Potential of ASR for the study of L2 learner corpora

Sarra El Ayari, Zhongjie Li

Potential of ASR for the study of L2 learner corpora

Sarra El Ayari
Structures Formelles du Langage
CNRS & Paris 8 University
sarra.elayari@cnrs.fr

Zhongjie Li
Structures Formelles du Langage
CNRS & Paris 8 University
lzj44010@gmail.com

Abstract

This study is at the crossroads of Natural Language Processing (NLP) and Second Language Acquistion (SLA). We used Whisper's speech recognition on a French L2 learner corpus to get automatic transcripts, and compared them with pre-existing manual transcripts. We then conducted quantitative and qualitative analysis of the issues which are inherent to the specificities of interlanguage for any automatic tool. We will discuss the different issues encountered by Whisper that are specific to learner corpora.

1 Introduction

The TranSLA project aims at analyzing to which extend Automatic Speech Recognition systems (ASR) can provide useful information on the distance between interlanguage and the internalized norm of those systems. Providing tools for corpus linguistics is an essential part of the research carried out in Second Language Acquisition (SLA). Recent technological advances raise new methodological questions. The act of transcribing involves an initial task of interpreting the discourse in L2, which is particularly delicate since it can influence the researcher's subsequent analysis (Benazzo and Watorek, 2021).

If the results obtained for speech recognition in general are very encouraging (Radford et al., 2023), we still need to be able to evaluate precisely their performance on non-standard languages, such as interlanguage of foreign learners (Selinker, 1972). Interlanguage is the idiolect developed by second language learners and it refers to the mental grammar constructed by a learner at a specific stage of the learning process (Ellis and Barkhuize, 2005). It is therefore intrinsically subject to variation and evolution simultaneously and possesses a unique linguistic organization.

This study aims firstly at measuring the performance of an ASR system on a L2 French learner corpora, and secondly to observe if ASR systems could be used as a tool to evaluate how close or distant learners speech productions can be from the language model that is used, and therefore to correlate it with learners' acquisition levels. We will discuss the discrepancies linked to SLA issues as well.

2 Transcription of learner corpora

The transcription process is a time-consuming phase for any researcher who wants to work on audio or multimodal data. It is also a very precise work that requires already to have clear thoughts about which linguistics phenomena will be analyzed, and therefore which elements have to be transcribed and how.

In Second Language Acquisition, this process is even more important because fine-grained access to information is crucial. To transcribe exactly what the learners are actually saying and pronouncing is the goal - even if it is not always attainable. In that way, how to transcribe is already a choice. It is even more complicated when the language has a wide gap between oral and written modalities, like French (Blanche-Benveniste, 2000). Thus choosing one form over the others carries the risk of over-estimate or under-estimating the knowledge of the learner (Benazzo and Watorek, 2021).

We present a few examples from the ESF (*European Science Foundation Second Language*) corpus (Perdue, 1993) which shows the problems of choosing a specific form for transcription in Figure 1.

The transcriptions presented (El Ayari and Wa-

- /ʒəparlɛ/ vs. /ʒeparle/ => 'je parlais' vs. 'j'ai parlé'
- /ʒəsɛ / vs. / ʒeesɛje / => 'j'essaie' vs. 'j'ai essayé'
- /ilese/' vs. / illese/ => 'il essaie' vs. 'il l'essaie' ou 'il les sait'
- /ilavole/ vs. /illavole/ => 'il a volé' vs. 'il l'a volé'

Figure 1: Examples of transcriptions

torek, 2021) show that transcribing a corpus is already choosing which linguistic form the learner has pronounced, even though we do not always have enough knowledge to decide.

3 State of art

Not many studies focused on ASR systems' performances on non-native languages. It is a very important matter as those systems have been trained predominantly on standard varieties (Graham and Roll, 2023). Studies focusing on learner corpora and ASR mostly focus on the global evaluation of ASR performances: (Graham and Roll, 2023; Cumbal et al., 2021) on Swedish or on providing pronunciation feedback with a focus on phonetic features: (Wei et al., 2022) on Dutch, (Ballier et al., 2023; Chanethom and Henderson, 2022) on French. We did not find any studies intertwining ASR performances and learners' proficiency.

4 The corpus

The LANGSNAP corpus[1] is based on the study abroad of 29 advanced learners of L2 French (mit) (eleven years length of French study). The learners are L1 English speakers and Anglophone university students, learning French over a 21-month period, including a 9-month stay abroad. This analysis is based on 14 participants. The audio data as well as the transcriptions are freely available on Talkbank[2] (CLARIN Knowledge Centre), an open access integrated repository for spoken language data.

The LANGSNAP corpus is longitudinal and therefore offers a good basis to compare the oral productions of the learners at different times. There are different linguistic tasks available: oral interviews (where participants took part in a semi-structured interview led by a member of the research team); story retelling (where participants retold a story guided by a sequence of pictures); argumentative writing (where participants wrote a timed 200-words response to a stimulus question). We chose to analyze the oral interviews where participants took part in a semi-structured interview led by a member of the research team, which have been conducted regularly through the project. We analyzed data at different times: October 2011 in stay abroad (T1), May 2012 in stay abroad (T2) and October 2012 post stay abroad (T3). The interviews have already been manually transcribed in chat format (MacWhinney, 2000), with speech alignment. The corpus contains also the same oral interviews performed by French native speakers manually transcribed too, which we will use as a baseline for the ASR performances on native French.

Examples of utterances:

(1) alors pour commencer décris moi où tu habites et les gens avec qui tu habites ?

(2) &-euh j'habite à City donc c'est une ville vers &-euh l'ouest <de la France>[//] &-euh de Paris.

This corpus is ideal for looking at the evolution of the interlanguage of the learners (Corder, 1980), as they have produced the same tasks at different times and as the data have been transcribed and analyzed beforehand. The data have a good audio quality without background noises, which is also something important to take into consideration for an automatic analysis.

5 Methodology

Our methodology consists in comparing the transcriptions obtained automatically by Whisper[3] to the ones produced manually to see precisely the differences and to evaluate the performance of the ASR system in general on this corpus.

5.1 ASR system

We used the ASR system Whisper, created by OpenAI. Our choice of ASR is a pragmatic one as Whisper is the only one freely available on governmental servers by the IR Huma-Num[4] and the CINES[5] in France (release 20231117). It has been

[1]LANGSNAP: https://web-archive.southampton.ac.uk
[2]Talkbank: https://www.talkbank.org
[3]Whisper: https://github.com/openai/whisper
[4]Huma-Num: https://www.huma-num.fr
[5]CINES: https://www.cines.fr

trained on French dataset, and therefore can produce speech recognition task and automatic transcriptions of oral data.

"Whisper architecture is a simple end-to-end approach, implemented as an encoder-decoder Transformer. [...]. A decoder is trained to predict the corresponding text caption, intermixed with special tokens that direct the single model to perform tasks such as language identification, phrase-level timestamps, multilingual speech transcription" (Radford et al., 2023).

The challenge here is to see how well the system performs on the particular oral data that are learner oral productions. As different linguistics levels are in the process of being acquired, the transitional aspect of interlanguage offers difficulties for any type of automatic process. Pronunciation, vocabulary, morphology and syntax will not be standard. As such, learner corpora can be considered as one type of less-resourced language, and specific resources might be needed to process them accurately.

5.2 Evaluation metrics

Different metrics are can be used to evaluate ASR systems. WER (*Word Error Rate*) evaluates the proportion of correct words compared to manual transcripts, while the CER (*Character Error Rate*) measures the proportion of correct characters. Both metrics are commonly used to quantify ASR performance. We are aware that those metrics have limitations such as only taking into account the word level and therefore not pondering the results linked to semantic similarity. Nevertheless they offer us a global metric to evaluate Whisper's performances despite the evolving nature of L2 data and interlanguage. We wanted to get a global overview of the results across time for a semi-control task. Nevertheless, we will deepen the analyse by looking closely at Whisper's corrections: insertions, substitutions and deletions in order to get a better understanding of the results. We did not look into Part-Of-Speech Error Rate because of the nature of the data, and particularly the bias created by the meaning idiosyncrasy where a form used by a learner does not imply that its linguistics function is also mastered (*proximity fallacy* (Perdue, 1993)).

5.3 Data processing

Our goal is to provide parallel corpora in order to compare manual and automatic transcripts. Figure 2 shows the pipeline for files normalization, in order to be able to compare the transcriptions.

The manual transcripts are in chat format, which belongs to the CLAN program (*Computerized Language ANalysis*) (MacWhinney, 2000). The speakers are introduced by a code and an asterisk and a pos-tagging has been automatically generated (line %mor) as shown on Figure 3.

We have encountered different issues during the process of the data. The first difficulty encountered when processing the transcriptions is the turn-taking. Long turns of speech are cut into several lines so it was difficult to combine the lines together in order to compare them. Secondly, as manual transcriptions have been done by different transcribers at different times, human errors and changes in the transcription guide had to be taken into consideration and were lacking regularities.

Another issue is linked to Whisper itself which creates bugs increasing the WER score of automatic transcriptions. The first bug relates to language changes detected in the middle of a French transcription. In the examples below, the transcription alternates between several languages and is not pronounced by the speaker at all (extra-hallucinated errors).

(3) Euh... Ça lui soulage la sensation. J is subi au passé. tardised beara. Hope you are okay now. Jean Apple. Brooke de Ney sur une locale.

The second type of bug concerns the repetition of a word or words in several lines. Here, as illustrated in the following example, the word "oui" is reproduced in several lines, which is not the case in the audio file. Word repetition degrades the quality of the automatic transcription by adding non-existent words or replacing several turns of speech.

(4) où il y a des élèves un peu difficiles. Oui. Il n'y a pas beaucoup... Une prof m'a dit qu'elle a... Oui. Oui. Oui. Oui. Oui. Oui.

The third type of bug is that Whisper fails to detect speech for certain audio sequences, leaving

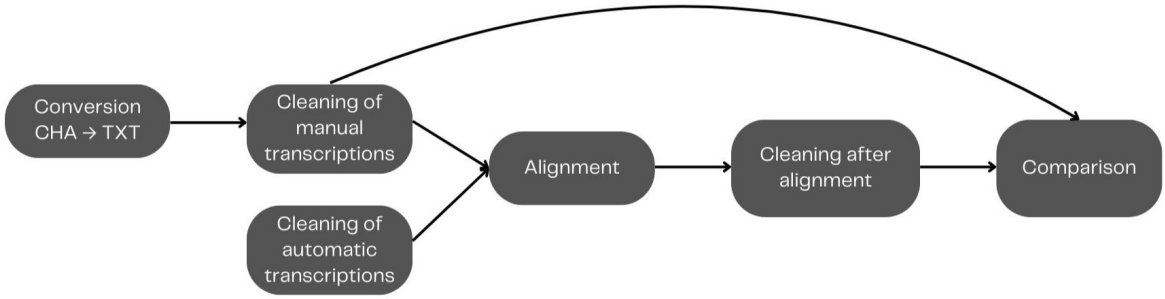

Figure 2: Pipeline

```
*100:  donc c'est une ville vers &-euh l'ouest <de la France> [//]
&-euh de Paris . •12032_16266•
%mor:  adv|donc pro:dem|ce$v:aux|être&PRES&3s det:art|une&f&sg
n|ville&f prep|vers det:art|le$n|ouest&m prep|de n:prop|Paris .
*KMCM:  hum hum . •16266_17279•
%mor:  co|hum co|hum .
*100:  j'habite avec ma tante et &-euh voilà . •17279_20299•
%mor:  pro:sub|je$v|habiter-IMP&2s prep|avec det:poss|ma&sg
n|tante&f conj|et adv:place|voilà .
```

Figure 3: Manual transcription format

white spaces instead - which requires also manual intervention in order to preserve the alignment between the two transcripts.

Most of the normalization process has been done automatically with python scripts, after analyzing the data. Nevertheless because of Whisper's irregularities, we had to manually check and sometimes manually correct the utterances to guarantee the accuracy of the alignment.

5.4 Harmonization of the data

The preparation of the data for metrics calculation has been done in two steps.

Manual transcriptions have been done by different transcribers and therefore do not always follow the same conventions. We had to address this, especially as oral conventions in Clan can have different formats. Different elements had to be removed, such a speaker codes, timestamps, punctuation and characters used for idioms. A conversion to lowercase has been done. A specific treatment on the numbers to convert them into words, as well as for the time. Table 1 illustrated the problems encountered to be able to compare the transcripts as accurately as possible.

5.5 Transcriptions' comparison

Transcription is the first stage in the study of any oral corpus and, as such, it implies theoretical choices. We assume that "learner varieties are not imperfect imitations of a 'real language' - the target language - but systems in their own right, error-free by definition" (Klein and Perdue, 1997). Indeed, there are notable differences, both quantitative and qualitative linguistic behavior between native and non-native languages (Dekydtspotter et al., 2006). For those reasons, two important points need to be kept in mind:

- **comparative fallacy** (Bley-Vroman, 1983): learners' language explained by reference to the target language system rather than as set of rules and performance characteristics ;
- **closeness fallacy** (Perdue, 1993) : learners' language explained by attributing references of the target language on the basis of their formal resemblance.

Those two elements are likely to cause difficulties for automatic tools processing on learner corpora.

We developed a framework to visualize both results at the same time, and automatically highlight and categorize the differences: elements inserted, replaced or deleted, and to be able to check the audio file for each utterance. The Figure 4 shows an excerpt of the interface.

Figure 4: Interface of the comparison framework

Issues	Manual transcripts	Correction
Speakers code	***109:** mais je suis pas sûre	mais je suis pas sûre
Timestamps	je suis pas sûre . **137805_139862**	je suis pas sûre
Compound words	rez **+** de **+**chaussé	rez de chaussé
Type case	j'habite au **City**	j'habite au **city**
Disfluencies	je suis content **&-euh** ici	je suis content ici
Ponctuations	Oui **,** et où **?**	Oui et où
Numbers	environ **3** minutes	environ **trois** minutes
Time	à **1h30** du matin	à **une heure et demie** du matin

Table 1: Transcripts' harmonization

5.6 Evaluation measures

WER metric, derived from Levenshtein's distance, provides a score based on the number of incorrectly transcribed words. The higher the score, the lower the similarity between the documents being compared similarity. CER metric indicates the percentage of characters that were incorrectly predicted. They are defined by the ratio between the number of incorrectly aligned words/characters and the total number of words/characters in the reference transcript:

$$WER|CER = \frac{s+i+d}{n}$$

where s, i and d are the number of substitutions, insertions and deletions and n is the total number of words/characters in the reference transcript. They both measure the overall word/character recognition performance without distinguishing between fluent and disfluent words (Lou and Johnson, 2020). Both calculations have been done with Python and the JiWER package [6].

5.7 Speech disfluencies

Speech disfluencies are non-pathological hesitances happening during speaking, like the use of fillers ("like" or "uh") or the repetition of a word or phrase. Unfortunately, "for faithful transcription of conversational speech, there remain challenges both in terms of the content predicted by [transformer based] models (hallucinations, unintended normalization of disfluencies and transcriptions of background noises) and in terms of alignment accuracy" (Yamasaki et al., 2023). The main reason being that the models of ASR systems are trained on fluent (and native) speech, the mismatch between training data and other types of corpora decreases their performance (Lou and Johnson, 2020).

[6] JiWER: https://pypi.org/project/jiwer

6 Results

In this section, we will be comparing the two types transcriptions: manual (MT) versus automatic (AT). Results are better for natives, a type of speech closer to the ones Whisper has been trained on - especially if we remove the speech disfluencies. Taking those into account make a real difference in the calculation of WER and CER for audio corpora, in tasks such as interviews where speakers are speaking freely and answering questions.

	+ disfluencies		- disfluencies	
Corpus	WER	CER	WER	CER
L-T1	0.31	0.25	0.25	0.19
L-T2	0.35	0.26	0.29	0.23
L-T3	0.28	0.21	0.22	0.18
Natives	0.36	0.28	0.23	0.17

Table 2: WER and CER measurements

A WER score between 0.1 and 0.2 is considered as good. The results without disfluencies, especially for natives and learners after stay abroad are good for that kind of corpora. We can conclude that Whisper's performances on the LANGSNAP corpus, for native speakers and advance learners are very decent.

6.1 Longitudinal scores

Our second research question concerns the hypothesis that ASR evaluation metrics can be correlated with learners' proficiency and should therefore decrease as learners get closer to the French speech Whisper has been trained on.

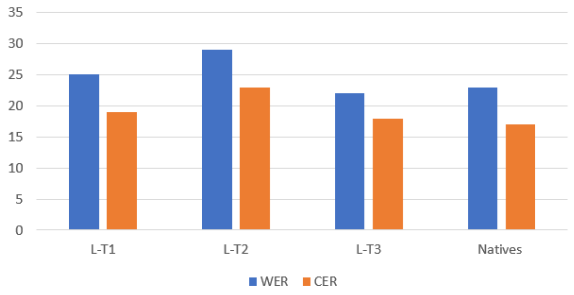

Figure 5: Longitudinal WER & CER metrics

As shows Figure 5 , both WER and CER metrics get lower as the learners improve their knowledge of French, and get closer to the results obtained on the natives speakers. This result is consistent with the improvement of the learners and their acquisition level in general.

If the results between T1 and T3 are decreasing (WER at T1: 0.25 / WER at T3: 0.22), we can also see that they are increasing at T2. To explain this phenomena from an acquisitional point of view, we can point out the critical rule hypothesis stated by W. Klein (Klein, 1989). The idea is that a linguistic rule inside interlanguage is not definitive and therefore is subject to change and evolve. So it could be possible that T2 would represent a specific acquisitional time where rules acquired by learners during language courses would evolve through the stay abroad, because of direct input from native speakers and that some linguistic rules would later be acquired in T3.

6.2 Overview of ASR process

In order to get a better understanding of the ASR results and correlate them to learners' proficiency, we took a closer look on the substitutions, insertions and deletions performed by Whisper. The Figure 6 shows the percentage of those three processes for each times.

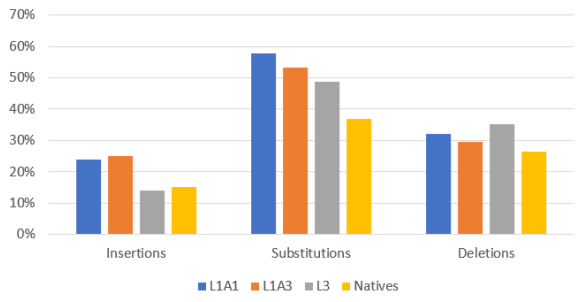

Figure 6: ASR processes

We will get a closet look to each of those three processes by comparing manual transcripts (MT) to automatic ones (AT).

6.2.1 Insertions

Insertions are what we define here as hyper-corrections or hyper-normalizations of the learners' speech.

(5) a. MT: je dois je me dis toujours c'est une expérience

b. AT: je dois je me dis toujours **que** c'est une expérience

Here Whisper adds a subordinating conjunction to the speaker's utterance.

6.2.2 Substitutions

Substitutions are mostly linked to morphology: number, gender, definiteness and verb tenses. French has a relatively complex orthography (van den Bosch et al., 1994) and contains a large number of silent letters which correspond to morphological markers, which make the transcription's process even more difficult.

(6) a. MT: j'ai imaginé que institute **était** à paris

b. AT: j'ai imaginé que j'ai institute **été** à paris

(7) a. MT: je me suis **inscrit** pour faire le marathon

b. AT: je me suis **inscrite** pour faire le marathon

(8) a. MT: danse **aérobic**

b. AT: danse **aérobique**

Most of the insertions are linked to negative forms:

(9) a. MT: je l'aime pas

b. AT: je **ne** l'aime pas

6.2.3 Deletions

Deletions are mostly about Whisper not processing normal speech disfluency, where people can repeat something twice while hesitating or thinking what to say next:

(10) a. MT: je fais la permanence du soir qui est **jusqu'au jusqu'à** dix neuf heures

b. AT: je fais la permanence du soir qui est **jusqu'à** dix neuf heures

We also encounter deletions where the learner produces a non-canonical form that is corrected by Whisper by deleting a character, such as contractions and speech pauses:

(11) a. MT: j'ai imaginé **que à** la bibliothèque je rencontrerais beaucoup **des** gens

b. AT: j'ai imaginé **qu'à** la bibliothèque je rencontrerais beaucoup **de** gens

Those three phenomena are expected in oral treatments. They show that Whisper has difficulties with elements linked to spontaneous speech, such as hesitations, repetitions, disfluencies, contractions. Those examples also show that it is difficult for the system to provide utterances that are not following a specific format, even when the pronunciation differs - like changing the words' order. There are typical corrections that one has to correct back in order to access interlanguage properly.

6.3 Specific SLA issues

Whisper tends to (hyper-)normalize the speech of the learners: Table 3 shows a few examples which are problematic when one is studying learners' productions for different reasons.

Those issues are extremely problematic for researchers who work in the SLA field, because it does not provide enough accuracy. The issues are linked to different specificities of a learner speech in L2 as pronunciation, prosody, fluency, pauses, morphology, syntax and different idiosyncrasies. The item **expériencer** for example is very important to acknowledge because it is a clear hint of the acquisition of verbal morphology from the learner.

Whisper rewrites the data according to the language's model deducted from the training dataset but does not provide a systematic treatment, as show the examples below from the same learner:

(12) a. MT: mais **le FLE** est vraiment similaire de le cours français

b. AT: mais **le fleur** est vraiment similaire au cours français

(13) a. MT: et puis **le FLE** c'est le français langue étrangère

b. AT: et puis **le bleu** c'est le français langue étrangère

FLE stands for Français Langue Etrangère (French as a Foreign Language). The system here provides two different proposals for the same unknown word.

We have also see that some learners had troubles with the sound /y/ in French, and pronounced it sometimes /u/. As it is a productive difference in French, Whisper sometimes misinterpreted the second-person pronoun :

(14) a. oui et **tout** marche très bien

b. oui et **tu** marches très bien

(15) a. maintenant **tout est tout passe** bien

b. maintenant **tu es tu passes** bien

The pronunciation of French for a L2 learner differs naturally from a native pronunciation, and might not have been encountered lots by Whisper during the data training phase. Without confidence scoring or any information to understand why the system chose *fleur* in one case and *bleu* in the other, it is difficult to understand which linguistics parameters have contributed. The system is perceived as a black box for the users as one can not know which patterns and rules are applied by the system.

Using those systems as a basis for a linguistic analysis of case studies could be interesting. Unfortunately, it does not provide such information outside of the result.

Linguistic levels	Manual transcripts	Automatic transcripts
Pronunciation	la **langue étranger**	le **long est rejeu**
Prosody	co-douche	coudouche
Morphology	entend**ons**	entend**ant**
Syntax	je **toujours parle** le français	je **parle toujours** le français
Semantics	expériencer	expérimenter

Table 3: Examples of errors

7 Conclusion

As *Tancoigne et al.* states, if we consider that transcribing is already analyzing then delegating this work to a machine can be seen as problematic in a number of cases (Tancoigne et al., 2022). This study aims at specifying which elements have to be taken into consideration for using such technologies on learner corpora.

Our study presents some limitations. An important one is linked to conducting this research with only one ASR system: the discrepancies we showed are inherent on Whisper. It would be needed to compare the results obtained with other ASR systems.
Secondly, we focused our analysis on advanced beginners which are one specific group of learners and would also need to add different level groups to get a bigger perspective on the performances of ASR and of the possible usage of this technology for SLA studies.

Nevertheless, Whisper appears as a good starting point for a manual correction of transcriptions. Its hyper-correction is not suited for the degree of precision needed on the actual production of the speakers. Thus it can provide a first version of the transcription, aligned on the audio with timestamps, and correcting transcriptions rather than creating them from scratch can diminish cognitive overload.

One very interesting feature that we found is that Whisper get very good results on inaudible speech for human ears, and therefore allows to double check manual transcriptions and complete them. This is something that can be useful in order to complete some data for which the sound is inaudible for the transcriber or to choose between different transcription possibilities.

Those reasons conducted us to add a new import feature on our transcription and annotation tool *Sarramanka* (El Ayari, 2022) to be able to take Whisper transcripts as a starting point for manual check before any annotation process. Nevertheless the data would have been reviewed in totality and checked thoroughly to correct the elements transcribed in a native manner.

As we have discussed, transcription is a crucial part of SLA researches on speech and a crucial step that is the basis for any linguistic analysis. Therefore an automatic system could really be a very helpful tool for the study of those corpora, especially as there are many corpora open-source and available which have been documented, transcripted, annotated and analyzed. Those are precious resources that could be used to fine-tune ASR systems.

Therefore it is important to keep in mind that those systems can provide help and facilitate some treatments but that a human check is always needed in order to guarantee the quality of the data processed automatically.

8 Perspectives

It is needed to train the system on data from learners of French, but the question arises of the impact of source languages, which induce specificities concerning the acquisition of the pronunciation of the target language - French in this case. Next step is to train the model on learners' corpora with specific dataset matching the SLA issues we present.

We want to conduct a similar study on beginners' productions and on learners with different L1 on the VILLA corpus. The corpus is issued from the project ANR ORA *Varieties of Initial Learners in Language Acquisition: controlled classroom input and elementary forms of linguistic organisation*. The researchers observed the

acquisitional path for L2 acquisition of Polish with only 14 hours of exposure for learners from five different L1: French, Italian, German, British English and Dutch. This corpus will allow us to see the impact of pronunciation and accent on the automatic transcription provided, with a similar level of acquisition. It will also be interesting to see how efficient the system can be on beginners' productions - as they should be more distant from ASR systems' inside norm.

Next step will be to train the model on learner corpora with specific datasets matching the issues specific to SLA we have presented in Table 3. It would be really interesting to fine-tune Whisper or another ASR system like *wav2vec* (Baevski et al., 2020) on learner corpora depending on the L2 or on the L1. As we said before, those corpora can be considered like a poorly endowed language due to their specificities.

Acknowledgments

The project **TranSLA** has been founded within the support of *Paris 8 University* and the research laboratory *Structures Formelles du Langage*.

References

A. Baevski, H.Zhou, A.Mohamed, and M. Auli. 2020. wav2vec 2.0: A framework for self-supervised learning of speech representations.

N. Ballier, A. Meli, M. Amand, and J.-B. Yunès. 2023. Using whisper LLM for automatic phonetic diagnosis of L2 speech, a case study with French learners of English. In *Proceedings of the 6th International Conference on Natural Language and Speech Processing (ICNLSP 2023)*, pages 282–292. Association for Computational Linguistics.

S. Benazzo and M. Watorek. 2021. Transcription de corpus oraux d'apprenants débutants en français l2 : quelques enjeux théoriques. In *L. Spreafico, G. Bernini, A. Valentini J. Saturno (éds.) Superare l'evanescenza del parlato. Un vademecum per il trattamento digitale di dati linguistici*, pages 127–165. Bergamo: Sestante.

C. Blanche-Benveniste. 2000. *Approches de la langue parlée en français*. Paris: Ophrys.

R. Bley-Vroman. 1983. The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning*, (1):1–17.

A. van den Bosch, A. Content, W. Daelemans, and B. de Gelder. 1994. Measuring the complexity of writing systems. *Journal of Quantitative Linguistics*, 1(3):178–188.

V. Chanethom and A. Henderson. 2022. Alignment in ASR and L1 listeners' recognition of L2 learner speech: A replication study. In *15th International Conference on Native and Non-native Accents of English*, Łódź, Poland. Université de Łódź.

S. P. Corder. 1980. La sollicitation de données d'interlangue. *Langages*, (57):29–27.

R. Cumbal, B. Moell, J. Lopes, and O. Engwall. 2021. "You don't understand me!": Comparing ASR Results for L1 and L2 Speakers of Swedish. In *Proceedings Interspeech 2021*, pages 2021–2140.

L. Dekydtspotter, B. Schwartz, and R. Sprouse. 2006. *The Comparative Fallacy in L2 Processing Research*. 8th Generative Approaches to Second Language Acquisition Conferences.

S. El Ayari. 2022. Sarramanka, une plateforme outillée de transcription, d'annotation et d'exploration de corpus. In *8ème Congrès Mondial de Linguistique Française (CMLF)*, volume 138, page 10006, Orléans, France.

S. El Ayari and M. Watorek. 2021. Exploration outillée pour un corpus de productions orales des apprenants débutants en L2. In *Colloque "Influence translinguisitique : où en est-on aujourd'hui ?"*, Toulouse, France.

R. Ellis and G. Barkhuize. 2005. *Analysing Learner Language*. Oxford:Oxford University Press.

Calbert Graham and Nathan Roll. 2023. Evaluating openai's whisper asr: Performance analysis across diverse accents and speaker traits. *Cambridge Open Engage*.

W. Klein. 1989. *L'Acquisition de langue étrangère*. Paris: Armand Colin.

W. Klein and C. Perdue. 1997. The Basic Variety (or: Couldn't natural languages be much simpler?). *Second Language Research*, 13(4):301–347.

P. J. Lou and M. Johnson. 2020. End-to-end speech recognition and disfluency removal. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2051–2061. Association for Computational Linguistics.

B. MacWhinney. 2000. *The CHILDES Project: Tools for Analyzing Talk. 3rd Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

C. Perdue. 1993. *Adult Language Acquisition. Vol 1: Field Methods.* Cambridge University Press.

A. Radford, J. Xu Kim, Brockman T., McLeavey G., C., and I. Sutskever. 2023. Robust speech recognition via large-scale weak supervision. *Proceedings of the 40th International Conference on Machine Learning*.

L. Selinker. 1972. Interlanguage. *International Review of Applied Linguistics*, (10):209–231.

Elise Tancoigne, Jean Philippe Corbellini, Gaëlle Deletraz, Laure Gayraud, Sandrine Ollinger, and Daniel Valero. 2022. Un mot pour un autre ? Analyse et comparaison de huit plateformes de transcription automatique. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 155(1):45–81.

X. Wei, C. Cucchiarini, R. van Hout, and H. Strik. 2022. Automatic speech recognition and pronunciation error detection of dutch non-native speech: cumulating speech resources in a pluricentric language. *Speech Communication*, 144:1–9.

H. Yamasaki, J. Louradour, J. Hunter, and L. Prevot. 2023. Transcribing and aligning conversational speech: A hybrid pipeline applied to french conversations. In *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, Taipei, Taiwan.