Using a bilingual concordancer to promote metalinguistic reflection in the learning of an additional language: The case of B1 learners of Catalan

Johannes Graën (Universität Zürich, Switzerland)
Carme Bach (Universitat Pompeu Fabra, Serra Húnter Fellow, Spain)
Daniel Cassany (Universitat Pompeu Fabra, Spain)

I. Introduction

In this work we describe the experimental use of a bilingual concordancer in a learning class of Catalan as an additional language. The concordancer is fed with translated sentences for several language pairs from a freely available parallel corpus. Users of the application can use simple search terms or regular expressions on source and target language in order to filter the available corpus examples.

Information technology is omnipresent in modern Western lives. It comes as no surprise that information and communication technology (ICT) also plays an important role in today's language learning settings, including learning a language as an additional language, as is the case of Catalan, a minority language next to Spanish or English. Proficiency in technology use is a prerequisite for the successful employment of ICT in settings and learner autonomy.

The objective of corpus exploration exercises on parallel texts is to encourage learners to develop hypotheses that serve to explain the examples they observe and evaluate them with further corpus queries. That process is designed to promote metalinguistic reflection and foster learner autonomy in line with the data-driven learning paradigm.

The bilingual concordancer PaCLE (Parallel Corpora for Learning Exercises) is part of a framework designed to let learners and teachers generate language learning exercises from parallel corpus material (Graën, submitted). We evaluate its use qualitatively by analyzing screen recordings of a class of B1 learners of Catalan following the methodology established by Vázquez-Calvo (2016) for describing technology use in language learning.

II. Theoretical framework

Information technology can also be used to foster metalinguistic reflection of learners of additional languages at basic levels, such as the didactic experience we show in this chapter.

Since learners began to play an active role in learning in general a few years ago, leaving the teacher's role as a merely complementary one, technologies have also become essential for language learning and for promoting learner autonomy.

> Technology opens up several potentials for language learning: access to native speakers and language peers around the world, easy 24/7 access to a wide range of authentic and learning-supportive language teaching and learning materials, building and exposure to engaging learning experiences and environments, and facilitating the construction of positive learner identities. (Lai and Gu, 2011, p. 317)

In this context, Data-Driven Learning was born as an interesting and effective method for language learning and teaching. According to Vyatkina and Boulton (2017), the use of corpora for language learning brings a significant gain in learning at the same time that its use motivates students to learn a new language.

Specifically, for teaching Spanish as a foreign language Buyse (2011) states that the use of corpora in the ELE classroom helps to clearly improve the quality of students' written expression. For Cobb and Boulton (2015), working with corpora is a "generally constructivist and inductive approach to language learning".

> The basic idea is that massive but controlled exposure to authentic input is of major importance, as learners gradually respond to and reproduce the underlying lexical, grammatical, pragmatic, and other patterns implicit in the languages they encounter. (Cobb and Boulton, 2015, p. 481)

First of all, it is necessary to talk about metalinguistic awareness, known as the process of introspective or explicit thinking through which learners are able to reflect on the use of some linguistic structures in their own language or in other languages they are going to use. As noted in Joan-Casademont et al. (in press), the degree to which learners of an additional language develop metalinguistic awareness depends, to some extent, on the tools and environment used for their learning (Festman, 2021; Witney and Dewaele, 2018). Clearly, the process of enduring language learning is enhanced if metalinguistic reflection takes into account the learners' prior linguistic knowledge as learners of other languages, their own or additional languages.

III. Methodology

We use a bilingual concordancer for our experiment. The software allows users to perform flexible searches on sentence pairs extracted from parallel corpora. The exercises that students had to do as part of their classroom activities mainly focus on the comparison of Catalan with Spanish (Castilian) and French in one exercise. However, students were encouraged to explore correspondences for any other language pair involving Catalan and also to use other language assistance resources to better understand the use of some grammatical aspects in Catalan and the similarities between Catalan and other dominant languages such as Spanish or French.

3.1 The bilingual concordancer

The PaCLE tool[1] (Graën, submitted) is designed as a tool to turn parallel corpus data into language learning exercises. Target user groups are language teachers and autonomous learners. The latter group can use the tool to automatically generate exercises or simply use the built-in concordancer to perform corpus searches. The early prototype used by our learners for this work comprises the parallel concordancer with sentence pairs from the OpenSubtitles corpus (see the following section) for any pair of the languages Catalan, English, French, German, Italian, Spanish and Swedish.

Corpus searches are performed on the whole sentences in one or both of the respective languages. We can specify one word, a fixed sequence of words or a pattern described by regular expressions. We expect that only a few (if any) of the envisaged users will be acquainted with regular expressions. To unlock the great potential that comes with such a versatile instrument, we would need to instruct users on how to use it, which is out of the scope of a regular language learning class. Searching the corpus with regular corpus query tools is

[1] PaCLE stands for "Parallel Corpora for Language Exercises"

equally challenging for language learners without basic education in linguistics. We cannot expect them to be acquainted with concepts such as tokens, lemmas or morphological features and their values, although they will probably be familiar with their manifestations in the target language.

Figure 1

Excerpt of the PaCLE search results for sentence pairs Spanish-Catalan where the sequence “un paseo” can be found in Spanish and “un passeig” in Catalan translation.

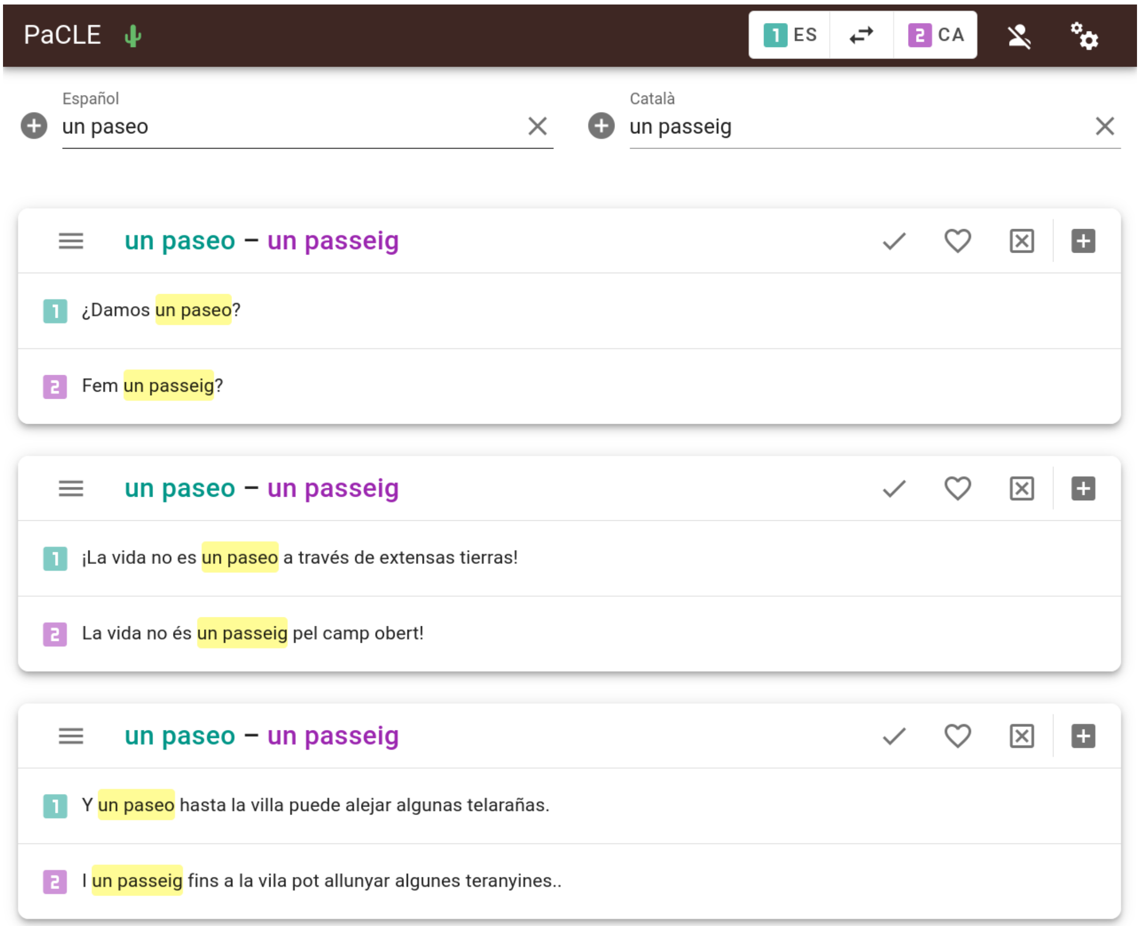
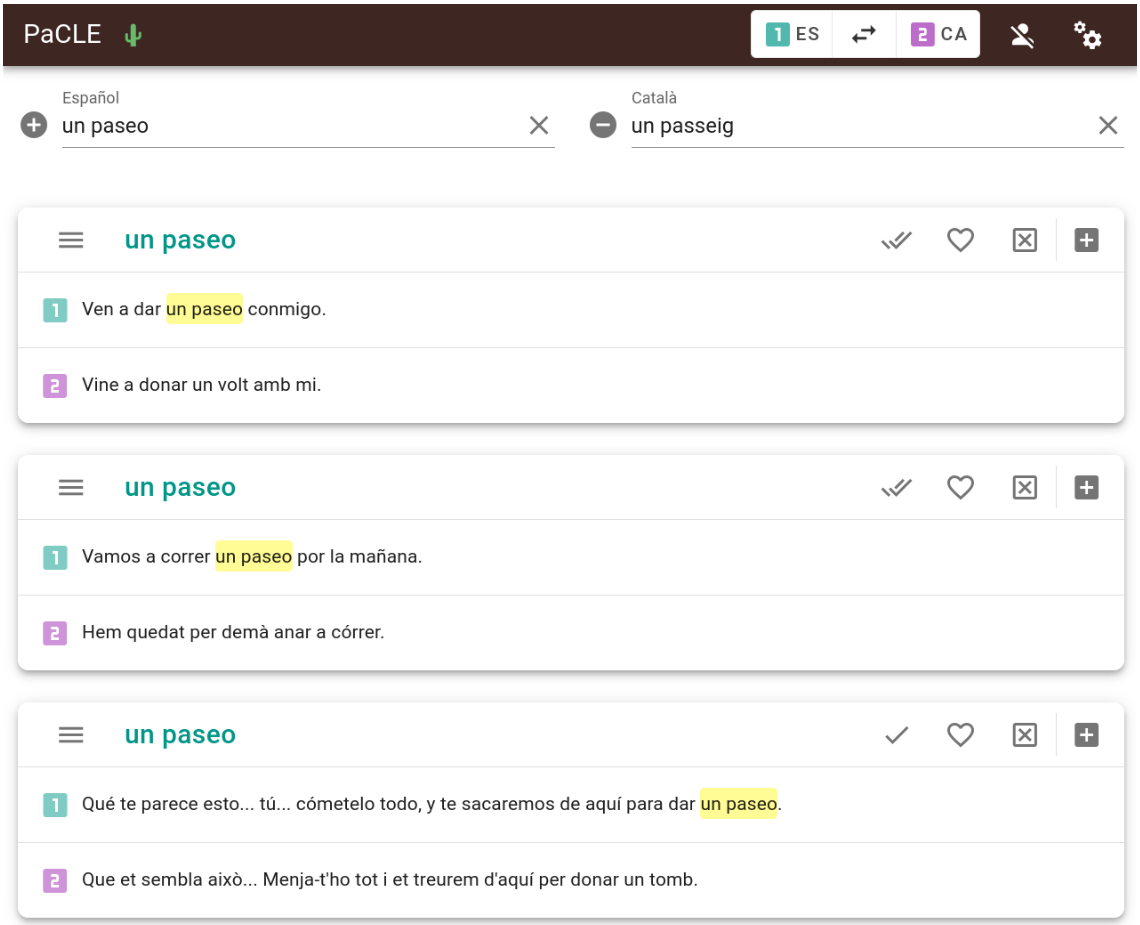


Figure 2

Excerpt of the PaCLE search results for sentence pairs Spanish-Catalan where the sequence "un paseo" can be found in Spanish, but "un passeig" not in the Catalan translation. Therefore, no text passage is marked on the Spanish translations.



A simple query that users can perform without any prerequisites is the search for a singular word or a continuous sequence of words in one or both languages. Figure 1 shows a query with some search results where the parallel occurrence of "un paseo" in Spanish and "un passeig" in Catalan is required. The search can be negated for each language, which means that the tool only shows sentence pairs where the negated search term or expression does not occur (as depicted in Figure 2). This function is particularly useful to search for example sentences with other translation variants than the default one.

3.2 The OpenSubtitles corpus

More and more parallel corpora have been made openly available in the past decades. The OPUS collection[2] (Tiedemann 2009, 2012) comprises numerous parallel corpora from a variety of sources with different text types and domains. For our experiments, we use the OpenSubtitles corpus (Lison and Tiedemann 2016) which comprises movie subtitles that have been, for the most part, translated to numerous other languages by volunteers or, alternatively, extracted from videos by means of OCR (Lison and Tiedemann 2016).

Subtitles are limited in space and time; they need to be shown for a sufficiently long time for the reader to be able to process them. For the same reason, they cannot be overly long. Longer sentences that need to be displayed in chunks are often marked with ellipses and can thus be reconstructed. While shorter sentences are, in general, easier to comprehend than longer ones, the necessity of shortness can bring about phenomena that run counter to comprehension such as the omission of parts that an experienced reader can reconstruct. We describe other potential sources of parallel corpora and their utility for language learning applications in (Graën, submitted, sec. 2.1 and 7.1).

3.3 Classroom experiment

We used the PaCLE application in a course entitled "Comparative use of Catalan and Spanish" (Ús comparat català-castellà),[3] which consisted of 20 learners of Catalan as an additional language (L2) with the CEFR target level B1. The course took place during the academic year 2020–2021 and was conducted exclusively online via the Zoom video conferencing software due to sanitary measures taken against the spreading of COVID-19. Participants were either L1 speakers of Spanish or had a proven competence of at least C1 level in Spanish. As language students, they also had other additional languages. All of them knew English, 13 were studying French and five were studying German. Three participants had Italian as their L1. For the participants, learning takes place in an environment of immersion as Catalan is the main language used at their place of study.

[2] https://opus.nlpl.eu/

[3] The course forms part of the degree in Translation and Interpreting and Applied Languages of Pompeu Fabra University.

One of the methodologies used in this course – as foreshadowed by its title – is the comparison of Catalan and Spanish, first and foremost with regard to syntactic aspects. That is why we considered the course setting ideal for the experimental use of PaCLE. The acquisition of one language by way of comparison with another language (be it L1 or a strong L2) and metalinguistic reflection about similarities and differences of the languages in question are central course objectives.

The PaCLE application was presented to the students in week five (out of ten), after other helpful ICT resources had been introduced. The application was showcased by means of guided exercises via screen sharing by the lecturer to demonstrate its functionality and point out its utility. Afterwards, the URL of the application was shared via the learning management system used in this course.

In week nine, the students were given an auto-reflexive exercise about three aspects of Catalan syntax, which had been part of the teaching and learning content of the course. As an incentive, those participants who did the exercise would earn extra credits for the final mark. The exercise consists of the following subtasks:

1. Explain the difference between the use of relative adjectives *cuyo*, *cuya*, *cuyos*, *cuyas* and the Catalan corresponding expressions *el qual*, *la qual*, *els quals*, *les quals*. Provide four useful examples obtained from searches in the PaCLE application, and explain the differences observed. Help yourself to any material you consider necessary.

2. Which are the syntactic functions provided by the weak pronoun *hi*? Remember to take into account both possible forms of the pronoun, *hi* and -*hi*. What happens in the Spanish texts?

   What if we make the comparison with French? What differences do you observe compared to Spanish? Provide four useful examples obtained from searches in the PaCLE application, and explain the differences observed between Catalan and Spanish, and Catalan and French. Help yourself to any material you consider necessary.

[3.] What type of conjunction is *perquè*? Does it only have one function? What are the differences with Spanish? Did you encounter any use of *perquè* in which it was not used as a conjunction? If so, what is it grammatically in that case? Provide four useful examples obtained from searches in the

PaCLE application, and explain the differences observed. Help yourself to any material you consider necessary.

As can be seen in the exercise description, it was suggested from the teacher's side to resort to a number of resources, beyond performing searches in PaCLE, in order to explain the observed differences. Students were given complete freedom to use any of the resources they consider helpful to answer the exercise questions:

1. Dictionaries
    a. **DIEC2**[4], dictionary of the Catalan language (diccionari de la llengua catalana), web portal for this normative monolingual dictionary of the Catalan language by the Institute for Catalan Studies (Institut d'Estudis Catalans).
    b. **DLE**[5], dictionary of the Spanish language (diccionario de la lengua española), web portal for this normative monolingual dictionary of the Spanish language by the Royal Spanish Academy (Real Academia Española).
2. Grammars
    a. **GEIEC**[6], Essential grammar of the Catalan language (Gramàtica essencial de la llengua catalana), scaled-down version adapted for online queries by the Institute for Catalan Studies (Institut d'Estudis Catalans)
3. Tools providing linguistic information
    a. **Optimot**[7], a service for linguistic queries provided by the Directorate General for Language Policy (Direcció General de Política Lingüística) in collaboration with the Institute for Catalan Studies (Institut d'Estudis Catalans) and the Centre for Terminology TERMCAT (Centre de Terminologia TERMCAT).

---

[4] Institut d'Estudis Catalans. Diccionari de la llengua catalana. 2a edició. [DIEC2]. http://dlc.iec.cat/

[5] Real Academia Española. Diccionario de la lengua española, 23.ª ed., [versión 23.5 en línea]. https://dle.rae.es

[6] Institut d'Estudis Catalans (2018). Gramàtica essencial de la llengua catalana. [GEIEC]. Barcelona: Institut d'Estudis Catalans. https://geiec.iec.cat/

[7] Generalitat de Catalunya. Optimot. Servei de consultes lingüístiques. https://aplicacions.llengua.gencat.cat/llc/AppJava/index.html

b. **esadir.cat**[8], language portal of the Catalan Media Corporation (Corporació Catalana de Mitjans Audiovisuals (CCMA)), which is a serviceable tool for anyone working in the field of communication and teaching, students, and, in general, everyone interested in the use of the Catalan language.

c. **Google Translate**[9], machine translation service offered by Google, often used as a complementary tool for language learning.

[d.] **Diccionario Panhispánico de dudas**[10], a website provided by the Royal Spanish Academy (Real Academia Española) addressing frequently asked questions concerning the use of the Spanish language.

3.4 Instruments

For evaluating the use of PaCLE in the course on "Comparative use of Catalan vs. Spanish", we considered two means:

1. The result of the auto-reflexive exercise that students were requested to do subsequent to performing searches in PaCLE in order to gain extra credits. We analyzed the degree of success students showed in response to the three exercise questions, the productive use of PaCLE, the capacity of metalinguistic reflection and the way other tools were employed to answer the questions, be it the aforementioned additional resources or other tools known to the students, independent of whether they are used as complementary or exclusive means.

2. The recordings of the students' desktops demonstrating how they used the respective tools, but also comprising their communication with the teacher in case of questions or doubts.[11]

For the analysis of the screen recordings, we used the schema proposed by Vázquez-Calvo (2016), which has been developed specifically for the description of technology use in language learning. This schema helps us study

[8] Corporació Catalana de Mitjans Audiovisuals. esadir. https://esadir.cat/

[9] Google Translate. https://translate.google.com/

[10] Real Academia Española. 2005. Diccionario panhispánico de dudas. Madrid: Santillana. https://www.rae.es/dpd/

[11] All screen recordings happened with the consent of the respective student encompassing their use for our research, always following the criteria of the Institutional Commission for Ethical Review of Research Project (Comissió Institucional de Revisió Ètica de Projectes (CIREP)) of the Pompeu Fabra University.

how digital resources are used by our language learners, in particular how each of the tasks has been solved, which techniques they carried out in order to solve the tasks, and how adequate the solutions they obtained from the respective resources have been.

We refrained from collecting feedback by means of a questionnaire as the number of participants was fairly low and we believed that a qualitative analysis of the students' activities would be considerably more adjuvant.

Figure 3

Schema proposed by Vázquez-Calvo (2016, p. 239) for analyzing how language learners make use of technology.

<table>
<tr><th colspan="5">Secuencia 5</th></tr>
<tr><th colspan="5">Análisis</th></tr>
<tr><th>Min.</th><th colspan="2">Datos</th><th>Descripción</th><th>Interpretación</th></tr>
<tr><td rowspan="5">Inicio: 08:00— Fin: 08:10</td><td>Tarea:</td><td>Composición escrita en EN. Resumen de un vídeo y continuación de la historia</td><td rowspan="5">El alumno:
1) Sin dilación, al abrir GT, el alumno comienza a usarlo.
2) La interfaz de GT está configurada de manera que la combinación lingüística es ES-CA.
3) El texto que inserta es «La hsitoria*».
4) Cambia de lengua de salida a EN.
5) A continuación acude al texto y comienza la escritura en EN con «The history».</td><td rowspan="5">El alumno:
— Usa el RLL que tiene más accesible para solucionar la duda que tiene en mente. Prima la inmediatez por el resultado. (1)
— No configura el RLL para su uso específico ante la duda que tiene en mente: configuración de idioma, dirección de lenguas, etc., en este caso realiza esta configuración después de introducir la duda y no antes (2, 4).
— Usa GT como un diccionario bilingüe unívoco, lo que le induce a calcos léxicos. (3)
— Usa GT como un diccionario autocorrectivo, y descuida el texto de entrada a nivel ortográfico. (5)</td></tr>
<tr><td>Duda lingüística:</td><td>Léxico</td></tr>
<tr><td>Finalidad:</td><td>Producción en EN</td></tr>
<tr><td>Recurso:</td><td>GT</td></tr>
<tr><td>Solución:</td><td>No adecuada</td></tr>
</table>

Figure 3 shows an example from Vázquez-Calvo (2016), where for a particular activity of a given task metadata has been annotated. Besides the timespan from the respective screen recording (on the left) and a running number assigned to each sequence as technical data (on top of the box), general information regarding the task is recorded in the first column, a description of which actions can be seen in the video in the second, and the interpretation of those actions in the third one.

As for general information, the schema stipulates five fields: 1) the task that the student is carrying out, which is typically a predefined one, 2) the type of linguistic doubt that the student tries to resolve, 3) the intended purpose of the task, 4) the resource or resources used, and 5) whether the selected solution is judged adequate, partially adequate or inadequate. The description lists sequentially all interactions that the student performs. The interpretation part refers to those actions while commenting on the presumable reasons why they led to an adequate, partially adequate or inadequate solution of the task set.

IV. Analysis

In order to carry out the analysis of the data that will allow us to evaluate the usefulness of the bilingual concordancer, we have triangulated the data from the recording of the students' desks, on the use of the tools proposed to promote learning based on the proposed metalinguistic reflection with the final task they have written, in which they explain the functioning of linguistic features of Catalan in comparison to similar features of other languages, in particular Spanish.

In the following, we provide details on some of the individual activities performed by the students to solve the tasks explained in the previous section. Figures 4 to 7 show different levels of adequacy as assessed by us.

Table 1

Example of the analysis schema showing a partially adequate solution for the task in question.

<table>
<tr><td colspan="5">Student 1 – Sequence 7</td></tr>
<tr><td colspan="5">Analysis</td></tr>
<tr><td>Min.</td><td colspan="2">Properties</td><td>Description</td><td>Interpretation</td></tr>
<tr><td rowspan="5">67 –– 76</td><td>Task</td><td>Writing down the solution found</td><td rowspan="5">After copying examples of haver-hi and -hi as adverbials of place, the student proceeds to answer the question.

The student does not investigate the full list of examples provided by PaCLE but stick to the first two pages (each comprising 10 examples) and does not realize that hi can be used regardless of the verb haver-hi in proclitic position, as an adverbial pronoun as "L'Anna estudia a la seva habitació? Sí hi estudia" or as a verb phrase complement, as in the example below "Ja penses en</td><td rowspan="5">This student is ignorant of how to use regular expressions and does not try to find hi independent of haver-hi in a proclitic position.

The comparison with French could have led to more fruitful results as a proclitic position il y a is not lexicalized as are hay and hi ha.

Since he/she does not know how to talk about lexicalization or grammaticalization, he/she uses the word partícula (particle) to refer to the lexicalized hi in hi ha / il y a and -hi. He/she could have referred to it as weak</td></tr>
<tr><td>Linguistic doubt</td><td>Differences in the way hi/-hi works in Catalan compared to Spanish and French</td></tr>
<tr><td>Objective</td><td>Solve the second task</td></tr>
<tr><td>Resource</td><td>PaCLE, MS Word</td></tr>
<tr><td>Solution</td><td>Partially adequate (very little)</td></tr>
</table>

			la teva mare? Sí *hi* penso". Using a different regular expression in order to exclude cases of *haver-hi* could have been worth trying out, but he/she apparently does not know how to do that and does not ask for support, which he/she would have received if such a question had been raised.	pronoun (*pronom feble*) which was the term used in the task description, but he/she did not use that term. He/she talks about interpretation, unable to syntactically analyze the pronoun. The student does not find cases of verb phrase complements, only one of an adverbial of place, in enclitic position. He/She does comprehend the typographic markers of metalinguistic uses of the words, which the student finds entertaining to format in italics.

Table 2

Example of the analysis schema showing an adequate solution for the task in question.

Student 14 – Sequence 4				
Analysis				
Min.	Properties		Description	Interpretation
7 –– 15	Task	Consult PaCLE	The student copies the search expressions provided in the task description to PaCLE, skims the results and quickly starts copying them to Word. Interestingly, he/she chooses examples that do not coincide in gender and number. Very important: he/she picks examples that are not complete, which we would have removed from the corpus examples in the first place, if we had seen them, but the student does not assess them, as he/she has already selected sufficient examples.	Very good selection of examples that do not coincide in gender and number between Catalan and Spanish. Pragmatic when choosing examples; he/she only focuses on differences between Catalan and Spanish, namely gender and number and the concordance between the possessor and its possessive in each language. The examples that we would consider deficient are not necessarily bad, all depends on the intention behind the corpus search.
	Linguistic doubt	Differences in the use of possessive relative pronouns in Catalan and Spanish		
	Objective	Solve the first task		
	Resource	PaCLE, MS Word		
	Solution	Adequate		

			The student walks through examples until reaching the ninth page (each comprising 10 examples). He/she does not copy example pairs as other students do, but copies them one by one.	

Table 3

Example of the analysis schema showing a perfectly adequate solution for the task in question.

<table>
<tr><td colspan="5">Student 11 – Sequence 13</td></tr>
<tr><td colspan="5">Analysis</td></tr>
<tr><td>Min.</td><td colspan="2">Properties</td><td>Description</td><td>Interpretation</td></tr>
<tr><td rowspan="5">76</td><td>Task</td><td>Consult PaCLE</td><td rowspan="5">The student opens PaCLE and knows how to search by modifying the search expression successfully.

He/she might have previous knowledge of regular expressions.

Already on the first page (each comprising 10 examples), the student finds all required examples, and even more.

He/she copies one example that serves him/her well and explains it in Word. Then, the student copies and pastes the example from PaCLE.

The student does not provide four different examples as there are actually no four distinguishable cases to explain.</td><td rowspan="5">The student is probably familiar with regular expressions; as a matter of fact, all of the students should have come in contact with them as part of the lecture “ICT resources” (recursos TIC).

This student is even aware that there are fewer than four different functions of perquè and therefore does not provide four different examples, but only three very appropriate ones.</td></tr>
<tr><td>Linguistic doubt</td><td>Functions of perquè</td></tr>
<tr><td>Objective</td><td>Solve the third task</td></tr>
<tr><td>Resource</td><td>PaCLE, MS Word</td></tr>
<tr><td>Solution</td><td>Very adequate</td></tr>
</table>

Table 4

Example of the analysis schema showing an inadequate solution for the task in question.

<table>
<tr><td colspan="5">Student 15 – Sequence 12</td></tr>
<tr><td colspan="5">Analysis</td></tr>
<tr><td>Min.</td><td colspan="2">Properties</td><td>Description</td><td>Interpretation</td></tr>
<tr><td rowspan="5">56</td><td>Task</td><td>Consult Optimot</td><td rowspan="5">The student opens the Optimot application to find out how to translate the Spanish word nombre to Catalan.

He/she searches for número instead and since the same word exists in Catalan, the student does not corroborate or look up any further information but accepts the result as it is.

What is more, this student chooses Catalan-Spanish in the application and the correct result is shown on the first page, but he/she ignores it or does not pay attention to it.</td><td rowspan="5">The student looks up número. Even if that word has a different meaning in Catalan, he/she takes it for granted. He/she uses Optimot only for corroborating what he/she thinks should be the translation, without actually revising the results obtained, even though he/she performs the right search for Catalan-Spanish translation and the result comprises the information he/she was looking for.</td></tr>
<tr><td>Linguistic doubt</td><td>Differences in the use of possessive relative pronouns in Catalan and Spanish</td></tr>
<tr><td>Objective</td><td>Find out how to say nombre in Catalan</td></tr>
<tr><td>Resource</td><td>Optimot</td></tr>
<tr><td>Solution</td><td>Inadequate</td></tr>
</table>

The respective analyses can be aggregated in two dimensions. First, we look at the individual student's activity records, which allows us to draw conclusions about their respective capabilities concerning metalinguistic reflection as well as technical proficiency, and second, we aggregate analyses over all participants regarding individual task in order to assess the technical resources used, in particular our bilingual concordancer.

The triangulation of the data, both from the recordings of the students' use of the tool and from the results observed in the final work they handed in, allowed us to establish three student profiles, according to the degree of use of the bilingual concordancer provided and of the other resources consulted:

A. The student who does not know how to take advantage of PaCLE: they look at the results, but do not know how to interpret them, nor do the additional linguistic resources provided seem to be of any help to them.

B. The student who is able to partially interpret the data without the need to consult any other type of linguistic resource, given that their knowledge of Spanish and their reasoning ability are good and allow them to comprehend how the grammatical element in question works in Catalan.

C. The student who not only draws good conclusions from the use of PaCLE but also attempts to seek further explanations in complementary linguistic tools provided and in other ones.

We classified 26,7% of the participants as type A, 33,3% as type B and 40% as type C students. No generalizations can be made from this observation as many factors potentially play a role with regard to a student's ability to draw conclusions from language samples obtained (such as linguistic background, previous experience with using such resources, and the assistance provided by the respective teacher). Nonetheless, we are convinced of the potential of parallel corpora for comparative language learning classes. Overall, a majority of the students who participated in our experiment were able to solve the tasks autonomously with as little preparation as one introductory session.

As for the concordancer PaCLE, we observed several issues on more than one occasion. Most notably are difficulties in understanding regular expressions, which is why we provided search expressions for each task (which the students could manipulate but did not need to in order to solve the tasks). In theory, the participants should know the concept from a previous lecture (see Figure 6, Interpretation), but expecting them to write regular expressions from scratch seemed to us very challenging when conceiving the exercises.

Since only a small subset of the features provided by regular expressions is needed for corpus searches in general (namely alternatives as in "cuyo|cuya|cuyos|cuyas", word boundaries as in "\yhi\y" or variable suffixes as in "*cuy\w+",* which we did avoid by explicitly listing the alternatives instead), one potential improvement could be the use of a query language with a reduced feature set. The downside of this, however, would be that more technically experienced users would not be able to make use of the full potential of regular expressions and the features of that new query language would still need to be explained to users.

A potential improvement of the PaCLE application that we derived from the analysis of the screen recordings concerns the way in which results are presented. Currently, we show all matching sentence pairs in an aleatory order and only highlight the matching parts in both languages. The idea behind not ordering the results is that users would go through the list of results and select those that they consider useful. In some cases, we observed exactly this behavior; in other cases, students were browsing through a large number of examples but were apparently unable to judge them.

Two improvements to the interface that we had sketched out but have not implemented to date would potentially improve use experience. First, we designed the visual presentation of the individual examples to allow for interactive presentation of linguistic features such as the part of speech, syntactic relations and morphological features. Enhanced with this feature, we expect users to spend more time on exploring the respective examples and facilitate the selection of good examples. Second, we planned to provide options for ordering results, e.g. by conflating similar sentence pairs to one that is most representative, deriving readability and proficiency level measures on sentence pairs, and list corresponding translation equivalents and their frequencies.

V. Final reflections

We have described how we employed a bilingual concordancer in a language learning class to foster learner autonomy and promote metalinguistic reflection. Students were given three exercises about Catalan syntax and we provided them with links to the concordancer PaCLE and several other language resources on the internet. The concordancer had already been introduced in a previous class.

For evaluating the use of the concordancer and other technical resources, we analyzed screen recordings provided by the participants using an already established schema and also consulted the written answers to the exercise questions. We classified the students into three groups by their ability to interpret the results obtained from the concordancer and draw the right conclusions with regard to the properties of Catalan syntax.

We could see that the majority of the students were able to use the tool in a beneficial way, the capability of metalinguistic reflection is key for their success. Our analysis revealed several minor problems concerning the use of our tool, some of which were not unique to an individual student. Those are of the utmost interest for us, as one of our goals is to address them in future versions of the concordancer and thus improve its usability.

We would have liked to provide feedback to the participants who provided their screen recordings to us. Unfortunately, the shift from a physical to an online format required additional time that we were lacking in the end.

VI. References

Buyse, K. (2011). ¿Qué corpus en línea utilizar y para qué fines en la clase de ELE? Del texto a la lengua. *La aplicación de los textos a la enseñanza-aprendizaje del español L2-LE,* 37, 277–288.

Cobb, T. and Boulton, A. (2015). *Classroom applications of corpus analysis. The Cambridge Handbook of English Corpus Linguistics.* Biber, D. and Reppen, R. (Eds.). Cambridge University Press, 478–497.

Festman, J. (2021). Learning and processing multiple languages: The more the easier? Language Learning. 71, 121–162.

Graën, J. (Submitted). *Learning Languages from Parallel Corpora: An Application for Turning Corpus Examples into Language Learning Exercises.* Slovenščina 2.0.

Joan-Casademont, A, Bach, C. and Viladrich, È. (In press). Analysis of training needs in coherence and cohesion based on corpora: some online didactic proposals for Catalan as an additional language (AL). Tyne, H. & Spina, S. (Eds.) *Applying Corpora in Teaching and Learning Romance Languages*. Amsterdam: John Benjamins.

Lai, C. & Gu, M. (2011). Self-regulated out-of-class language learning with technology. *Computer Assisted Language Learning, 24*(4), 317–335. https://doi.org/10.1080/09588221.2011.568417

Lison, P. & Tiedemann, J. (2016). OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*.

Tiedemann, J. (2009). News from OPUS – a Collection of Multilingual Parallel Corpora with Tools and Interfaces. *Proceedings of Recent Advances in Natural Language Processing (RANLP), 5*, 237–248.

Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC).*

Vyatkina, N. and Boulton, A. (2017). Corpora in Language learning and teaching. *Language Learning & Technology, 21*(3), 1–8.

Vázquez-Calvo, B. (2016). *Digital Language Learning from a Multilingual Perspective: The Use of Online Language Resources in the One-to-One Classroom*. PhD thesis, Universitat Pompeu Fabra.

Witney, J. and Dewaele, J. M. (2018). Learning two or more languages. In A. Burns & J. C. Richards (Eds.), *The Cambridge guide to learning English as a second language*, 43–52. Cambridge University Press.