Towards a Verb Profile: distribution of verbal tenses in FFL textbooks and in learner productions

Nami Yamaguchi[1,2], **David Alfter**[1], **Kaori Sugiyama**[2] and **Thomas François**[1]
[1] Université catholique de Louvain, Belgium
[2] Seinan Gakuin University, Japan
first.last@uclouvain.be, sugiyama@seinan-gakuin.jp

Abstract

Morphological inflection is known to be difficult to master for L2 learners. In this paper, we examine the state of the use of inflection in the verbal tense system among learners of French, and contrast it with the use in FFL textbooks. The objectives of our study are threefold: 1) To establish the distribution of verbal tenses on French textbooks in an automatic way, in order to obtain the first fully empirical and extensive resource on French verbal tenses; 2) To objectively describe the use of verbal tenses by learners of different CEFR levels; 3) To identify the tenses that learners struggle with. Through the description of the use of the tenses in the learners, we found that they had difficulty with the past perfect indicative, even at advanced levels. The proposed Verb Profile summarizes which tenses should be understood at which level, and as such can guide teachers and learners, as well as help pinpoint tenses that learners are underperforming on.

1 Introduction

In second language acquisition (SLA), the construct of complexity is frequently used to measure learner development (Bulté and Housen, 2012; Pallotti, 2015) and has been mainly addressed through lexicon and syntax. Measures of morphological complexity, on the contrary, have been overlooked to some extent, as argued by De Clercq and Housen (2019). Yet studies remind us that learning morphological inflection remains a challenge even for advanced learners (DeKeyser, 2005; Larsen-Freeman, 2010; Lardiere, 2016), and, for morphologically rich languages such as French, the use of morphological complexity measures seems even more justified (Brezina and Pallotti, 2019).

Furthermore, beyond the field of complexity, there are a number of studies and theories that focus on morphological development of learners and discuss the development of language and its order of acquisition (Dulay and Burt, 1974; Pienemann, 1998; Bartning and Schlyter, 2004). For French, a large part of the morphological complexity lies at the level of the verbal system and the inflectional morphology of verbs (De Clercq and Housen, 2019, p. 76). Among the many features of verbal inflection, the verb tense is one of the main components of the complexity of the system. Therefore, in this study, we focus on the acquisition of morphological inflection in relation to verbal tenses in learners of French as a foreign language (FFL).

Here is a list of the tenses existing in French (Grevisse and Goosse, 2007) and the moods that invariably accompany them:

- Indicative: present, imperfect, simple past, past perfect, double compound past, pluperfect, double compound pluperfect, anterior past, simple future, anterior future or future perfect, and double compound future perfect;
- Conditional: present and past;[1]
- Imperative: present and past;
- Subjunctive: present, past, double compound past, imperfect, and pluperfect;
- Infinitive: present, past, and double compound past;
- Participle: present, past, past perfect, and double compound past;
- Gerund: present and past.

For second language learning, it is clear that one cause of difficulty is related to L1 interference;

[1] Grevisse and Goosse (2007, p. 980) classify the tenses of the conditional mood within those of the indicative, following the tendency of the linguists. However, in this article, we distinguish them, because this is normally the case in the FFL textbooks.

Nami Yamaguchi, David Alfter, Kaori Sugiyama and Thomas François. Towards a Verb Profile: distribution of verbal tenses in FFL textbooks and in learner productions. *Proceedings of the 11th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2022)*. Linköping Electronic Conference Proceedings 190: 123–142.

however, examinations of learners' actual use of the language have revealed that many errors come from the target language itself, not from the L1 (Richards, 1970). Since then, the focus has been on the similarities that L2 learners have, regardless of their L1 (Spada and Lightbown, 2020, p. 118).

There are many theories about the learning steps that learners are generally expected to follow. One of the most representative theories is the *Processability Theory* (PT) proposed by Pienemann (1998). It is a theory that formally predicts which structures can be processed by the learner at a given level of development based on human psycholinguistic constraints on language processing. Table 1 presents the order of development according to this theory:

Order of development	Processing procedures	Structural outcome
5	Subordinate clause procedure	Main and subordinate clause
4	S-procedure	Interphrasal information exchange
3	Phrasal procedure	Phrasal information exchange
2	Category procedure	Lexical morphemes
1	Word or lemma access	Words

Table 1: Processing procedures and their structural outcome according to PT (based on Tables 1 and 2 in Pienemann and Håkansson (1999))

However, it is not possible to explain in detail the sequence of acquisition of each linguistic phenomenon (or grammatical rule), because PT only has five steps and therefore lacks granularity for this purpose. For example, if we apply this theory to verbal tenses, which are the focus of this paper, simple tenses – composed of a single verb – belong to the second stage, because they are processed in the word category. On the other hand, compound verbs, which are composed of an auxiliary and a verb, are at stage 3 (sentence level, beyond the word category). Therefore, according to PT, simple verbs are acquired at an earlier stage than compound verbs. The fact that the French indicative present is easier than the indicative past perfect is indeed consistent with FFL teachers' practices. However, it is unlikely that all simple tenses, such as the simple past, are easier than the past perfect. In addition, PT does not tell us which tense is acquired first among simple or compound tenses.

Pragmatically, what L2 teachers and learners are interested in is to know which linguistic elements should be mastered at what stage of learning. After the introduction of the Common European Framework of Reference for Languages (CEFR; Council of Europe 2001) in 2001, this framework became widely used in Europe as well as the proficiency scale of the CEFR. The CEFR scale provides "can-do" descriptors for the five skills (listening, reading, two subcategories for speaking, and writing) spread over six levels (A1 to C2). However, since the CEFR was developed to be compatible with different European languages, these "can-do" descriptors remain rather general and do not specify the details corresponding to each language (Hawkins and Buttery, 2010, p. 2). As a result, a number of research projects have attempted to link precise lexical or grammatical elements to the CEFR scale for various languages, as outlined in Section 2.

In the present study, we attempt to explain the acquisition of morphological inflections of verbal tenses by FFL learners. We use an empirical approach that relies on two datasets (textbooks and learners essays) and natural language processing (NLP) techniques to automatically annotate the large amounts of data. We hope that this approach will lead to more robust and generalizable results. Our research questions are:

1. **In the corpus of FFL textbooks, which verb tenses appear at which CEFR level?**
   Based on analysis of a textbook corpus, we will study the distribution of tenses according to the CEFR levels. Then, we will propose a "Verb Profile", a resource which will be the first fully empirical and extensive resource describing the distribution of verbal tenses in FFL pedagogical texts.[2]

2. **How do learners at different levels use verbal tenses?**
   In the long term, we plan to also establish the

[2]The name "Verb Profile" was chosen based on existing *grammar profiles*, of which it can be seen as a subcomponent.

"Verb Profile" of learners based on a large amount of written production data. In this paper, as a first step, we attempt to describe the use of tenses by FFL learners using a manual annotation of a small corpus of written productions. We will also discuss the challenges of using NLP for the automatic identification of verbal tenses.

3. **Which tenses do learners have difficulties with?**
   Inspired by the CEFR-J project (Tono, 2013), in case the learners have made some errors with the tense forms, we will also annotate the form that they should have written. These annotations will allow us to identify the tenses causing the most difficulties to learners.

The next section (Section 2) describes previous work on grammatical profiling of learners and on the acquisition steps of the morphology of verb tenses. It is followed by Section 3, which describes our research object, the corpus used, and the two annotation methods (automatic and manual). Section 4 presents the results of the textbook and learner data analysis respectively. Then, in Section 5, we enter into a discussion of the results and future research perspectives, and we conclude this paper in Section 6.

2 Related work

In both SLA and NLP, English is the dominant language in research, and this is also the case for grammar profiling projects. We can therefore mention several projects for English, such as the Core Inventory for General English by British Council (North et al., 2010), the English Grammar Profile[3] (O'Keeffe and Mark, 2017), the Pearson's Global Scale of English[4], and CEFR-J (Tono, 2013).

For French, there is a limited amount of work, including the study of Bartning and Schlyter (2004), the reference level descriptors (RLD) of Beacco and his collaborators (Beacco, 2008; Beacco and Porquier, 2007; Beacco et al., 2011, 2004; Riba, 2016) and the reference level descriptors of North (2015).

Bartning and Schlyter (2004) are among the first that summarized the acquisitional stages in the style of a grammatical profile, by analyzing a corpora of Swedish FFL learners' oral production with an empirical perspective. Based on these stages, they described French grammatical phenomena from the morphosyntactic point of view, specifying the phenomena expected at each developmental stage, from beginners (level 1) to Quasi-natives (level 6).

As mentioned in the introduction, it is worth remembering that the CEFR descriptors lack a detailed description of acquisitional stages for the linguistic phenomena. To overcome this problem, the Council of Europe, which published the CEFR, also supported the creation of Reference Level Descriptions (RLDs) with the aim of offering more detailed language guides (Abel 2014, p. 112; Dürlich and François 2018, p. 873). The French version of the RLD is the referentials of Beacco and collaborators (Beacco, 2008; Beacco and Porquier, 2007; Beacco et al., 2011, 2004). Their RLD describe, for each level of the CEFR, the linguistic phenomena that are supposed to be mastered, and organize them within several distinct categories (basic lexicon, specialized notions, syntactic structures, phonemes, graphemes, functions, etc.). However, in the end, the knowledge of experts seems to often have been the primary criterion influencing the decision to assign a given language item to a given level (Dürlich and François, 2018).

According to North (2015, p. 5), what we teach, what learners can do, and what we measure in exams are not the same. Beacco et al.'s RLD have not sufficiently resolved the teachers' question about what content to teach at what level of the CEFR (North, 2015, p. 5), as they focus more on what learners are supposed to be able to do. North's work, then, focused on activities within the classroom. With the goal of making the CEFR accessible to teachers and learners, he established the linguistic inventory of key content for levels A1 to C1. These key elements were determined through the analysis of several types of data: the CEFR descriptors, some curricula, the French RLD and other similar sources, as well as a survey addressed to FFL teachers. By outlining these key contents, North's work provides teachers with support for selecting classroom activities and learners with support for independent learning.

In Appendix A, we have summarized the acquisitional stage of the French moods and tenses

[3] http://englishprofile.org
[4] https://www.pearson.com/english

as described in these three studies. Based on the comparison of these resources, we can say that Bartning and Schlyter (2004)'s study is interesting in that the results are based on actual learner data. However, because the tenses discussed are not comprehensive, we cannot see the overall picture of verb acquisition for French. In addition, the results are not aligned with the CEFR scale, so they do not address recent needs. In contrast, being based on the CEFR, Beacco and North's RLD are more applicable to practical situations. In addition, they cover many elements. Nevertheless, their inventories are based on various sources of information, including expert or teacher opinion, but not on large corpora. References of this nature are very informative in some respects, but it is not clear whether that they accurately reflect usage in textbooks and by learners. Thus, to the best of our knowledge, there is no study that is at the same time data-driven, comprehensive, and based on the CEFR scale. This study attempts to fill this gap by applying NLP to this challenging issue.

3 Methodology

In this section, we first describe the study proper (3.1). We then give an overview of the corpora used (3.2), the automatic annotation pipeline (3.3), and conclude with a description of the manual annotation process (3.4).

3.1 Overview of the study

Our study focuses on the use of verbal tenses in French. It is therefore necessary to define our object of study, that is, the tenses that will be the subject of our analysis. Among the tenses we presented in the introduction, the double compound tenses, the participles and the gerund have been excluded from our study for the following reasons:

- Double compound tenses: They are almost never used and are not taught in French textbooks.
- Participles: Present and past participles belong to one of the grammatical categories that are difficult to classify because of the ambiguity between the participle and the adjective, when they are in epithet (after nouns) or predicate position (after the verb *être* 'to be').[5] Moreover, by performing tests that we will detail later, we observed that parsers/taggers sometimes detect nouns that end in *ant* (e.g. étudiant 'student', enseignant 'teacher', etc.) as present participles, which would bias the results. Therefore, the present participle was also excluded.
- Gerund: The gerund consists of the preposition *en* and a present participle. It is difficult to find the link between these two elements automatically. This is an area for improvement that can be explored in the future.

Thus, we look at 18 tenses in this paper.

3.2 Corpus

In this study, we use two corpora: the first one is a FFL corpus of textbooks, and the second one is a French learner corpus.

The textbook corpus is identical to the one used in the study by Yancey et al. (2021). The corpus contains 20 textbooks published since 2015, as well as the *Annales du niveau C* publicly available on the Internet.[6] The selected texts target reading comprehension tasks, and the CEFR level assigned to them is that of the textbook from which they were taken. In total, the corpus contains 2769 texts distributed over five levels (A1 to C) – levels C1 and C2 having been merged –, for a total of 369,170 words, as detailed in Table 2.

Level	Texts	Words	Books
A1	764	48,639	6
A2	865	77,255	6
B1	507	82,728	4
B2	345	81,171	3
C	288	79,377	3
Total	2769	369,170	22

Table 2: Number of texts, words and textbooks by level in the textbook corpus

The learner corpus includes written productions from the TCF exam (Test de connaissance du français).[7] In this exam, candidates respond to three tasks, which are varied in topic and can be given to candidates of any level. Such a corpus

[5] "The past and present participles, which by their nature can be used as epithets, are often confused with adjectives" (Grevisse and Goosse, 2007)

[6] https://www.france-education-international.fr/diplome/dalf/exemples-sujets http://www.delfdalf.fr/exemples-sujets-dilf-delf-dalf.html

[7] The data was obtained through an agreement with France Education International and currently cannot be published.

is advantageous in that we can compare data produced by learners of various levels on the same tasks. First of all, it should be noted that each answer was evaluated by two or three trained evaluators, which provides a reliable CEFR level for each production. Moreover, the corpus also includes the candidates' usual language. We selected texts written by learners of five common but different languages, namely Arabic, Chinese, English, Russian and Spanish. Then, we extracted prototypical productions, i.e. those productions whose levels assigned by the two evaluators are the same and whose rounded Multi-faceted Rasch Analysis values correspond well to the level assigned by the evaluators.[8] Concerning the levels, having combined the C1 and C2 levels which were poorly represented in some of our five languages, we obtain five different levels (A1, A2, B1, B2 and C), following the example of the textbook corpus. Finally, as the topic of the tasks can influence the use of verbal tenses, we controlled for the number of tasks oriented towards the past (e.g. telling about one's last weekend), the present (e.g. talking about one's preferences about something such as how to shop (online or on the spot)) and the future (e.g. proposing an activity to friends). In concrete terms, in the 25 prepared sub-corpora (i.e. five levels for each of the five common languages), we randomly selected texts from a larger corpus and retained texts until we had two per task type (past, future and present).[9] Table 3 gives an overview over the learner corpus used.

Level	Texts	Words
A1	26	1253
A2	30	1452
B1	30	2793
B2	30	3002
C	30	2943
Total	146	11,443

Table 3: Number of texts and words by level in the learner corpus

[8]Multi-faceted Rasch Analysis (Linacre, 1989) is used to calculate an adjusted score for each production which takes into account rater severity and test taker competence.

[9]On the lowest level A1, there were not enough future-oriented tasks. This is why the number of texts in this level is 26 instead of 30.

3.3 Automatic annotation

In order to process large amounts of data, we created a script that identifies verbal tenses automatically based on several automatic language processing tools that we evaluated. We first performed a preliminary evaluation with five popular parsers and taggers, namely Stanza (Qi et al., 2020), UDpipe (Straka and Straková, 2017), spaCy (Honnibal et al., 2020), TreeTagger (Schmid, 1994), and RNNtagger (Schmid, 2019). In this preliminary study, we noticed that both UDpipe and RNNtagger failed at detecting several verbs. TreeTagger seemed promising, but its main limitation lies in the fact that it is a tagger and not a parser (i.e., it lacks the dependency information which is necessary to detect compound tenses). Following this preliminary analysis, we performed a more detailed evaluation of Treetagger, Stanza and spaCy. For this purpose, we prepared 10 sentences for each tense to be detected. The sentences were selected from French grammars (Asakura, 2002; Beacco et al., 2004; Cherdon, 2005; Grevisse and Goosse, 2007; Machida, 2015); we choose sentences that were as diverse as possible at the lexical level (both verbs and auxiliaries), at the usage level (complex sentences, as well as basic sentences that are suitable for language learners) and at the syntactic level (with or without adverbs such as negation, and inversion).

However, none of the taggers and parsers used in this study can detect French compound tenses, and, except for a system described in de Alencar (2017) that focuses on identifying the past perfect and passive constructions, but seems to be unavailable at the time of writing, we are not aware of previous work focusing on the detecting of composed tenses in French. Hence, we created a custom script that identifies compound tenses based on dependencies and part-of-speech information. The script identifies dependencies between auxiliary verbs and participles and uses a set of rules to derive the composed tense. While not a focus of this study, we included the detection of passive tenses, since they sometimes resemble active tenses and thus might lead to erroneous counts.

Based on this comparative evaluation of the three tools, we chose spaCy as main parser, TreeTagger for the present conditional and imperfect subjunctive, and Stanza for the past imperative. Table 4 shows the recall, precision and F1 score of the final script. The script can be accessed through

a dedicated web interface.[10] The low precision for ind_pres and sbj_pres may be due to the fact that many verb forms of these tense have identical surface forms (e.g., *qu'il marche*-SBJ_PRES and *il marche*-IND_PRES). Furthermore, we noticed that sbj_pres was often mistagged as sbj_imp.

Tense	Recall	Precision	F1 score
Simple tenses			
ind_pres	1	0.56	0.71
ind_imp	1	1	1
ind_ps	0.8	1	0.89
ind_fs	1	1	1
cnd_pres	1	1	1
impe_pres	0.7	0.85	0.78
sbj_pres	0.8	0.5	0.61
sbj_imp	0.9	1	0.95
inf_pres	1	1	1
Compound tenses			
ind_pc	1	1	1
ind_pqp	0.9	1	0.95
ind_pa	0.9	1	0.95
ind_fa	1	1	1
cnd_pass	1	1	1
impe_pass	1	1	1
sbj_pass	0.9	0.9	0.9
sbj_pqp	1	1	1
inf_pass	1	1	1

Table 4: Precision, recall and F1 score on the different tenses

3.4 Manual annotation

We will first perform the annotation of verbal tenses in both corpora using our script. However, since automatic language processing tools are developed on the basis of well-formed data, it is to be expected that learner corpora, due to the inclusion of errors, will lead to annotation errors (Granger, 2011; Štindlová et al., 2011; Krivanek and Meurers, 2013; Rubin, 2021; Volodina et al., 2022). Therefore, we decided to also manually annotate the learner corpus.

According to Volodina et al. (2022, p. 152), a common pitfall when annotating learner corpora is to start by annotating what the learners meant, which is subjective in nature, rather than objectively describing what was used. Therefore, we started with manual annotation by scrupulously respecting the forms that the learners produced. That is to say, when we found a correctly written verb whose conjugated form exists, we annotated this verbal tense in square brackets ([ ])[11], regardless of whether its usage in relation to the context is correct or not. In this step, we did not take into account the learners' intention in order to capture only what they are able to produce.

(1) Je vous écris [ind_pres] pour vous informer [inf_pres] que la fête du sport aura [ind_fs] lieu dans ma ville le 01/04/2022. (C-fut-chi2)[12,13]

As has been done in the CEFR-J project, it would also be interesting to clarify what the learners wanted/needed to produce. In addition to the annotation that was based purely on form, we chose to include additional information. In some cases, it is clear that the verb form used was not the one that the learners were trying to use. That is, when the verb is in a form that exists but its usage is grammatically incorrect, due to errors such as a spelling mistake and/or a lack of grammatical competence, we added the error label *E!* or *E*. The former was added when the conjugated form that the learner wanted/needed to write was identifiable. In this case, we added next to it the tense they would have wanted/needed to write in curly braces ({ }).The second was used when the learner's intention was not certain or when the verb has no subject, except in the imperative form. As explained above, the present and past participles are not included in this study. However, it happens that the learner writes a verb in the participle when they probably wanted to form another verbal tense. In this case too, we annotated it with these labels.

(2) Après j'ai [ind_pres_E!] fais [ind_pres_E!] {ind_pc} le longue couries (B1-pre-ang1)

(3) Bonjour, moi ecrîte un proposer [inf_pres_E] pour tu. (A1-fut-ang1)

[10] https://cental.uclouvain.be/verbprofile

[11] See Appendix B for tense abbreviations used in the annotation.

[12] The label identifies the learners; it is composed of their level (A1, A2, B1, B2, C), the task orientation (*pas* for past, *fut* for future and *pre* for present), and their everyday language (*ang* for English, *ara* for Arabic, *chi* for Chinese, *esp* for Spanish and *rus* for Russian, followed by the id (1 or 2).

[13] Since most of the presented examples contain errors that make their translation difficult to impossible, we have opted not to gloss the sentences in English.

(4) J'aime [ind_pres] mangé [prt_pass_E!] {inf_pres} dans le restaurant familial (B2-pre-ang2)

Sometimes, a conjugated form corresponds to more than one tense. This is mainly the case of ambiguity between the present indicative and the present subjunctive. The subjunctive is mainly used with the conjunction *que*. When this ambiguous case occurs without such markers in the situations mentioned just before (annotation *E!* or *E*), the present indicative was noted as a temporary annotation before the correction. The reason is that it is evident from previous research and our textbook corpus results presented later that the present subjunctive is taught and learned later and is much less frequent than the present indicative.

(5) je aime [ind_pres] bien reste [ind_pres_E!] {inf_pres} avec soliei du campagne (A2-pas-ang2)

When a misspelled word is found that can be assumed to be a verb, we have annotated ∅ plus the correction between curly braces ({ }). This annotation is only used when the learner's intention is deemed sufficiently certain. In the opposite case, i.e. when we cannot determine the tense the learner has used, we used the annotation <E>.

(6) Donc j'ai [ind_pres_E!] règardè [∅] {ind_pc} le netflix (B1-pas-ang-2)

(7) Nous fair <E> le skis fon avec mes enfants. (A1-pre-rus1)

In cases where it is impossible to tell whether a word is a verb or another part-of-speech, we added the label <NV>.

(8) L'ecole est [ind_pres] pas lion pour enfant, just marche <NV> (A1-pre-chi1) [The word *marche* can be a noun or a verb.]

Our correction (between { }) acts on the change of form and mode of the verbs if we can formulate a hypothesis based on what the learners have written. The choice of tense is linked to the writing style and it is therefore delicate to determine whether a tense is appropriate or not (e.g. use of the present tense instead of the past tense). Therefore, in general, our correction does not change the tense that the learners used.

As mentioned above, the passive is not included in our analysis, so we had to distinguish between passive and active cases. In situations where it was difficult to judge whether it was a passive or active voice, we asked three experts to decide. These experts are native French speakers and have already worked on projects that also encountered this difficulty. They annotated one of the two voices, following the annotation guide we had prepared based on the definition of voices according to the Bon Usage (Grevisse and Goosse, 2007) and the annotation guide of the French Treebank (Abeillé et al., 2003).

4 Results

In this section, we first describe the results from the textbook corpus (4.1). We then focus on the learner productions (4.2), including an in-depth analysis of both the automatic (4.2.1) and manual (4.2.2) annotation.

4.1 Textbook corpus

Table 8 in the appendix presents the results of the automatic analysis of the textbook corpus. To attach a level to a phenomenon, several approaches have been used, like the first occurrence (Gala et al., 2014; Alfter et al., 2016), but also threshold methods (Hawkins and Filipović, 2012; Gala et al., 2014; Alfter et al., 2016), and since we are dealing with learner language, observing a phenomenon only once or twice at a certain level is not sufficient to claim that it is of this level (Hawkins and Filipović, 2012; Alfter, 2021). In order to assign a level to each tense, we looked both at frequency and dispersion: we only took into account frequencies of tenses that occurred in *all* textbooks of that level; indeed, if only one textbook introduces a tense at a certain level, it is less likely to be globally of this level than if multiple/all textbooks introduce it. For frequency, we explored threshold methods, with thresholds of 1,3,5,10, and found that for our corpus, a threshold of 5 gives consistent results.

For the tenses that have not been sufficiently covered in some textbooks up to level C, namely the anterior past, the anterior future, the past imperative, the past subjunctive, the imperfect subjunctive, and the pluperfect subjunctive, it is very unlikely to find them in learner production tasks, and we can assume that their learning is a very low priority. For all the other tenses, we can observe that they are used a certain number of times until level B1, and more prominently at level B2.

The proposed Verb Profile based on the number of occurrences in the textbooks is shown in Table 5. Light colored cells indicate levels at which the tense may be observed sporadically, while dark shaded cells indicate levels at which the tense should be understood by learners of the corresponding level.

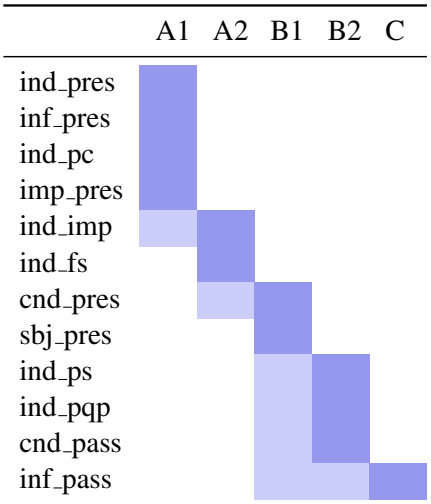

Table 5: Proposed textbook verb profile

4.2 Learner productions

In this section, we first describe the automatic analysis of learner productions, followed by the manual analysis of learner productions.

4.2.1 Automatic analysis

After the textbook corpus analysis, we performed an automatic analysis of the learner corpus. Several problems were identified, especially in the lowest CEFR level productions. This is due to the fact that the syntactic parser is misled by learner errors. By trying to interpret the texts despite its errors, our script will try to recognize as verbs words that are not, but that are in the expected position for a verb. In the following examples, the tense in parentheses is the one identified by the script.

(9) elle ne pa (ind_pres) de grave. (A1-pas-ang-1)

This "feature" causes other misidentifications. For example, it tends to judge words ending in *er* or *ir* as present infinitives. This error is probably caused by the fact that verbs of the first and second groups, which represent the majority of French verbs, end with these suffixes respectively.

(10) ôûî je puex alleer (inf_pres) al aniversarie de paula (A2-fut-esp1)

(11) j' ai more rir (inf_pres) pour lui (A1-pas-ara1)

We have observed this same phenomenon for other tenses. For example, when there are erroneous words that end with a suffix of a certain verb conjugated to a certain tense, the script can identify that tense, even though the word does not exist. Here are some examples that were misidentified as *simple past*, whose conjugated form of the first group verbs end with *ai*, *as*, *a*, *âmes*, *âtes*, and *èrent* depending on the person.

(12) Bonjour Maris ! sava (ind_ps) toi (A1-pas-ang2)

(13) j aimrai (ind_ps) bien passe mon anniversaire a la maison (A1-pre-ara2)

(14) Les doctors dirent (ind_ps) regarder le télé deux heures par jours par plus, (A2-pre-rus2)

Misidentifications of the *past perfect* have also been frequent. This tense consists of the auxiliary *avoir* 'to have' or *être* 'to be' conjugated in the present indicative and a past participle verb. However, when the word that follows the auxiliary is close to or identical with a certain verb form, the script may identify it as *past perfect*.

(15) çava matte noi, j'ai trie (ind_pc) mal lu venti. (A1-pas-ang1)

(16) Après j'ai fais (ind_pc) le longue couries (B1-pre-ang1)

(17) Et pour le dessert j'ai preparer (ind_pc) gâteux au framboise au créme anglaise. (A2-fut-rus2)

Likewise, the scripts assign a certain tense to verbs even if the accent is missing or on the contrary with an accent added in a wrong way.

(18) la concentration à pris (ind_pc) place. (A2-pas-esp2) [expected form: a pris]

(19) je vousley achèter (inf_pres) une pair de nouveau chausseurs. (A2-fut-ang2) [expected form: acheter]

(20) sa me fe reflechir (inf_pres) bocou (A1-pas-esp2) [expected form: réfléchir]

Thus, the script tends to infer irrelevant results by interpreting erroneous data, which would lead to an over-assessment of learners' results. To clarify the picture, we performed annotation manually as detailed below.

4.2.2 Manual analysis

The results from the manual analysis are presented in Table 9 in the appendix. Colored zones reflect the results from Table 8.

The percentages of each verbal tense were calculated on the basis of the numbers found in **Total 1**, which are the sums of all correctly conjugated verbs. The "other" values correspond to the numbers of words we labeled ∅, <E>, <NV>as well as words accidentally formed as past/present participles. Except for <NV>labels, which are infrequently present, all other words that are classified as "other" could be considered errors. The percentages found in the "others" row were calculated on the total numbers including these errors, i.e., the **Total 2** row. It is important to note that this percentage is considerably high at the lower levels. In particular, at level A1, we see that half of the verbs that learners tried to produce are there, and were not included in the first part of the table, the one showing the distribution of tenses.

We can observe tenses that are present in the textbooks, but which are not produced by the learners, even at the higher levels, namely the past simple, the indicative pluperfect, the anterior future, the past conditional and the past infinitive. This does not necessarily mean that they are not acquired, but it may simply mean that they are used less frequently in the everyday context corresponding to the tasks that the TCF exam calls for. For example, in tasks describing a past weekend, the imperative is expected to appear less frequently. In addition to the influence of opportunity, it is not excluded that learners avoid certain grammatical elements as a consequence of an avoidance strategy (Granger, 2011). As O'Keeffe and Mark (2017, p. 462) point out, zero occurrences in the native speaker data can be interpreted as resulting from choice, whereas in the learner data, this can be seen more as due to lack of proficiency. It is therefore important to be careful about the interpretation of the absence of a feature in a learner corpus. To know when they are learning the tenses, we will need other tests such as a grammaticality test. But here, our results are interpretable in the sense that we were able to observe the use of tenses for written production in a context with few constraints, thus with freedom of choice for the learner.

We see several uses of the present conditional and present subjunctive at a level below that expected according to the Verb Profile. For the first of these two tenses, all five uses were relevant. However, they were only *je voudrais*, a boilerplate that shows a modal value for politeness. This is consistent with what previous work had mentioned (Bartning and Schlyter, 2004; Beacco et al., 2004; Beacco and Porquier, 2007; Beacco, 2008; Beacco et al., 2011; North, 2015).

(21) Je voudrais [cnd_pres] visiter à Paris, (A1-fut-esp1)

(22) Je voudrais [cnd_pres] manger les repas typics (A1-fut-esp1)

(23) je voudrais [cnd_pres] faire amies là bas, (A1-fut-esp1)

(24) Je voudrais [cnd_pres] participé à activité pour marcher en samedi, (A2-pas-chi2)

(25) tu voudrais [cnd_pres] participé avez moi? (A2-pas-chi2)

Regarding the other tense, the present subjunctive, we observed three uses at the A1 level, whereas this tense was used very little in the textbooks at this level and not often at the next level either. These three uses are as follows:

(26) j' ai [ind_pres_E!] sorte [sbj_pres_E!] {ind_pc} avec Mohammed a jaddhe (A1-pas-ara1)

(27) il set sorte [sbj_pres_E] son ficag Parsra les basa bonne (A1-pas-ara1)

(28) J' aime ma ville ,paceque avec juli montange , vive [sbj_pres_E!] {inf_pres} est facile , magasins es a cote ,la lac pas trop lion , (A1-pre-chi1)

In fact, these three uses are the results of a spelling error and/or chance. We cannot therefore consider that the learners were able to produce it.

In the annotation so far presented, we annotated by respecting the form of the verbs that the learners wrote. This allowed us to know what they wrote without overestimating their skills, which was one of the problems in the previous results

with the automated approach. Moreover, thanks to this annotation, we were also able to better identify erroneous verbs by level, which was not the case in the automatic annotation. On the other hand, as we have just shown via the three examples of incorrect use of the present subjunctive, there are sometimes cases where verbs are not assigned to the correct verbal tense. However, it would be interesting to know what the learners wanted to write. This would allow us to clarify the difficulties they have in producing certain forms. Therefore, we prepared another table that was modified by the correction made with our estimation. Table 10 in the appendix shows the results of our correction hypotheses. As for Table 9, the colored areas reflect the results of Table 8.

Here, *correction* refers to the fact that we have removed the number of verbs annotated with *E*, *E!* and ∅. Moreover, for the last two, it is the learner's intended tense (and not the tense detected in the previous manual annotation) that is counted in Table 10.

Table 6 below shows the percentage change between the original and corrected annotation. For example, -11.22% in the present indicative in level A1 represents the percentage difference between the original table (Table 9, 73.50%) and the corrected table (Table 10, 62.29%).

	A1	A2	B1	B2	C
ind_pres	-11.2	0.3	-3.4	-3.99	0.2
ind_imp	0	-1.44	-0.15	-0.57	0.27
ind_ps	0	0	0	-0.22	0
ind_pc	5.15	7.31	3.13	2.09	-0.02
ind_pqp	0	0	0.39	0.2	-0.03
ind_pa	0	0	0	0	0
ind_fs	-0.28	-0.33	-0.56	0.36	-0.13
ind_fa	0	0	-0.02	0	0
cnd_pres	0.29	-0.16	-0.19	0.01	-0.08
cnd_pass	0	0	-0.02	-0.03	0
imp_pres	0	0.47	0.6	0.54	0.18
imp_pass	0	0	0	0	0
sbj_pres	-1.99	0.47	-0.32	0.09	0
sbj_pass	0	0	0	0	0
sbj_imp	0	0	0	0	0
sbj_pqp	0	0	0	0	0
inf_pres	8.05	-6.61	0.57	1.54	-0.38
inf_pass	0	0	-0.02	-0.04	-0.01

Table 6: Change in percentage before and after correction

We would like to draw attention to the past perfect (*passé composé*), whose numbers generally increase even at the higher levels, meaning that learners wanted to use and should have produced more past perfect but failed to do so. We have therefore studied the case where learners failed to produce the past perfect although they intended to do so.

As mentioned earlier, the past perfect is composed of the auxiliary *avoir* 'to have' or *être* 'to be' conjugated in the present indicative and a verb in the past participle. At A1 level, they had construction problems where the auxiliary was missing.

(29) je parti [prt_pass_E!] {ind_pc} week-end à la compagne chez ma grand parents. (A1-pas-ang2)

(30) Nou bian sortie [prt_pass_E!] {ind_pc}. (A1-pas-rus1)

Être and *avoir* are verbs that are learned from the beginning. From level A2 on, the construction is stabilized. The auxiliary was present and the learners were generally able to conjugate it correctly. However, we found that they had difficulty conjugating the second part, the past participle.

(31) les enfants se sont [ind_pres_E!] amuser [inf_pres_E!] {ind_pc}, (A2-pas-ara1)

(32) pour le dessert j'ai [ind_pres_E!] preparer [∅] {ind_pc} gâteux au framboise (A2-fut-rus-2)

(33) parceque j ai [ind_pres_E!] remarquer [inf_pres_E!] {ind_pc} (B1-pas-ara1)

(34) s'il y a quelqu'un qui dèja a [ind_pres_E!] connais [ind_pres_E!] {ind_pc}, (B1-pre-esp2)

(35) le film dont tu m'a [ind_pres_E!] parler [inf_pres_E!] {ind_pc} la semaine dèrnière. (B2-fut-ara1)

(36) Le chef nous a [ind_pres_E!] preparé [∅] {ind_pc} le plat japonais (B2-pas-rus1)

We can see that it was the inability to correctly conjugate the past participle that prevented the realization of the past perfect. In the first manual annotation, in Table 9, we annotated the well-conjugated auxiliary as being in the present indicative instead of assigning it to the past perfect.

This partly explains the drop in the percentages of the present tense after the correction. In addition, as we see in the examples above, there are many cases where the present indicative or present infinitive, the tenses we learn right at the beginning of the learning process (see Table 5 and 8) was used in place of the past participle. This also contributed to the decrease for these two tenses. Moreover, the proportion of errors in relation to the total number of verbs in the past perfect tense decreases as learners progress. But it is important to note that this type of error is still present at the higher levels.

Comparing the numbers of different levels in Table 6, the decrease of the present indicative tense in A1 is noticeable. In order to create Table 10, we have deleted error-labeled verbs to avoid counting verbs whose verbal tense used is too difficult or impossible to estimate in its context. The written productions of low level learners are not always comprehensible because of incorrect construction and wrong words. At A1 level, this tendency was obviously pronounced and a large proportion of verbs in the indicative present tense labeled as errors were observed. In addition to the difficulty of the past perfect tense, this could be a factor in the marked decrease. It is interesting, however, that even after eliminating the inappropriate use of the present indicative tense, its presence remains dominant at low levels and decreases as learners' level increases. This is consistent with the trend observed in the textbook data.

5 Discussion

Our automated annotation of a corpus of FFL textbooks made it possible to create a Verb Profile based on a large amount of data. It is an indicator that represents a form of consensus in the teaching of FFL, as it was created by considering the number of occurrences of verbal tenses in a large sample of textbooks, which may have different characteristics and objectives. The Verb Profile clearly indicates which elements are taught at which level. It can therefore be useful for teachers and those creating computer-based learning systems, such as an intelligent tutoring system, to select texts and the right tenses to cover, and to think about how much time to spend on explanations.

From a didactic point of view, we found that the indicative past perfect continues to cause errors from the beginning of learning to a fairly high level in learners. As our study has validated, it is a tense used quite frequently in the textbook corpus, as well as in the learner corpus. It is therefore necessary to teach it strategically to learners. For example, A1 learners were found to have difficulty producing the correct form of the past perfect. Therefore, it would probably be effective to offer them tasks that focus on its structure. From A2 onward, their difficulty is with the second verb, which is supposed to be conjugated as a past participle. Multiple-choice questions requiring them to select the past participle from several options would allow learners to practice the correct conjugation. Later, tasks that require them to spell the verbs themselves would further anchor their use.

To see how our results relate to existing studies, we compared our textbook and learner profiles to previous studies based on the CEFR, i.e., Beacco et al. (2004); Beacco and Porquier (2007); Beacco (2008); Beacco et al. (2011) as well as North (2015). Specifically, we checked whether the first level in which each of the two references marks a given tense as acquired corresponds (1) for the textbooks, to the first level we indicated with the dark shade in Table 8, and (2) for the learners' productions, to the first level in which we checked the usage of a given tense by five learners in Table 10 after correction. For both the textbook profile as well as the learners' profile, we find that they are closer to North than to Beacco. For our textbook profile, this trend makes sense, as North's inventory is more oriented towards reception. For learner production, on the other hand, we expected it to be closer to the inventory of Beacco et al. that describes the acquisition of tenses from a production point of view. We therefore need a more detailed interpretation of our results and also of these referentials.

Finally, from a NLP perspective, through our study, we confirmed that analyzing learner data in an automatic way is not easy. One way to improve on the study would be to integrate existing dictionaries into the script. As shown in Section 4.2.1, it is the overly bold assumptions of the parser that lead to errors. It is likely that most of these problems can be solved by adding a check against dictionary entries. However, even with this improvement, the problem of undesirable identification that leads to the over-estimation of verbal tense usage remains when a word is accidentally conjugated to an existing verbal form as a result

of learner error. As we can see from the example of the present subjunctive, discussed in Section 4.2.2, when a tense appears when it should not yet have been learned in a textbook at a given level, it is likely to be an inappropriate use. If a semi-automatic approach is considered, manual verification could make the results more reliable when a learner is using a tense that they are not yet expected to know at their current level. The Textbook Verb Profile could serve as a reference resource for estimating the tenses known at a given CEFR level and therefore usable by learners.

We would like to mention some limitations of our study and suggest directions for further research. First, we studied only the tenses found in the verb conjugation table. Thus, the periphrastic near future tense *futur proche* (e.g., *je vais manger* 'I'm going to eat') and the recent past tense *passé récent* (e.g., *je viens de manger* 'I just ate'), both constructed with the verb *venir* 'to come', are counted as two separate verbs in this study – instead of as a compound tense – even though they behave like compound tenses. In addition to the passives and the gerund, which we excluded from the analysis, these automatic identifications and examinations must be addressed.

Second, the sampling was done with the aim of being able to generalize the results, therefore the impact of the learners' native language was not examined. In view of previous studies on the acquisition stages, the consensus is that language acquisition is not affected by the learner's native language. On the other hand, many teachers and researchers are empirically or intuitively convinced that L1 influences L2 acquisition (Izumi et al., 2005; Spada and Lightbown, 2020, p. 119). Therefore, the impact of the language used by the learner (and possibly the language of instruction, although this information is not present in our data) should be taken into account in the analysis.

Third, the present research was conducted from a purely morphological perspective and therefore remains at a one-dimensional level. The English Grammar Profile, for example, provides, in addition to a CEFR level assigned to the items concerned, much more in-depth information such as lexical range. In the future, we would like to also create a Verb Profile for learners, taking into account the lexical and syntactic difficulty of a given verbal tense.

Finally, our attempt to apply automatic analyses to learner data has again highlighted the difficulties of automatically processing data containing errors. However, manual annotation is time consuming and necessarily involves subjective judgments; it seems inevitable to use NLP to treat large amounts of texts in order to produce a Verb Profile for learners, which would give a robust and generalizable perspective based on a large amount of data. Two observations arise from these statements: first, there is a need for a more systematic and in-depth analysis of taggers and parsers in order to tackle the problem of correctly identifying verb tenses in learner language; second, we should seek ways in which to handle learner language in order to make it compatible with our scripts. A potential solution to these problems may lie in the normalization of learner productions, either manually, semi-automatically or automatically.

6 Conclusion

Thanks to our script that automatically identifies verbal tenses we have made it possible to process a large amount of data to establish a Verb Profile of FFL textbooks. It can serve as a resource in a different way from others that already existed, as it is purely data-driven, and thus does not rely on (subjective) human judgments as to which tenses ought to be known at which levels. Another remarkable aspect of this resource is the comprehensive treatment of tenses. Tenses that are not covered in the existing resources, i.e., those that were thought not to need to be taught or were not considered to be used by learners, were also included in the study. This allowed us to verify whether or not these tenses were covered in the textbooks that underpin learners' learning. That said, biases inherent in the data may affect the analyses. Therefore, it should be noted that the quality of our resources depend on the nature of the data used in this study.

Our two methods of analyzing learner productions, one that shows what they wrote and another that shows what they would have wanted/needed to produce, allowed us to describe the state of the use of verbal tenses according to CEFR levels. Furthermore, the comparison of the two annotations revealed that learners, even at advanced levels, had difficulties with the past perfect. We also found a gradation in difficulty, depending on the level, meaning that learners at A1 level had difficulties with the auxiliary verb, while learners at

A2 level had more difficulties with the inflection of the main verb. These results can help teachers focus on areas that need addressing in learners of different levels.

7 Acknowledgement

This study was enabled through a research agreement with France Education International. We would like to express our deepest gratitude to Eva Rolin and Adeline Müller, who willingly accepted to help us during the annotation process. Nami Yamaguchi is supported by the Japan Student Services Organization (JASSO). David Alfter is supported by the Fonds de la Recherche Scientifique de Belgique (F.R.S-FNRS) under grant MIS/PGY F.4518.21.

References

Anne Abeillé, Lionel Clément, and François Toussenel. 2003. Building a treebank for French. In *Treebanks*, pages 165–187. Springer.

Andrea Abel. 2014. A trilingual learner corpus illustrating european reference levels. *RiCOGNIZIONI. Rivista di Lingue e Letterature straniere e Culture moderne*, 1(2):111–126.

Leonel Figueiredo de Alencar. 2017. A computational implementation of periphrastic verb constructions in French. *Alfa: Revista de Lingüística*, 61(2):437–467.

David Alfter. 2021. *Exploring natural language processing for single-word and multi-word lexical complexity from a second language learner perspective*. Ph.D. thesis, University of Gothenburg, Sweden.

David Alfter, Yuri Bizzoni, Anders Agebjörn, Elena Volodina, and Ildikó Pilán. 2016. From distributions to labels: A lexical proficiency analysis using learner corpora. In *Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition*, pages 1–7.

Sueo Asakura. 2002. *Dictionnaire des difficultés grammaticales de la langue française (Shin Furansu bunpō jiten)*. Hakusuisha, Tokyo.

Inge Bartning and Suzanne Schlyter. 2004. Itinéraires acquisitionnels et stades de développement en français L2. *Journal of French language studies*, 14(3):281–299.

Jean-Claude Beacco. 2008. *Niveau A2 pour le français*. Didier.

Jean-Claude Beacco, Béatrice Blin, Emmanuelle Houles, Sylvie Lepage, and Patrick Riba. 2011. *Niveau B1 pour le français*. Didier.

Jean-Claude Beacco, Simon Bouquet, and Rémy Porquier. 2004. *Niveau B2 pour le français: utilisateur-apprenant indépendant: textes et références*. Didier.

Jean-Claude Beacco and Rémy Porquier. 2007. *Niveau A1 pour le français*. Didier.

Vaclav Brezina and Gabriele Pallotti. 2019. Morphological complexity in written L2 texts. *Second language research*, 35(1):99–119.

Bram Bulté and Alex Housen. 2012. Defining and operationalising L2 complexity. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, pages 23–46.

Christian Cherdon. 2005. *Guide de grammaire française*. De Boeck.

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Press Syndicate of the University of Cambridge.

Bastien De Clercq and Alex Housen. 2019. The development of morphological complexity: A cross-linguistic study of L2 French and English. *Second Language Research*, 35(1):71–97.

Robert M DeKeyser. 2005. What makes learning second-language grammar difficult? a review of issues. *Language learning*, 55(S1):1–25.

Heidi C Dulay and Marina K Burt. 1974. Natural sequences in child second language acquisition 1. *Language learning*, 24(1):37–53.

Luise Dürlich and Thomas François. 2018. EFLLex: A graded lexical resource for learners of English as a foreign language. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Núria Gala, Thomas François, Delphine Bernhard, and Cédrick Fairon. 2014. Un modèle pour prédire la complexité lexicale et graduer les mots. In *TALN'2014*, pages 91–102.

Sylviane Granger. 2011. How to use foreign and second language learner corpora. *Research methods in second language acquisition: A practical guide*, pages 5–29.

Maurice Grevisse and André Goosse. 2007. *Le bon usage*. De Boeck, Duculot.

John A Hawkins and Paula Buttery. 2010. Criterial features in learner corpora: Theory and illustrations. *English Profile Journal*, 1.

John A Hawkins and Luna Filipović. 2012. *Criterial features in L2 English: Specifying the reference levels of the Common European Framework*, volume 1. Cambridge University Press.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.

Emi Izumi, Koyotaka Uchimoto, and Hitoshi Isahara. 2005. Investigation into Japanese learners' acquisition order of major grammatical morphemes using error-tagged learner corpus (erā tagu tsuki gakushūsha kōpasu wo mochiita nihonjin eigo gakushūsha no shuyō bunpō keitaiso no shūtoku junjo ni kansuru bunseki). *Journal of Natural Language Processing*, 12(4):211–225.

Julia Krivanek and Detmar Meurers. 2013. Comparing rule-based and data-driven dependency parsing of learner language. *Computational dependency theory*, 258:207.

Donna Lardiere. 2016. Missing the trees for the forest: Morphology in second language acquisition. *Second language*, 15:5–28.

Diane Larsen-Freeman. 2010. Not so fast: A discussion of L2 morpheme processing and acquisition. *Language Learning*, 60(1):221–230.

John Michael Linacre. 1989. *Many-faceted Rasch measurement*. Ph.D. thesis, The University of Chicago.

Ken Machida. 2015. *Compendium de la grammaire française (Furansugo bunpō sō kaisetsu)*. Kenkyusya, Tokyo.

Brian North. 2015. Inventaire linguistique des contenus clés des niveaux du cecrl.

Brian North, Angeles Ortega, and Susan Sheehan. 2010. Eaquals core inventory for general english.

Anne O'Keeffe and Geraldine Mark. 2017. The english grammar profile of learner competence: Methodology and key findings. *International Journal of Corpus Linguistics*, 22(4):457–489.

Gabriele Pallotti. 2015. A simple view of linguistic complexity. *Second Language Research*, 31(1):117–134.

Manfred Pienemann. 1998. *Language processing and second language development*, volume 10. Amsterdam: John Benjamins.

Manfred Pienemann and Gisela Håkansson. 1999. A unified approach toward the development of Swedish as L2: A Processability Account. *Studies in second language acquisition*, 21(3):383–420.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.

Patrick Riba. 2016. *Niveaux C1/C2 pour le français*. Paris, Didier.

Jack C Richards. 1970. A non-contrastive approach to error analysis. In *TESOL Convention*, pages 1–37.

Rachel Rubin. 2021. Assessing the impact of automatic dependency annotation on the measurement of phraseological complexity in L2 Dutch. *International Journal of Learner Corpus Research*, 7(1):131–162.

Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *International Conference on New Methods in Language Processing*.

Helmut Schmid. 2019. Deep learning-based morphological taggers and lemmatizers for annotating historical texts. In *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*, pages 133–137.

Nina Spada and Patsy M Lightbown. 2020. Second language acquisition. In Norbert Schmitt and Michael P.H. Rodgers, editors, *An introduction to applied linguistics*, third edition, chapter 7, pages 172–188. Routledge, London and New York.

Barbora Štindlová, Svatava Škodová, Jirka Hana, and Alexandr Rosen. 2011. CzeSL – an error tagged corpus of Czech as a second language. In *PALC*, pages 13–15.

Milan Straka and Jana Straková. 2017. Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 shared task: Multilingual Parsing from raw text to universal dependencies*, pages 88–99.

Yukio Tono. 2013. *The CEFR-J Handbook : a resource book for using CAN-DO descriptors for English language teaching*. Taishukan Shoten.

Elena Volodina, David Alfter, Therese Lindström Tiedemann, Maisa Susanna Lauriala, and Daniela Helena Piipponen. 2022. Reliability of automatic linguistic annotation: native vs non-native texts. In *Selected papers from the CLARINAnnual Conference 2021*. Linköping University Electronic Press (LiU E-Press).

Kevin Yancey, Alice Pintard, and Thomas Francois. 2021. Investigating readability of french as a foreign language with deep learning and cognitive and pedagogical features. *Lingue e linguaggio*, 20(2):229–258.

A Link between tense/mood and learner proficiency levels according to previous studies

Bold-faced tenses and moods show newly introduced tenses/moods at this level. An asterisk indicates that usage is sporadic at this level according to the original study. Underspecified tenses in the original study such as "future" or "past" are indicated in double quotation marks.

A.1 Summary of moods and tenses by levels according to Bartning and Schlyter (2004)

Stage	Mood and tense
Stage 1	Indicative: **past perfect***
Stage 2	Indicative: past perfect, **imperfect***
Stage 3	Indicative: **future simple***
	Subjunctive*
	"Past"
Stage 4	Indicative: **pluperfect**
	Conditional
	Subjunctive
Stage 5	Indicative: pluperfect, future simple
	Conditional
	Subjunctive
Stage 6	"stabilized inflectional morphology"

A.2 Summary of moods and tenses by levels according to Beacco et al. (2004); Beacco and Porquier (2007); Beacco (2008); Beacco and Riba (2011)

Level	Mood and tense
A1	Indicative: **present**, **imperfect***, **past perfect***
	Infinitive
	Imperative
	Conditional: present
A2	"Main tenses for certain verbs"
	"imperfect"*
	"future"*
B1	Indicative: present, imperfect, past perfect, **pluperfect**, "future"
	Conditional: present
	Subjunctive: present*;
	Imperative: present
	Infinitive: present
	Participle: present*, **past***
B2	Indicative: present, imperfect, past perfect, "future", **future anterior**
	Conditional: present, **past**
	Subjunctive: present, **past**
	Imperative : present, **past**
	Infinitive : present, **past**
	Participle: present, past, **past perfect**

A.3 Summary of moods and tenses by levels according to North (2015)

Level	Mood and tense
A1	Indicative: **present**, **imperfect***, **past perfect***
	Conditional: **present***
	Imperative: **present***
	Infinitive: **present**
A2	Indicative: present, imperfect, past perfect, **simple future**
	Conditional: present*
	Imperative: present
	Subjunctive: **present***
	Infinitive: present
B1	Indicative: imperfect, **pluperfect**
	Conditional: present, **past**
	Imperative: present
	Subjunctive: present
	Infinitive: present, **past**
B2	Indicative: **simple past**, pluperfect, **anterior future**
	Conditional: present, past
	Subjunctive: present, **past (receptive)**
	Infinitive: past
C	Indicative: simple past
	Subjunctive: present, past

B Tenses and their abbreviations

Tense	English name	Abbreviation
Indicatif présent	Indicative present	ind_pres
Indicatif imparfait	Indicative imperfect	ind_imp
Indicatif passé simple	Indicative simple past	ind_ps
Indicatif passé composé	Indicative past perfect	ind_pc
Indicatif plus-que-parfait	Indicative pluperfect	ind_pqp
Indicatif passé antérieur	Indicative anterior past	ind_pa
Indicatif futur simple	Indicative simple future	ind_fs
Indicatif futur antérieur	Indicative anterior future	ind_fa
Conditionnel présent	Conditional present	cnd_pres
Conditionnel passé	Conditional past	cnd_pass
Impératif présent	Imperative present	impe_pres
Impératif passé	Imperative past	impe_pass
Subjonctif présent	Subjunctive present	sbj_pres
Subjonctif passé	Subjunctive past	sbj_pass
Subjonctif imparfait	Subjunctive imperfect	sbj_imp
Subjonctif plus-que-parfait	Subjunctive pluperfect	sbj_pqp
Infinitif présent	Infinitive present	inf_pres
Infinitif passé	Infinitive past	inf_pass
Participe présent	Participle present	part_pres
Participe passé	Participle past	part_pass

Table 7: Tenses and abbreviations

C Textbook and learner corpus annotation results

C.1 Textbook corpus annotation results

	A1		A2		B1		B2		C	
	A1	A1%	A2	A2%	B1	B1%	B2	B2%	C	C%
ind_pres	4501	66.81	5334	49.81	5262	49.63	4725	47.38	4318	48.31
ind_imp	88	1.31	492	4.59	414	3.90	504	5.05	216	2.42
ind_ps	3	0.04	8	0.07	34	0.32	185	1.86	85	0.95
ind_pc	459	6.81	1018	9.51	913	8.61	702	7.04	591	6.61
ind_pqp	1	0.01	29	0.27	74	0.70	48	0.48	24	0.27
ind_pa	1	0.01	0	0	1	0.01	4	0.04	1	0.01
ind_fs	35	0.52	274	2.56	227	2.14	185	1.86	285	3.19
ind_fa	0	0	1	0.01	14	0.13	14	0.14	3	0.03
cnd_pres	51	0.76	128	1.20	180	1.70	203	2.04	170	1.90
cnd_pass	0	0	0	0	27	0.25	23	0.23	16	0.18
imp_pres	233	3.46	476	4.44	201	1.90	121	1.21	290	3.24
imp_pass	0	0	1	0.01	10	0.09	1	0.01	0	0
sbj_pres	3	0.04	34	0.32	139	1.31	135	1.35	82	0.92
sbj_pass	0	0	1	0.01	14	0.13	5	0.05	6	0.07
sbj_imp	0	0	0	0	0	0	0	0	0	0
sbj_pqp	0	0	0	0	0	0	3	0.03	0	0
inf_pres	1361	20.20	2902	27.10	3052	28.78	3087	30.96	2832	31.68
inf_pass	1	0.01	11	0.10	41	0.39	27	0.27	19	0.21
Total	6737		10709		10603		9972		8938	

Table 8: Results of the textbook corpus analysis. Light shaded cells indicate levels at which the tense was used at least once in all of the textbooks of this level. Dark shaded cells indicate levels at which the tense was used at least five times in all of the textbooks of this level.

C.2 Learner corpus annotation results (original)

	A1				A2				B1				B2				C			
	V	V%	C	C%	V	V%	C	C%	V	V%	C	C%	V	V%	C	C%	V	V%	C	C%
ind_pres	86	73.50	25	96.15	105	57.38	30	100	192	49.61	30	100	205	44.28	30	100	164	36.36	29	96.67
ind_imp	0	0	0	0	12	6.56	5	16.67	18	4.65	9	30	27	5.83	10	33.33	24	5.32	10	33.33
ind_ps	0	0	0	0	0	0	0	0	0	0	0	0	1	0.22	1	3.33	0	0	0	0
ind_pc	2	1.71	2	7.69	13	7.10	6	20	42	10.85	13	43.33	40	8.64	16	53.33	68	15.08	17	56.67
ind_pqp	0	0	0	0	0	0	0	0	4	1.03	4	13.33	0	0	0	0	4	0.89	3	10
ind_pa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ind_fs	1	0.85	1	3.85	4	2.19	2	6.67	15	3.88	9	30	18	3.89	9	30	20	4.43	8	26.67
ind_fa	0	0	0	0	0	0	0	0	1	0.26	1	3.33	0	0	0	0	0	0	0	0
cnd_pres	3	2.56	1	3.85	2	1.09	1	3.33	9	2.33	6	20	14	3.02	10	33.33	12	2.66	8	26.67
cnd_pass	0	0	0	0	0	0	0	0	1	0.26	1	3.33	2	0.43	2	6.67	0	0	0	0
imp_pres	0	0	0	0	0	0	0	0	5	1.29	4	13.33	5	1.08	4	13.33	5	1.11	4	13.33
imp_pass	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sbj_pres	3	2.56	2	7.69	0	0	0	0	4	1.03	3	10	8	1.73	6	20	0	0	0	0
sbj_pass	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sbj_imp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sbj_pqp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
inf_pres	22	18.80	9	34.62	47	25.68	24	80	95	24.55	28	93.33	140	30.24	29	96.67	153	33.92	29	96.67
inf_pass	0	0	0	0	0	0	0	0	1	0.26	1	3.33	3	0.65	3	10	1	0.22	1	3.33
Total 1	117				183				387				463				451			
part_pres	0				1				0				0				0			
part_pass	12				10				10				3				1			
∅	76				56				55				39				9			
E	25				17				6				1				0			
NV	8				2				0				0				2			
Other	121	50.84			86	31.97			71	15.50			43	8.50			12	2.59		
Total 2	238				269				458				506				463			

Table 9: Results of the original learner corpus annotation. V: counts for a given tense; V%: percentage of V with regards to all verbs, i.e., Total 1; C: number of learners who produced this tense; C%: percentage of C with regards to all learners (26 for level A1, 30 for the other levels)

C.3 Learner corpus annotation results (corrected)

	A1				A2				B1				B2				C			
	V	V%	C	C%	V	V%	C	C%	V	V%	C	C%	V	V%	C	C%	V	V%	C	C%
ind_pres	109	62.29	25	96.15	124	57.67	29	96.67	195	46.21	29	96.67	199	40.28	30	100	170	36.56	29	96.67
ind_imp	0	0	0	0	11	5.12	5	16.67	19	4.50	9	30	26	5.26	9	30	26	5.59	11	36.67
ind_ps	0	0	0	0	0	0	0	0	6	1.42	5	16.67	1	0.20	1	3.33	4	0.86	3	10
ind_pc	1	0.57	1	3.85	4	1.86	2	6.67	14	3.32	8	26.67	21	4.25	10	33.33	20	4.30	8	26.67
ind_pqp	0	0	0	0	0	0	0	0	1	0.24	1	3.33	0	0	0	0	0	0	0	0
ind_pa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ind_fs	1	0.57	1	3.85	4	1.86	2	6.67	14	3.32	8	26.67	21	4.25	10	33.33	20	4.43	8	26.67
ind_fa	0	0	0	0	0	0	0	0	1	0.24	1	3.33	0	0	0	0	0	0	0	0
cnd_pres	5	2.86	2	7.69	2	0.93	1	3.33	9	2.13	6	20	15	3.04	11	36.67	12	2.58	8	26.67
cnd_pass	0	0	0	0	0	0	0	0	1	0.24	1	3.33	2	0.40	2	6.67	0	0	0	0
imp_pres	0	0	0	0	1	0.47	1	3.33	8	1.90	6	20	8	1.62	7	23.33	6	1.29	5	16.67
imp_pass	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sbj_pres	1	0.57	1	3.85	1	0.47	1	3.33	3	0.71	2	6.67	9	1.82	7	23.33	0	0	0	0
sbj_pass	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sbj_imp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sbj_pqp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
inf_pres	47	26.86	19	73.08	41	19.07	20	66.67	106	25.12	27	90	157	31.78	29	96.67	156	33.55	29	96.67
inf_pass	0	0	0	0	0	0	0	0	1	0.24	1	3.33	3	0.61	3	10	1	0.22	1	3.33
Total	175				215				422				494				465			

Table 10: Results of the corrected learner corpus annotation. V: counts for a given tense; V%: percentage of V with regards to all verbs, i.e., Total 1; C: number of learners who produced this tense; C%: percentage of C with regards to all learners (26 for level A1, 30 for the other levels)