



SMILLE for Portuguese: Annotation and Analysis of Grammatical Structures in a Pedagogical Context

Leonardo Zilio^(✉), Rodrigo Wilkens, and Cédric Fairon

Centre de traitement automatique du langage – CENTAL,
Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium
{leonardo.zilio,rodrigo.wilkens,cedrick.fairon}@uclouvain.be

Abstract. In Second Language Acquisition (SLA), the exposure of learners to authentic material is an important learning step, but the use of raw text may pose problems, because the information that the learner should be focusing on may be overlooked. In this paper, we present SMILLE for Portuguese, a system for detecting pedagogically relevant grammatical structures in raw texts. SMILLE's rules for recognizing grammatical structures were evaluated in random sentences from three different genres, achieving an overall precision of 84%. The automatic recognition of pedagogically relevant grammatical structures can help teachers and course coordinators to better inform the choice of texts to be used in language courses, while also allowing for the analysis of grammar profiles for SLA. As a case study, we used SMILLE to analyze pedagogical material used in a Portuguese as foreign language course and to observe how the predominance of grammatical content in the texts is related to the described grammatical focus of the language levels.

Keywords: Second Language Acquisition · Grammatical structures
Natural Language Processing · Grammatical parsing for Portuguese

1 Introduction

Research on the field of Second Language Acquisition (SLA) has already shown that the mere presentation of input to a language learner is not enough for ensuring that something will be learned [9]. This means that the language learner may process the input for its meaning alone, without noticing its linguistic structures, because there is no salient language information. Input is understood as language data that is potentially processable and made available, by chance or by design, to language learners [15]. On the other hand, the intake is the part of the input which is actually internalized by a learner and that can potentially be stored in the long-term memory [11].

Supported by the Walloon Region (Projects BEWARE 1510637 and 1610378) and Altissia International.

An input in its raw form has lower chances of being converted into intake by the learner, and may thus not bring any new linguistic information. In the early 90's, Schmidt [12] developed the hypothesis that, in order for language learners to convert input into intake, they have to notice the relevant information in the input. Schmidt would later state, in a less controversial way, that “people learn about the things that they attend to and do not learn much about the things they do not attend to” [13]. There is much discussion regarding the assumptions of the Noticing Hypothesis, and it has its contesters (e.g. Truscott [18]). Nevertheless, it seems to be of general agreement that noticing is at least a facilitator of the language learning process, even though there are differences in the way that authors view the process of noticing, either as a purely conscious process or as a possibly unconscious process [5].

In this context, raw input may not present properties for drawing the learner's attention, especially when one deals with authentic texts, which are normally not meant as a language learning object. In an attempt to solve the lack of salience in raw input, Smith and Truscott [16] suggested the use of what they called “input enhancements”, so as to give prominence to the relevant linguistic information. This “focus-on-form strategy” [6] has provided a new way to assist language learners, and some studies have shown that input enhancements represent a positive step in transforming input into intake (e.g. [10, 14]).

In this paper, we present SMILLE (Smart and Immersive Language Learning Environment) for Portuguese¹, a system that can analyze and enhance written texts by employing Natural Language Processing (NLP) techniques for automatically retrieving pedagogically relevant grammatical structures. By highlighting pedagogically interesting structures in texts, SMILLE can be used by language teachers to automatically locate specific grammatical structures in texts and to evaluate if the texts that they are going to use with their learners are adequate to their language level. For instance, some structures introduced in a given moment may not be fully understood by the learners until they reach a more advanced level – this is specially the case for structures that are seen at the end of each level. As such, using a text that exposes the learner to all structures of a level may lead to over-exposure and cause the learner to lose focus. In this context, it is important to analyze to which structures the learner is being exposed during the second language learning process. Besides the more directly pedagogical application of SMILLE (i.e., enhancing grammatical structures), its association with a second language learning curriculum makes of it a useful tool for analyzing the pedagogical material of second language courses.

Since SMILLE prioritizes a pedagogical approach to information extraction, some of the automatically annotated grammatical structures are not directly recognized from part-of-speech-tagger or dependency-parser information (such as hidden and explicit subject and passive voice) and, thus, they are recognized through rules based on the more generic parser information. These rules are not trivial, and, thus, in this paper we want to especially observe the precision of

¹ The system is available for testing at https://cental.uclouvain.be/resources/smilla_smille/smille/.

SMILLE’s rules, because, as stated by Meurers et al. [9], in language learning, precision tends to be more important than recall.

In addition to observing SMILLE’s precision, we also put it to test in analyzing the pedagogical material used in a Portuguese as foreign language course. This analysis was designed to take into account the distribution of grammatical structures in the pedagogical material, in order to observe how the textual content presented to the learners aligns with the grammatical content that is taught in the different levels. In brief, this analysis is a profile of grammatical structures that occur in the texts of the pedagogical material, and this profile will then be contrasted with the structures presented in the course’s handbooks. Our hypothesis is that the grammatical content of the handbook for the basic level will be significantly more prominent (95% confidence) in the basic level texts. Conversely, the grammar of the advanced level will be significantly more prominent in the texts for the advanced level.

This paper is organized as follows: we present information on related work in Sect. 2; next, we describe SMILLE’s annotated structures in Sect. 3; Sect. 4 presents evaluation of SMILLE’s annotation in different corpora; in Sect. 5, we explain and present the results of our experiment with the pedagogical material; and, finally, in Sect. 6 we present our final remarks and future work.

2 Related Work

In this section, we describe applications developed in the context of Second Language Acquisition (SLA) that can retrieve pedagogical information from raw texts and enhance it to learners.

The REAP project [3] is a tutoring system for English that focuses on finding authentic, Web-based texts that are suitable for the user in terms of reading level. The system also highlights words that are supposedly not known by the user. There is also a REAP.PT project [8] that was ported from English to European Portuguese and then further developed to also encompass gamification and user interaction in a 3D environment.

The WERTi system [9] allows for text enhancements of selected linguistic elements of English, Spanish and German. WERTi uses NLP tools combined with rules and regular expressions to retrieve text information for each of seven linguistic subjects: articles, determiners, gerunds, noun countability, phrasal verbs, prepositions, wh-questions.

SmartReader, developed by Azab et al. [1, 2], provides a reading assistant tool that is fully based on the structures annotated by Stanford CoreNLP [7]. Each word in the text can also be clicked on to display semantic, syntactic and other information. It also displays syntactic function of selected words in the given sentence and generates simple questions about named entities, provided the answers are in the near context.

The *FLAIR* system is an online information retrieval system that annotates and re-ranks Web documents based on user-selected grammatical constructions [4]. It can recognize 87 different types of grammatical structures described in the official curriculum of English as foreign language used in German schools.

SMILLE distinguishes itself from these systems in how the enhancements are selected. For instance, in WERTi, only a few grammatical structures are available, while SmartReader and REAP present information more relevant to the meaning of the words and lexical units, and rely exclusively on parsing for retrieving grammatical structures. And, although FLAIR and SMILLE share a bigger scope in terms of detected grammatical information, the focus of FLAIR lies on text retrieval, while SMILLE focuses on the recognition process. In addition, with the exception of REAP, the other systems do not focus on Portuguese.

3 Grammatical Structures in SMILLE

SMILLE was originally developed for English [20,22,23], and then was further extended to Portuguese, both understood as a foreign language for the learner. This extension to Portuguese included a fully new set of rules that were developed based on the grammatical structures that are deemed important in a Portuguese-as-foreign-language course².

SMILLE links the detected information to the guidelines of the Common European Framework of Reference for Languages (CEFR) [19], so that the grammatical enhancements are not limited to isolated linguistic structures, but covers the needs for a given language level and for specific linguistic knowledge that is required from the learner in proficiency tests. By applying rules on top of the parser annotation³, SMILLE also detects grammatical structures that are not directly retrieved from parsing. As such, for instance, teachers can select texts that are interesting according to their learner’s preferences, while keeping an eye on important information in terms of linguistic structures that are relevant for their process of acquiring a second language.

SMILLE uses the PassPort system [21], a dependency parsing system based on Universal Dependency tags and PALAVRAS part-of-speech tags, as basis. Thus, much of the grammatical information that is detected by SMILLE for Portuguese requires only that the underlying parser correctly analyze the word or structure in question. Such is the case, for instance, of some adverbs, adjectives and simple verb tenses. However, some structures require rules on top of part-of-speech and dependency tags for retrieving more complex grammatical constructions, such as compound verb tenses, passive voice and relative clauses. And other structures still, such as comparatives and some adverbial phrases, are retrieved based on specific rules. As such, SMILLE combines the analysis done by the parser with hand-written rules to extract text information that would not be easily identified, and would not be salient, in a raw input. Several rules in SMILLE are the mere association of different part-of-speech tags and dependency tags or attachments (e.g., compound tenses), other rules require much more complex pattern matching, with multiple possibilities, especially in the case of comparatives, which can appear in several different forms.

² Our grammatical structures were based on the course developed by Altissia International (www.altissia.com).

³ SMILLE for Portuguese makes use of the PassPort system [21].

While developing SMILLE, we had to make a decision regarding the granularity of grammatical structures and the escalation of knowledge associated to each language level. In a language course, different grammatical structures can be learned in progressive steps, so, for instance, today a language learner may study the relative pronoun “que” (approx. “that/which”) and later, during another session, it is possible to learn the relative pronoun “quem” (approx. “who”). In authentic texts, the chances are that different pronouns will appear at the same time, interwoven in the text. To address this fine-grained differentiation, SMILLE would have to encompass specific rules for each case, sometimes for each word in a grammatical category. This would require more processing and an increase in the number of rules. So, although SMILLE respects the escalation related to different language levels (e.g., grammatical structures from different levels were separated in specific rules), the progression of content in the same level was overruled and generalized in overarching classes of grammatical structures, such as “relative pronouns”.

SMILLE for Portuguese contains a total of 71 rules for recognizing pedagogically relevant grammatical structures in written texts. These rules encompass both the Brazilian and the European variants and are based on the CEFR levels from A1 to B2, and each rule is linked to a specific level. Here is a list of grammatical structures that SMILLE can detect in Portuguese texts⁴: prepositions, articles, use of pronouns “tu” and “você”, pronouns used as indirect and direct complements, possessive pronouns, demonstrative pronouns, comparatives, adjectives, plural forms, nouns, expression of preferences, imperative, expressions of obligation, various verb tenses (including progressive ones), interrogative sentences, irregular verbs, uses of “ser”, “estar”, “ter” and “haver”, diminutives, direct and indirect complements, superlative, final clauses, relative clauses and pronouns, verbal periphrases, numbers, possessives, indefinite pronouns, use of the pronoun “si”, several types of adverbs and adverbs derived from adjectives, passive voice, hidden and explicit subjects, and use of clitics.

4 Evaluation of Selected Structures

For evaluating the quality of the rules used for recognizing the different, pedagogically relevant grammatical structures that SMILLE can display to the user, and to see which ones can be trusted for further analyses, we used random sentences from three different genres and applied SMILLE’s pipeline of parsing and rules annotation.

First, we selected three different genres: literature⁵, newspaper articles (from the Diário Gaúcho corpus⁶) and subtitles (from the Portuguese corpus of subti-

⁴ We do not present here the 71 rules because many of the grammatical structures are divided along the CEFR levels, presenting some basic content in lower levels and reinforcing them in higher levels, and others are divided in different categories, such as the verb tenses, the comparative forms, the types of adverbs, etc.

⁵ Selected romances from www.dominiopublico.gov.br.

⁶ This corpus was compiled in the scope of the project PorPopular (www.ufrgs.br/textecc/porlexbras/porpopular/index.php).

tles compiled by Tiedemann [17]). We then annotated the corpora with SMILLE and randomly extracted 25 sentences for each structure from each corpus to be manually evaluated, totaling 75 sentences/instances per structure⁷. Finally, a manual evaluation was carried out by one linguist. Here are example sentences for the 20 structures that were manually evaluated in terms of precision (the main words associated with the structure are marked in *italic*):

1. **Adverbs of manner:** A mulher bateu *fortemente* no assaltante.
2. **Adverbs derived from adjectives:** Os papéis desapareceram *rapidamente*.
3. **Interrogative pronouns and adverbs:** *De que* estás a falar?
4. **Relative pronouns:** A árvore *que* está no centro do parque é a mais bonita.
5. **Superlative form of adjectives:** O Benfica é *o mais forte*.
6. **Adjectival comparative forms:** A Paula é *mais velha do que* a Susana.
7. **Extended comparative forms:** Um bebé *dorme mais do que* um adulto.
8. **Compound future tense:** *Vou estudar* muito para o exame.
9. **Forms of expressing obligation:** Elas *têm de* estudar muito
10. **Interrogative clauses in the present tense:** *Onde está* o teu amigo?
11. **Verbal periphrases:** *Começo a traduzir* agora mesmo.
12. **Compound pluperfect tense:** Eu já *tinha jantado* quando tu chegaste.
13. **Present continuous tense:** A Sofia *está trabalhando*.
14. **Reflexive pronouns:** O professor explica-*se* aos seus superiores.
15. **Final clauses:** Vim *para te ver*.
16. **Relative clauses:** O rapaz *com quem te encontraste* é muito giro.
17. **Hidden and explicit subjects:** *Disseram-me* que *eles* iam dormir aqui.
18. **Progressive tenses:** A Rita *esteve lavando* a cara.
19. **Passive voice:** O trabalho *foi terminado* ontem.

Table 1 shows results of the annotation divided by structure and genre. As we can see, most of the structures have high precision, so that the mean precision lies at 84.07% for the evaluated structures, and the median is 88%. The literature genre seems to pose more problems for the annotation, with a mean of 83.58% and a median of 84%; newspaper articles were worse in the mean precision, with 82.32%, but the median was much higher, at 92%; finally, the subtitles had the best mean precision, at 86.32%, and median, at 96%. In terms of individual structures, there were very few for which the genre seems decisive, and most of them had either generally bad performance, like reflexives, or generally good performance, like the progressive tenses. Even so, we see some structures, like the compound future or the hidden or explicit pronominal subjects, that present an imbalance in the precision evaluation towards one genre.

This precision evaluation showed us which structures can be used in further analyzing the pedagogical material in terms of content and organization per level. The material and the analysis are described in the next section.

⁷ Sentences with more than one instance of the selected structure were evaluated only based on the first instance.

Table 1. Precision of automatically annotated grammatical structures

#	Structure	Newspaper	Literature	Subtitles	Total
1	Adverbs of manner	72%	92%	76%	80.0%
2	Adverbs from adjectives	96%	100%	100%	98.7%
3	Interrogative pronouns	92%	80%	96%	89.3%
4	Relative pronouns	80%	80%	52%	70.7%
5	Superlative	92%	100%	100%	97.3%
6	Comparative	100%	100%	100%	100.0%
7	Extended comparative	80%	76%	84%	80.0%
8	Compound future	92%	48%	92%	77.0%
9	Expression of obligation	96%	100%	100%	98.7%
10	Questions	80%	80%	92%	84.0%
11	Verbal periphrases	28%	48%	60%	45.3%
12	Plusperfect tense	100%	100%	100%	100.0%
13	Present continuous	100%	100%	100%	100.0%
14	Reflexive pronouns	44%	60%	28%	44.0%
15	Final clauses	92%	84%	96%	90.7%
16	Relative clauses	76%	68%	68%	70.7%
17	Hidden and explicit subject	56%	96%	100%	84.0%
18	Progressive tenses	100%	96%	100%	98.7%
19	Passive voice	88%	80%	96%	88.0%

5 Pedagogical Material: Analysis of Grammatical Distribution

Having evaluated the structures that SMILLE annotates by means of rules, we turned ourselves to the task of analyzing how courses of Portuguese as foreign language are organized in terms of grammar and how they present this information in the texts that exist in their pedagogical material. As a case study, we selected the material developed at the Universidade Federal de Juiz de Fora (UFJF) and used in its course for Brazilian Portuguese learners⁸.

The corpus is composed of texts from handbooks used for teaching Brazilian Portuguese to foreigners at the UFJF and covers basic and intermediate levels. Since the rules of SMILLE covers those two levels of the CEFR, the non-existence of an advanced level was not a problem for our analysis. We also ignored the levels to which each of the SMILLE's structures are linked in the CEFR and considered only the curriculum of the specific course, as stated in the handbooks. The corpus contains texts used in reading activities, but we excluded texts with gaps used for exercises or texts that explained grammatical content of the language course.

⁸ <https://oportuguesdobrasil.wordpress.com/musicas-apresentadas-na-sala-de-aula/>.

We also excluded lyrics and poetry, since these genres tend to use different punctuation and structure. The texts are skewed to the informative genre, with more emphasis to magazine and history articles, but there are also dialogues, literature, opinion, and general descriptions. The corpus contains 19,741 tokens (8,031 for basic and 11,710 for intermediate) distributed along 957 sentences (421 for basic and 536 for intermediate). It was annotated with all SMILLE structures for Portuguese and the frequencies were then standardized according to the number of sentences per document (in the case of syntactic structures) or tokens per document (for morphological or lexical structures).

For the analysis, we excluded grammatical structures that had less than 80% precision in the evaluation, but we added structures that are directly based on parser information (like simple verb tenses and word classes), since the parser has around 94% of accuracy for part-of-speech tagging and around 85% of accuracy for dependency parsing [21]. We analyzed the corpus in terms of distribution of grammatical structures at each level, looking at the most prominent structures in each of them in terms of significant differences in relative frequency between both levels. We then contrasted the different prominences with the presented grammatical structure of the handbooks to see if they are in consonance in terms of presentation of structures and exposure of learners to the structures.

By running a Mann-Whitney U test, we could see that some structures have a significantly different occurrence ($p < 0.05$) in both corpora. Some grammatical structures, like the present tense overall and, in specific, the present tense of the verb “ter”, and the use of personal pronouns as subject, are significantly more prominent in the sentences of the basic level, as expected according to the pedagogical content of the handbook. Looking at the intermediate level, we have structures like the past future and the past imperfect tense as more prominent. According to the dispositions in the handbooks, both these tenses are taught at the end of the basic level and are reviewed with more emphasis during the intermediate level, so their predominance at the review level is understandable.

Probably due to the size of the corpus, some structures that have a different occurrence in both levels did not achieve a significant level of confidence ($p < 0.05$). So, for instance, the present forms of the verb “ser” and all forms of the present of the conjunctive occur more than double in texts from the basic level, where they are indeed emphasized according to the grammar content of the handbooks. On the other hand, structures associated to questions, like interrogative pronouns and simple questions in the present tense were over twice more frequent per sentence at the intermediate level, even though the formation of questions are emphasized as a topic at the basic level. Most of the observed structures did not show a significant difference and occur in a similar way in both basic and the intermediate levels, and this suggests that most of the texts used in the corpus are not there to emphasize grammatical aspects of the language.

This type of analysis can aid teachers and pedagogical coordinators in the task of preparing a Portuguese as foreign language course in a way that the texts can better reflect and emphasize the grammatical content that is being

taught. In this analysis, our hypothesis that the texts would match their level were confirmed for the structures for which we had significant differences, but there are clues that some of the content from the intermediate level may be, in fact, too basic. A more in detail evaluation of each text would need to be performed to gauge the full extent to which the texts match their level, but this type of analysis is beyond the scope of this paper.

6 Final Remarks

In this paper, we presented SMILLE for Portuguese, a system that can recognize pedagogically relevant grammatical structures in raw texts. SMILLE covers structures from levels A1 to B2 following a CEFR categorization, corresponding to the basic and intermediate levels. It can be used not only to enhance texts that are to be used with language learners, but it can also be applied in the selection of these texts. As such, a teacher would have a help in selecting more appropriate texts based on their grammatical profile and the grammatical structures that need to be emphasized for the learners.

Since many of the recognized structures use rules on top of the parser annotation, we carried out a precision evaluation in three different genres: newspaper articles, literature and subtitles. Most of the structures scored as high as 100% of precision, such as the comparatives and the present continuous tense, but some of them, such as reflexive pronouns and verbal periphrases, scored much lower (respectively, 44% and 45.3%). Overall, the system achieved an average precision of 84% in the evaluated structures. For the structures that presented bad performance, we saw a mix of bad parsing performance and bad rules, so that we will be addressing these issues for the future versions of the system.

We also presented a case study of how SMILLE can be applied to observe the adequacy of texts used in Portuguese as foreign language courses. For that, we analyzed the texts presented in handbooks of the language course held at the Universidade Federal de Juiz de Fora. This analysis was performed to observe if the texts in each of the two available levels (basic and intermediate) actually emphasize the grammatical content that is described in the handbooks. From the structures that had significantly different use in both levels, we could observe that they do follow the description provided in the handbooks for the basic and the intermediate levels. Nevertheless, for some of the structures, there was a large difference (more than double) in terms of average relative frequency in the texts (but with no significant difference), pointing to a possible mismatch with the level's grammatical content and suggesting that further investigation would be needed to evaluate the adequacy of their distribution in the texts.

SMILLE can detect grammatical structures that are relevant for the learning of Portuguese as a foreign language and it can help analyze texts used in language courses, but it could also be applied to analyze a full profile of, for instance, how learners of Portuguese tend to write their texts in terms of grammatical organization. As future work, we are interested in expanding the corpus of handbook texts and include also learners' texts, to be able to compare how both these instances of language learning behave in terms of grammar.

References

1. Azab, M., Salama, A., Oflazer, K., Shima, H., Araki, J., Mitamura, T.: An english reading tool as a NLP showcase. In: The Companion Volume of the Proceedings of IJCNLP 2013: System Demonstrations, pp. 5–8. Asian Federation of Natural Language Processing, Nagoya, Japan, October 2013. <http://www.aclweb.org/anthology/I13-2002>
2. Azab, M., Salama, A., Oflazer, K., Shima, H., Araki, J., Mitamura, T.: An NLP-based reading tool for aiding non-native english readers. *Recent Advances in Natural Language Processing*, p. 41 (2013)
3. Brown, J., Eskenazi, M.: Retrieval of authentic documents for reader-specific lexical practice. In: *InSTIL/ICALL Symposium 2004* (2004)
4. Chinkina, M., Kannan, M., Meurers, D.: Online information retrieval for language learning. In: *ACL 2016*, p. 7 (2016)
5. Cross, J.: Noticing'in sla: Is it a valid concept. *TESL-EJ* **6**(3), 1–9 (2002)
6. Doughty, C.: Second language instruction does make a difference. *Stud. Second Lang. Acquisition* **13**(04), 431–469 (1991)
7. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D.: The Stanford CoreNLP natural language processing toolkit. In: Association for Computational Linguistics (ACL) System Demonstrations, pp. 55–60 (2014). <http://www.aclweb.org/anthology/P/P14/P14-5010>
8. Marujo, L., et al.: Porting reap to european portuguese. In: *SLaTE*, pp. 69–72 (2009)
9. Meurers, D., et al.: Enhancing authentic web pages for language learners. In: Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications, pp. 10–18. Association for Computational Linguistics (2010)
10. Plonsky, L., Ziegler, N.: The CALL-SLA interface: Insights from a second-order synthesis (2016)
11. Reinders, H.: Towards a definition of intake in second language acquisition (2012)
12. Schmidt, R.: The role of consciousness in second language learning1. *Appl. Linguistics* **11**(2), 129–158 (1990)
13. Schmidt, R.: Attention, awareness, and individual differences in language learning. *Perspect. Individ. Characteristics Foreign Lang. Educ.* **6**, 27 (2012)
14. Simard, D.: Differential effects of textual enhancement formats on intake. *System* **37**(1), 124–135 (2009)
15. Smith, M.S.: Input enhancement in instructed sla. *Stud. Second Lang. Acquisition* **15**(02), 165–179 (1993)
16. Smith, M.S., Truscott, J.: Explaining input enhancement: a mogul perspective. *Int. Rev. Appl. Linguistics Lang. Teach.* **52**(3), 253–281 (2014)
17. Tiedemann, J.: Finding alternative translations in a large corpus of movie subtitle. In: *International Conference on Language Resources and Evaluation* (2016)
18. Truscott, J.: Noticing in second language acquisition: a critical review. *Second Lang. Res.* **14**(2), 103–135 (1998)
19. Verhelst, N., Van Averaet, P., Takala, S., Figueras, N., North, B.: *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge University Press, Cambridge (2009)
20. Zilio, L., Fairon, C.: Adaptive system for language learning. In: *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, pp. 47–49. IEEE (2017)

21. Zilio, L., Wilkens, R., Fairon, C.: Passport: a dependency parsing model for portuguese
22. Zilio, L., Wilkens, R., Fairon, C.: Enhancing grammatical structures in web-based texts. In: Proceedings of the 25th EUROCALL, pp. 839–846, Accepted, 2017
23. Zilio, L., Wilkens, R., Fairon, C.: Using NLP for enhancing second language acquisition. In: Proceedings of Recent Advances in Natural Language Processing, pp. 839–846 (2017)