundaries, which are conventionally delimited in writing, complicates the task of distinguishing each segment. For our study, we have relied on the sentential presegmentation provided by ORFÉO-CEFC.

Edit	Level	ID	Linguistic unit(s) affected, operation
Deletion		1	Repetitions
		2	Affirmation and negation words
		3	Interjections
		4	Conjunctions
		5	Discourse markers
		6	Restarts and reformulations
		7	Adverbs and adjectives
		8	Incomplete words
		9	Statement verbs
		10	Pronouns
		11	Verbs with little semantic value
Replacement	Lexical	12	Simpler synonyms for content words
		13	Compression of nominal phrases
		14	More standard equivalents for content words
		15	Smoothing of swear words
	Verbal morphology	16	Intransitive to transitive verbs
		17	Pronominal to non-pronominal verbs
		18	Change of verbal tense
		19	Compression of verbal locutions
	Syntactic	20	Passive to active voice
		21	Cleft to canonical constructions
		22	Neutralization of dislocated subjects
		23	Pronoun transformations
Restructuration		24	Reorder
		25	Sentence splitting
		26	Sentence merging
Addition		27	Explicitation or disambiguation of a word
		28	Completion of truncated sentences
		29	Clarification of uncommon terms
Copy		30	Input sentence is left unchanged

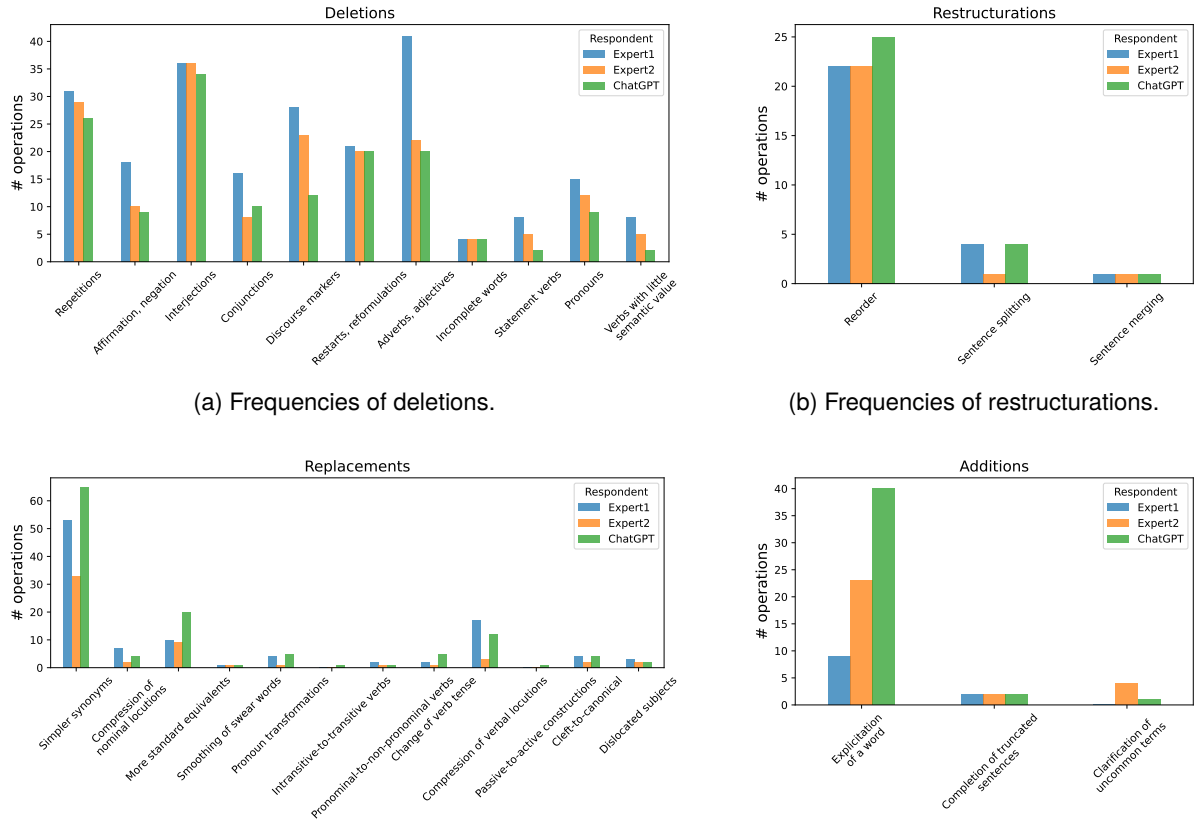
Table 2: Taxonomy of edit operations observed in the data sample, reflecting the simplification process from spoken-based transcripts.

of elements in the utterance, often driven by the search for a canonical subject-verb-object order (as seen in Example II in Table 4). As for the additions, displayed on Plot (d), the most notable category is the explicitation of a word. In this regard, the generative model seemed more inclined than humans to add extra information, with the aim of resolving eventual ambiguities from the source sentence.

Replacement operations, shown in Plot (c), were probably the most interesting edit type. We distinguished 3 linguistic levels of modification: *lexical*, *morphological*, and *syntactic*. Upon closer examination, we uncovered a preference for lexical-based edits (as shown in Figure 3), which were the second most common after deletions. Among these, the most occurring subcategory is the substitution of content words (nouns, adjectives, adverbs, and verbs), in favor of more common alternatives (*i.e.*, *confrérie* → *association* in Example III in Table 4). It is relevant to note in this regard that ChatGPT was the most prone to make changes of this type. This may have been triggered by the provided prompt, where we mentioned lexical substitution of complex terms as an example operation (see Instruction 2 shown in Figure 7).

Besides, we have noticed that the replacement of lexical units does not always stem from complex terms, but rather from slang ones. In these cases, the 3 respondents tended to use more standard equivalents, probably under the hypothesis that colloquialisms may be less familiar terms for foreign speakers. Some examples include: *gosses* → *enfants*, *monde* → *personnes* or *bouquins* → *livres*. The tendency to adopt a more formal register in the crafted simplification is also evidenced in the smoothing of profanity (as illustrated in Example IV in Table 4).

In addition, we observed a propensity to compress the constituents of phrases that do not convey much semantic content, probably on the assumption that a shorter sentence is also often perceived as simpler. This phenomenon can be observed in the shortening of nominal groups (*i.e.*, *monde du travail* → *travail*). That same principle seems to



(a) Frequencies of deletions.

(b) Frequencies of restructurations.

(c) Frequencies of replacements.

(d) Frequencies of additions.

Figure 4: A closer look to the distribution of the edit operations performed on the data sample extracted from ORFÉO-TEST.

apply in the morphology dimension, where verbal locutions or periphrases are often compressed into shorter forms (*i.e.*, *faire la demande* → *demander*), and compound or less frequent tenses tend to be converted into simple or more frequent ones (*elle jouait déjà de la guitare* → *elle joue la guitare*), if meaning is not altered. In addition, even to a lesser extent, we find instances in which pronominal verb forms are replaced by non-pronominal alternatives, and intransitive ones are substituted by transitive variants.

As for changes in syntax, these occur far less frequently than lexical transformations. This may be due to the fact that syntactic edits typically affect a larger span within the utterance than lexical ones, making them intrinsically less numerous. In any case, it is interesting to note that syntactic operations have mainly been applied to constructions that exhibit a marked information structure. For this reason, cleft clauses and dislocated subjects, common in spoken French, are reverted to their canonical non-marked forms (see Example V in Table 4). Finally, it is worth noting that the conversion of passive constructions into active voice is anecdotal. Although diathesis change is a well-established operation in the field of ATS, the use of passive voice is inherently rare in French, and is even less common in a spoken modality.

Overall, the results show that the most common edit operations in spontaneous speech simplification are deletions. The proposed simplifications are often sentential equivalents stripped of any oral marks such as enunciation elements (discourse markers, interjections), hesitations (inherent to the live construction of a message), or the use of slang and profanity (infrequent in a written form). As a result, the proposed simplified outputs are often *writified*, register-standardized versions of the inputs that strictly include their propositional content.

4.1.2. Respondents' Agreement

In the next evaluation step, we seek to understand the level of agreement among participants. We opted for the Jaccard Index because of its adeptness in quantifying the similarity between different answers. This choice was made since participants' responses are not limited to single, mutually exclusive categories but can include multiple selections. This metric calculates the ratio of

the intersection to the union of the sets of choices, providing a clear, normalized value ranging from 0 to 1. Our approach involves comparing the selections of each respondent with every other respondent for the same sentence pair to assess how similar their choices are. The results of this pairwise comparison are: $J(Exp_1, Exp_2) = 0.54$, $J(Exp_1, GPT) = 0.52$, $J(Exp_2, GPT) = 0.51$. Overall, the values suggest a moderate level of agreement among the respondents, with none of the pairs showing a particularly high or low level of consensus. This indicates a generally consistent understanding or interpretation of the operations for making simplifications, but it also highlights the subjective nature of the task (Dmitrieva et al., 2021; Ormaechea et al., 2023b), where individual differences in judgment can lead to variations in the chosen operations.

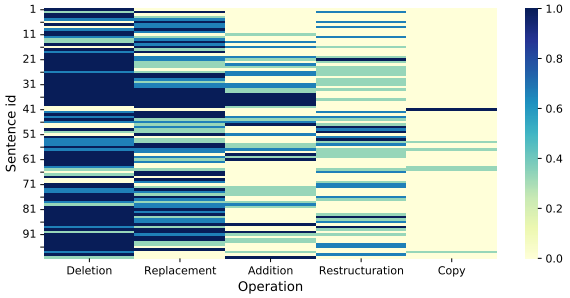


Figure 5: Agreement heatmap across sentences based on the 5 macro-categories.

Generally, there is a trade-off between the level of detail in the taxonomy definition and the desired level of agreement, as finer granularity often leads to more diverse operations and, consequently, reduced consensus (Heineman et al., 2023). To obtain a better insight into the chosen operations, we resorted to the analysis shown in the heatmap of Figure 5, where each cell represents the consensus among the respondents for one specific macro-category and input sentence. The darkest color indicates that all participants performed the same macro-operation on a given sentence. Based on the heatmap analysis, we observed that the agreement among respondents varied depending on the executed operation: in 69.7% of cases, deletion was performed on the same sentence by the three participants, which signifies a consensus on its utility for simplification. For the other cases, we observed a lesser consensus: 41.7% for replacement, 25% for restructuration, 12.2% for addition, and 16.7% for copy.

4.2. Qualitative Evaluation

To assess the suitability of the produced human- and machine-based simplification sentences for a foreign-speaking audience, we conducted a qualitative intrinsic evaluation with three master-level non-native French students. Specifically, they were asked to score the given simplification on a five-point Likert scale (see Table 3), on the basis of two criteria: *i*) *simplicity gain* (S_G): how much simpler is the candidate simplification compared to the original sentence?; and *ii*) *meaning preservation* (M_P): how much of the meaning in the original sentence is preserved in the candidate simplification?

Simplicity gain	Meaning preservation
5 – Much simpler	5 – Fully preserved
4 – Somewhat simpler	4 – Mostly preserved
3 – Same difficult	3 – Partially preserved
2 – More difficult	2 – Completely different
1 – Unintelligible	1 – Unintelligible

Table 3: Labels assigned to each score. Inspired on the taxonomy by Yamaguchi et al. (2023).

Judges were shown the original sentences along with the simplified versions proposed by one of the three respondents in a random order (see Figure 8 in Appendix C). Based on their assessment, we observed that three judges, each with slightly different but closely aligned evaluations, agreed that Expert 1 was the most proficient at providing simpler sentences (see Figure 6). Whereas Expert 1 achieves high S_G scores, Expert 2 makes more conservative decisions, leading to a lower gain, yet obtaining a higher average than Expert 1 in the M_P dimension. ChatGPT receives an intermediate mean score for both criteria, and seems to find a trade-off between these two seemingly inverse tendencies.

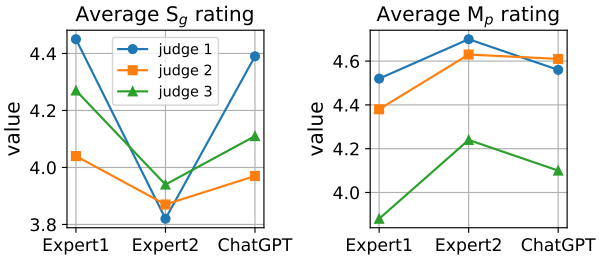


Figure 6: Average rating on S_G and M_P dimensions from the judges.

5. Conclusions and Further Work

In this paper, we have presented a taxonomy of the simplification strategies applied on the basis of French spontaneous transcripts for a non-native audience. To date, research on simplification has been primarily based on written sources, but seldom on spoken-based ones. Due to the lack of guidelines allowing us to steer this process, we

adopted an *intuitive* approach to characterize the strategies employed to simplify this kind of data. By means of a survey-based study, we collected a set of simplifications from 2 native French-speaking linguists. More precisely, we asked them to provide an explainable simplified version of 100 spoken utterances randomly selected from ORFÉO-TEST. Additionally, we have compared human-crafted speech simplifications, with machine-generated ones. Based on the quantitative evaluation, ChatGPT tends to suppress fewer elements when generating simplified outputs compared to human experts. As for the qualitative evaluation, it suggests an inverse correlation between S_G and M_P criteria. Results show that Expert 1 achieves higher S_G than ChatGPT, but the latter strikes a more balanced compromise between the two dimensions.

With this work, we provide a multi-reference set that allows to map the existing ORFÉO-TEST audio-transcript pairs with simpler counterparts. Assuming that the intuitions provided by experts serve as ground truth simplified sentences, this resource can be further used to assess automated solutions for generating spontaneous speech simplifications. For these reasons, we have released on a GitHub repository the resulting set mapping the original transcripts to their corresponding expert simplifications, named PROPICTO-ORFÉO-SIMPLE[5].

Furthermore, by annotating edit operations, we enable a finer-grained evaluation and a better understanding of the patterns that a model would have applied. This can promote greater explainability compared to conventional scores used to assess model performance (*i.e.*, BLEU or SARI). These overall metrics often provide little information about the simplification operations that the system has learned. Additionally, this in-depth examination can further serve as the groundwork for defining guidelines on speech-based simplification.

As for the limitations of the study, the lack of context in the manual sentence-level simplification was pointed out by the experts as a difficulty for its completion. Of course, providing context would have facilitated the task, especially within spontaneous speech, which is by nature interactive and conversational. However, we chose random proportionate sampling with the aim of favoring a better representativeness of the extracted sample. Consequently, the resulting data being analyzed lacked context as the sentences comprising it originated from various strata and were not linked to a single conversation.

[5] PROPICTO-ORFÉO-SIMPLE is made available on the following link: https://www.ortolang.fr/market/corpora/propicto.

Acknowledgements

This work is part of the PROPICTO (French acronym standing for *PRojection du langage Oral vers des unités PICTOgraphiques*) project, funded by the Swiss National Science Foundation (N°197864) and the French National Research Agency (ANR-20-CE93-0005).

We would also like to express our sincere gratitude to the linguists who took the time to perform the manual simplification task, and to the participants who completed the qualitative intrinsic evaluation.

6. Bibliographical References

David Allen. 2009. A study of the role of relative clauses in the simplification of news texts for learners of English. *System*, 37:585–599.

Katrin Angerbauer, Heike Adel, and Ngoc Thang Vu. 2019. Automatic Compression of Subtitles with Neural Networks and its Effect on User Experience. In *Interspeech 2019*, pages 594–598. ISCA.

Thierry Bazillon, Vincent Jousse, Frédéric Béchet, Yannick Estève, Georges Linarès, and Daniel Luzzati. 2008. La parole spontanée : transcription et traitement. In *Traitement Automatique des Langues, Volume 49, Numéro 3 : Recherches actuelles en phonologie et en phonétique : interfaces avec le traitement automatique des langues*, pages 47–76. ATALA.

Daniel Becquemont. 1996. Le sous-titrage cinématographique : contraintes, sens, servitudes. In Yves Gambier, editor, *Les transferts linguistiques dans les médias audiovisuels*, pages 145–155. Presses universitaires du Septentrion.

Christophe Benzitoun, Jeanne-Marie Debaisieux, and Henri-José Deulofeu. 2016. Le projet ORFÉO : un corpus d'étude pour le français contemporain. *Corpus, 15*.

Rocío Bernabé and Piero Cavallo. 2021. Easy-to-Understand Access Services: Easy Subtitles. In *Universal Access in Human-Computer Interaction. Access to Media, Learning and Assistive Environments*, pages 241–254. Springer International Publishing.

Claire Blanche-Benveniste. 1997. *Approches de la langue parlée en français*. Ophrys.

Dominique Brunato, Felice Dell'Orletta, and Giulia Venturi. 2022. Linguistically-Based Comparison of Different Approaches to Building Corpora for Text Simplification: A Case Study on Italian. *Frontiers in Psychology*, 13.

François Buet and François Yvon. 2021. Toward Genre Adapted Closed Captioning. In *Interspeech 2021*, pages 4403–4407. ISCA.

Andrew Caines, Michael McCarthy, and Paula Buttery. 2017. Parsing Transcripts of Speech. In *Proceedings of the Workshop on Speech-Centric Natural Language Processing*, pages 27–36. Association for Computational Linguistics.

Arnaldo Candido, Erick Maziero, Lucia Specia, Caroline Gasperin, Thiago Pardo, and Sandra Aluisio. 2009. Supporting the Adaptation of Texts for Poor Literacy Readers: A Text Simplification Editor for Brazilian Portuguese. In *NAACL HLT Workshop on Innovative Use of NLP for Building Educational Applications*, pages 34–42.

Ronald Carter and Michael McCarthy. 2017. Spoken Grammar: Where Are We and Where Are We Going? *Applied Linguistics*, 38(1):1–20.

Walter Daelemans, Anja Höthker, and Erik Tjong Kim Sang. 2004. Automatic Sentence Simplification for Subtitling in Dutch and English. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association (ELRA).

Jan De Belder and Marie-Francine Moens. 2010. Text Simplification for Children. In *Workshop on Accessible Search Systems*, pages 19–26.

Anna Dmitrieva, Antonina Laposhina, and Maria Lebedeva. 2021. A Comparative Study of Educational Texts for Native, Foreign, and Bilingual Young Speakers of Russian: Are Simplified Texts Equally Simple? *Frontiers in Psychology*, 12.

Solene Evain, Solange Rossato, Benjamin Lecouteux, and François Portet. 2022. Typologie de la parole spontanée à des fins d'analyse linguistique et de développement de systèmes de reconnaissance automatique de la parole. In *Proc. XXXIVe Journées d'Études sur la Parole*, pages 212–221.

Núria Gala, Anaïs Tack, Ludivine Javourey-Drevet, Thomas François, and Johannes C. Ziegler. 2020. ALECTOR: A Parallel Corpus of Simplified French Texts with Alignments of Misreadings by Poor and Dyslexic Readers. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1353–1361.

Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. 2023. ChatGPT Outperforms Crowd-Workers for Text-Annotation Tasks. *Proceedings of the National Academy of Sciences*, 120(30).

Michael Alexander Kirkwood Halliday. 1985. *Spoken and written language*. Oxford University Press.

David Heineman, Yao Dou, Mounica Maddela, and Wei Xu. 2023. Dancing Between Success and Failure: Edit-level Simplification Evaluation using SALSA. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 3466–3495. Association for Computational Linguistics.

Colby Horn, Cathryn Manduca, and David Kauchak. 2014. Learning a Lexical Simplifier Using Wikipedia. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pages 458–463.

William Hwang, Hannaneh Hajishirzi, Mari Ostendorf, and Wei Wu. 2015. Aligning Sentences from Standard Wikipedia to Simple Wikipedia. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 211–217.

Juhani Luotolahti and Filip Ginter. 2015. Sentence Compression For Automatic Subtitling. In *Proceedings of the 20th Nordic Conference of Computational Linguistics*, pages 135–143. Linköping University Electronic Press, Sweden.

Daniel Luzzati. 1998. Rhétorique et description de l'oral. *VERBA*, 25:7–30.

Daniel Luzzati. 2013. Enseigner l'oral spontané ? In Beacco J.C., editor, *Ethique et politique en didactique des langues*, pages 188–207. Éditions Didier.

Anna Matamala. 2022. Easy-to-understand language in audiovisual translation and accessibility: State of the art and future challenges. *XLinguae*, 15(2):130–144.

Sneha Mehta, Bahareh Azarnoush, Boris Chen, Avneesh Singh Saluja, Vinith Misra, Ballav Bihani, and Ritwik K. Kumar. 2020. Simplify-Then-Translate: Automatic Preprocessing for Black-Box Translation. In *AAAI Conference on Artificial Intelligence*.

Lucía Ormaechea, Pierrette Bouillon, Maximin Coavoux, Emmanuelle Esperança-Rodier, Johanna Gerlach, Jerôme Goulian, Benjamin Lecouteux, Cécile Macaire, Jonathan Mutal, Magali Norré, Adrien Pupier, and Didier Schwab. 2023a. PROPICTO: Developing Speech-to-Pictograph Translation Systems to Enhance Communication Accessibility. In *Proceedings of the 24th Annual Conference of the European*

Association for Machine Translation, pages 515–516. European Association for Machine Translation.

Lucía Ormaechea and Nikos Tsourakis. 2023. Extracting Sentence Simplification Pairs from French Comparable Corpora Using a Two-Step Filtering Method. In *Proceedings of the 8th Swiss Text Analytics Conference 2023*. Association for Computational Linguistics.

Lucía Ormaechea, Nikos Tsourakis, Didier Schwab, Pierrette Bouillon, and Benjamin Lecouteux. 2023b. Simple, Simpler and Beyond: A Fine-Tuning BERT-Based Approach to Enhance Sentence Complexity Assessment for Text Simplification. In *Proceedings of the 6th International Conference on Natural Language and Speech Processing (ICNLSP)*, pages 120—133.

Adrien Pupier, Maximin Coavoux, Benjamin Lecouteux, and Jerome Goulian. 2022. End-to-End Dependency Parsing of Spoken French. In *Proc. Interspeech 2022*, pages 1816–1820.

Luz Rello, Ricardo Baeza-Yates, and Horacio Saggion. 2013. DysWebxia: Textos Más Accesibles Para Personas con Dislexia. *Procesamiento del Lenguaje Natural*, 51.

Horacio Saggion. 2017. *Automatic Text Simplification*. Morgan & Claypool Publishers.

Matthew Shardlow and Raheel Nawaz. 2019. Neural Text Simplification of Clinical Letters with a Domain Specific Phrase Table. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 380–389. Association for Computational Linguistics.

Sanja Stajner and Maja Popovic. 2016. Can Text Simplification Help Machine Translation? *Proceedings of the 19th Annual Conference of the European Association for Machine Translation*, pages 230–242.

Laurens Van den Bercken, R. H. J. Sips, and C. Lofi. 2019. Evaluating Neural Text Simplification in the Medical Domain. *The World Wide Web Conference*.

Wei Xu, Chris Callison-Burch, and Courtney Napoles. 2015. Problems in Current Text Simplification Research: New Data Can Help. *Transactions of the Association for Computational Linguistics*, 3:283–297.

Daichi Yamaguchi, Rei Miyata, Sayuka Shimada, and Satoshi Sato. 2023. Gauging the Gap Between Human and Machine Text Simplification Through Analytical Evaluation of Simplification Strategies and Errors. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 359–375. Association for Computational Linguistics.

Xingxing Zhang and Mirella Lapata. 2017. Sentence Simplification with Deep Reinforcement Learning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 584–594.

7. Language Resource References

Benzitoun, Christophe and Debaisieux, Jeanne-Marie and Deulofeu, Henri-José. 2016. *Corpus d'Étude pour le Français Contemporain (CEFC)*. Distributed via ORTOLANG. PID https://repository.ortolang.fr/api/content/cefc-orfeo/11/documentation/site-orfeo/index.html.

A. Expert Simplification Task: LimeSurvey Example

Figure 7: A sample of a spontaneous utterance, along with the instructions for its simplification, on the LimeSurvey platform.

B. Simplification Examples: Respondents' Proposed Answers

Ex. I	**Input**	*ouais c'est ça sauf que moi on m'a jamais expliqué le rythme du coup*
	Expert1	On ne m'a pas expliqué le rythme
	Expert2	Oui, c'est ça, sauf qu'on ne m'a jamais expliqué le rythme.
	ChatGPT	Oui, c'est vrai, mais personne ne m'a jamais expliqué le rythme.
Ex. II	**Input**	*donc il a adapté euh cet univers cette légende à sa façon le réalisateur qui s' appelle Alexandre Astier*
	Expert1	Le réalisateur Alexandre Astier a adapté cette légende à sa manière
	Expert2	Le réalisateur, qui s' appelle Alexandre Astier, a adapté cet univers à sa façon.
	ChatGPT	Donc, Alexandre Astier, le réalisateur, a changé cette légende, cet univers, à sa manière.
Ex. III	**Input**	*mais enfin cette confrérie elle avait quand même un un avantage c' est que elle venait en aide euh aux malheureux*
	Expert1	Cette association avait un avantage, elle a aidé les démunis
	Expert2	Cette association avait l'avantage de venir en aide aux malheureux
	ChatGPT	Mais cette confrérie aidait quand même les malheureux.
Ex. IV	**Input**	*et c' était pas son style de bouffer de la merde*
	Expert1	c'était pas à son goût de manger mal
	Expert2	Et il/elle n'est pas habitué à manger des aliments de mauvaise qualité
	ChatGPT	Il n'aimait pas manger de mauvaises choses.
Ex. V	**Input**	*on sent que la prise de conscience de ce genre de choses elle s' est faite tard*
	Expert1	Nous pensons que la compréhension de ce problème est arrivée tard
	Expert2	La prise de conscience de ces choses-là est arrivée tard
	ChatGPT	Les gens ont commencé à comprendre ces choses tard.

Table 4: A set of examples extracted from the data sample along with the proposed simplifications.

C. Qualitative Evaluation: LimeSurvey Example

Figure 8: An example (comprising the original spontaneous transcript and a candidate simplification) of the qualitative evaluation task on the LimeSurvey platform.