

Design, implementation and evaluation of a data-driven learning didactic unit based on an online series corpus*

Diseño, implementación y evaluación de una unidad didáctica de data-driven learning basada en un corpus de series online

Néstor Singer**, José Luis Poblete*** y Carlos Velozo****

ABSTRACT

This paper presents the implementation of a didactic unit based on Singer's proposal for data-driven learning (DDL) within a task-based (TB) framework in the context of Chilean translation education. The didactic intervention has four stages: 1) data collection from students' preferred multimodal consumption in order to generate a written corpus based on the transcription of popular online TV series; 2) the design of a didactic unit that graduates concordance software use and student autonomy by means of linguistic tasks; 3) the application of the approach in an English Language 7 course unit with 24 participants at the Translation Studies undergraduate program at Universidad de Santiago de Chile, and 4) the evaluation of the intervention by means of an online questionnaire. Students' and lecturers' perceptions suggest that a DDL-TB approach is suitable for language teaching in translator education. Further recommendations to improve future DDL-TB projects are also provided.

Palabras clave:
data-driven learning, language teaching, multimodal consumption, corpus, translator education.

* Publicación adscrita al Proyecto de Innovación Docente (PID) N°033-2018 de la Vicerrectoría Académica de la Universidad de Santiago de Chile, Usach. Los autores quieren agradecer el apoyo del Instituto Profesional EATRI y del Departamento de Lingüística y Literatura de la Usach.

** Chileno. MA TESOL por The University of Manchester. Profesor asociado en Universidad de Santiago de Chile, Santiago, Chile. nestor.singer@usach.cl

*** Chileno. MA in English Linguistics and Literature por KU Leuven. Profesor adjunto en Universidad de Santiago, Santiago, Chile. jose.poblete.b@usach.cl

**** Chileno. Magíster en Docencia para la Educación Superior por la Universidad Andrés Bello. Profesor jornada completa en Instituto Profesional EATRI, Santiago, Chile. carlos.velozo@eatri.cl

RESUMEN

Este artículo presenta la implementación de una unidad didáctica basada en la propuesta de Singer para *data-driven learning* (DDL) en un enfoque por tareas (TB) en el contexto de la formación de traductores. La intervención didáctica consta de cuatro fases: 1) recolección de datos en torno al consumo multimodal preferente de los alumnos para generar un corpus escrito basado en las transcripciones de series de televisión online populares; 2) el diseño de una unidad didáctica que gradúa el uso del software de *concordancer* y la autonomía del alumno mediante tareas lingüísticas; 3) la aplicación del enfoque en una asignatura de Lengua Inglesa VII con 24 participantes en el programa de Lingüística Aplicada a la Traducción de la Universidad de Santiago de Chile; y 4) la evaluación de la intervención mediante un cuestionario online. Las percepciones de alumnos y profesores sugieren que un enfoque DDL-TB es adecuado para la enseñanza de idiomas en la formación de traductores. También se entregan recomendaciones adicionales para mejorar futuros proyectos basados en DDL-TB.

Keywords:
consumo multimodal, corpus, data-driven learning, enseñanza de lenguas, formación de traductores.

Introduction

Language teaching in translator education is still a relatively under-researched topic in the literature (Singer, 2016). A possible explanation is that guidelines for translator education usually come from contexts in which potential translation students already display proficiency in the languages with which they intend to work. For instance, in the European context, translation is taught at master's level and acceptance to these programs is conditioned by the applicants' language proficiency in at least two languages (European Master's in Translation, 2017). This allows translation MA programs to mostly focus on the development of students' *translator competence* (TC), understood as the set of skills and knowledge required to translate (Hurtado Albir, 2011).

Translator education in Latin-American contexts differs significantly. Most countries share the same linguistic background: Spanish is the first language, which means that opportunities to learn a second language by travelling to neighboring countries are limited. Second languages, particularly English, are taught in primary and secondary education. The latest EF English Proficiency Index (Education First, 2020) affirms that 13 out of 19 Latin-American countries display a low level of English language proficiency. In Chile, the Agencia de Calidad de la Educación (2018), which advises the Chilean Ministry of Education, reports that in 2017 68% of students nationwide conclude their secondary studies with an A2 level of the Common European Framework of Reference for languages (CEFR) (Council of Europe, 2020) or lower.

This low English language proficiency demands that translation students develop their second language proficiency concurrently with their TC during their undergraduate studies. This issue poses challenges in terms of translation educational curricula and the implementation of language course units that effectively promote language proficiency in a relatively short span of time. This raises questions as to how these language course units for translators could be taught and, perhaps more critically, what rationale should underlie teachers' planning and decision-making in this particular context.

In Chile, most English language teachers working in translation undergraduate programs at tertiary level have been trained in lan-

guage pedagogy but not in translation. This makes it difficult for them to grasp the ultimate purpose of their course unit (Singer & Basaure, 2018), namely the development of linguistic proficiency *and* TC-related skills. This apparent inability to articulate their course units with TC seems to be further worsened by a possible disconnection between language teaching and translator teaching areas in the programs. To address these issues, Singer (2016) proposes the adoption of Johns' (1991, 2000) data-driven learning (DDL) in translator language teaching by means of a task-based (TB) approach (Bygate, 2015; Ellis, 2003; Long, 2015). His approach consists of using corpora to allow student-driven discovery of language patterns while explicitly developing other TC-related skills.

This paper describes the application of a didactic unit using Singer's (2016) DDL-TB approach. Its main purpose is to systematize the implementation of this approach and assess its feasibility and appropriateness for the translator education context. In order to do this, a group of 24 undergraduate students enrolled in an English Language course unit participate in a DDL-TB didactic unit. After the intervention, the participants are enquired about their perceptions of the approach in five dimensions: 1) overall impressions, 2) corpus suitability, 3) unit curricular design, 4) learner autonomy and 5) transferable skills. To make the corpus more meaningful to students, this study uses their multimodal consumption in English language, i.e. online series (Singer et al., 2018), to generate a corpus of nearly 6,000,000 words to be used and explored in the classroom.

The conceptual underpinning of this study is presented in Section 2. Section 3 details the stages involved in the process of designing the corpus and materials for the didactic unit: contextualization, corpus and material design and application. Section 4 problematizes students' perceptions about their experiences in a DDL classroom. Section 5 discusses the authors' reflections about the implementation of the approach. Finally, Section 6 concludes this paper with a Summary of the findings and recommendations for future possible research.

Data-driven learning: affordances and challenges

The idea of using a corpus, i.e. "an electronically stored, searchable collection of texts" (Jones & Waller, 2015, p. 5), for language teaching was

proposed almost three decades ago. Back then, Johns (1991) proposed the notion of *data-driven learning* (DDL): a student-driven search of language items and patterns in corpus from which they could hypothesize about real language use. This conceptualization outlines two critical underlying principles in DDL, namely learner autonomy and exposure to 'real' language.

In terms of learner autonomy, DDL offers a potential means whereby students can take control of their own learning, while teachers become facilitators and co-researchers (Tribble, 2012). This repositioning of learners as the ones discovering the language is captured in Johns's (2000, p. 108) famous words, 'every learner is a Sherlock Holmes' in DDL. The searches in the corpus are carried out by means of a piece of software known as *concordancer*, which filters language items according to key words. The concordancer displays the key word in context (KWIC) at the center of the screen and highlights the relevant collocations matching the search criteria to the left or right of the word.

According to Flowerdew (2015), DDL autonomy is underpinned by Krashen's Noticing Hypothesis, Constructivism, and Vygotsky's Socio-cultural Theory: as learners explore the corpus, they notice particularities in language patterns and then work collectively to rationalize the linguistic phenomenon. By working in groups, they maximize opportunities to share their points of view and contribute to negotiation of meaning with other participants. Thus, learners are required to leave their passive role and actively engage with their peers in search of collective meaning construction surrounding the L2 linguistic system and culture.

DDL also allows students to be exposed to real and rich language (Braun, 2005). Boulton and Cobb (2017) add that DDL exemplar-based learning replaces the simplification of language that more traditional approaches tend to use to explain language patterns and uses. In other words, the second DDL key asset lies in input *assembly*, i.e. students attempting to make sense of how real and authentic language works. According to Robinson (1991), it is this language awareness that would allow learners to have a *feel* for the language they are learning.

In spite of these affordances, DDL presents some challenges that may hinder the implementation of this approach in language course

units. First, students' protagonist role in their language learning process might conflict with students' beliefs about language teaching and learning, as pointed out by Boulton (2009). Bernardini (2001, p. 23) adds that requesting this level of learner autonomy could in some cases imply that students are 'asked to abandon deeply rooted norms of classroom behavior.'

The second issue is related to the genuineness and authenticity of the texts in the corpus. Widdowson (1978) suggests that although the texts that constitute corpus are genuine, i.e. originally produced in the L2, this does not guarantee that the learners will recognize them as such and identify themselves as the original target audience for which the texts were originally produced would. Chen and Flowerdew (2018) point out that one of learners' main concerns using DDL is related to the reliability of these corpora, as these could contain language produced by both native and non-native speakers and could influence their hypotheses and results emerging from the data.

Third, the pedagogical implications of the approach itself could be potentially problematic. Boulton and Cobb (2017) state that most corpora are composed of authentic native language well beyond the comfort level of many learners. In addition, the resulting hints may be confusing and overwhelming due to the quantity of information acquired (Meunier, 2002), which can lead learners to flawed conclusions regarding language use. Similarly, Whistle (1999) emphasizes that DDL-based activities cannot exceed a maximum of 30 minutes, and must have additional types of tasks and dynamics so that learners do not find the approach repetitive, mechanic and hard. If that were the case, this could result in learners feeling frustrated and abandoning the search and tasks altogether (Kennedy & Miceli, 2001). Other factors to ponder when planning a DDL-based lesson involve the control transfer (from teachers to learners), the construction of syllabuses and class sequences. All these considerations presuppose a substantial amount of student training and careful preparation and planning on part of teachers.

To address these issues and facilitate the adoption of this approach in Chilean translation programs, lesson planning and corpus selection should consider current linguistic trends so that learners can explore and, at the same time, be motivated enough to overcome the aforementioned challenges. This is where a corpus based on the

learners' multimodal consumption (Singer et al., 2018) can contribute to DDL didactics.

Multimodal consumption as a corpus resource

Chen and Flowerdew (2018) state that students need to be helped to achieve a level of autonomy that will allow them to use corpora on their own. Literature points to some suggestions for the construction of the corpus so that it is approachable from learners' perspective. Some authors recommend orienting the corpus towards discourse- or genre-based approaches to give coherence to the corpus through a determined common theme (Braun, 2005; Sinclair, 2003). Others, such as Boulton (2011), propose designing corpora that engages with students' interests to address the issues of authentication and motivation.

This study draws on these considerations to create a corpus based on the learners' multimodal consumption. According to Singer et al. (2018), *multimodal consumption* refers to learners' appropriation of a collection of L2 multimodal texts as a means of satisfying their own personal belonging and identity interests. Kress (2010) regards these texts as the combined result of many different semiotic modes that circulate in a specific culture and that, as a result, encapsulate implicit and explicit meanings about subjects, objects and the relationships between them. It is hypothesized that when learners engage with these texts, they reconfigure their own understanding of the world. A prime example of multimodal consumption would be online television series on platforms, such as Netflix, YouTube or Amazon.

A corpus containing this kind of material would easily allow learners to authenticate themselves with it. Learners could then experience the same subjective identification process which they normally perform on a daily basis when consuming these types of texts. This would promote the generation of common coherent axes to articulate the construction of the corpus around series, types of narratives or musical genres for their linguistic analysis. Lastly, a corpus of this nature would foster learner motivation throughout time, allowing prolonged search tasks to be realized.

It is important to point out that, even though this kind of corpus would solve the main DDL-related issues, effective task articulation and graduation of control transfer from teachers to learners is necessary to guarantee a successful implementation.

DDL-task sequencing and graduation

Cresswell (2007) establishes that there are two possible educational approaches within DDL: deductive and inductive. While the deductive stance has to do with teachers' selection of the corpus and graduation of the tasks, using either printed or digital worksheets, the inductive perspective proposes that learners create their own questions to guide their study of the corpus in the search for answers. According to Singer (2016), a deductive stance should be considered as a starting point from which a more inductive approach could gradually be implemented. Based on this critical consideration, he proposes a series of gradual stages in which learners can develop the skills and abilities to successfully conduct a search and obtain the confidence to carry out independent corpus research. Figure 1 illustrates these four stages and their descriptions.

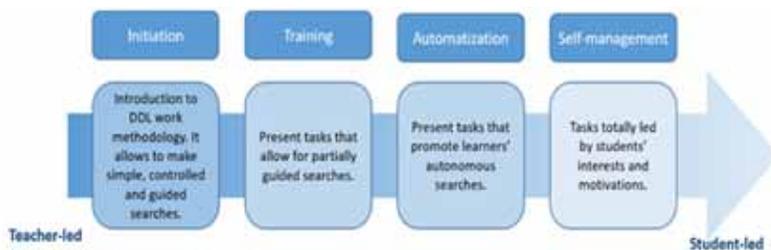


Figure 1. Singer's (2016) adaptation of the DDL task graduation

Source: Self-made.

The initiation and training phases also involve familiarizing learners with the tools available in the concordancer software. This is done by gradually presenting students with search-refining strategies and data visualization tools, such as data plotting. By the time the training stage finishes, learners should be able to autonomously conduct searches on their own.

Additionally, DDL tasks can be combined with other linguistic tasks to promote the development of other production and comprehension skills. Such sequence can be done through a task-based approach (TB) (Bygate, 2015; Ellis, 2003; Long, 2015). This approach consists of a preparation pre-task, a main task where the searches and analysis are conducted, and a post-task for presenting results, reviewing and reflecting on the process. However, it is possible to apply these stages in the creation of a general framework for DDL-TB lesson sequencing, as proposed in Figure 2.

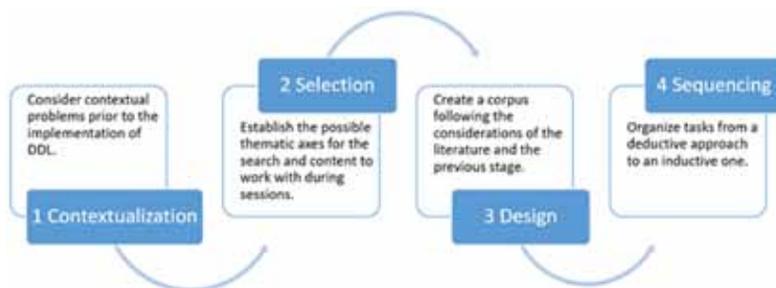


Figure 2. Adaptation of Singer's (2016) framework for lesson planning using a DDL-TB approach

Source: Self-made.

First, a context and needs analysis is conducted to determine the feasibility of applying a DDL approach in the educational setting. Second, based on the course unit outline, grammar points or lexis are selected, which are used to construct the learning outcomes of the course unit. Both contents and learning outcomes outline the type of corpus that would be required to be incorporated into the classroom. Third, surveys or interviews are applied to define the type of multimodal consumption that satisfies the needs mentioned in the previous stages. Once the texts that could be used to generate the corpus are determined, data is sourced, compiled and stored to be used with a concordancer software. Fourth, learners can be given different tasks, firstly from a deductive stance to an inductive one in a gradual transition. These four stages are used in the design of a didactic unit that is applied in a Chilean translation program.

Designing a DDL-TB didactic unit

This exploratory-descriptive study systematizes the design of a DDL-TB unit in an English language course unit of a Chilean translation undergraduate program. This section discusses the three consecutive stages involved in creating and implementing the unit planner and related materials: contextualization, corpus and unit design, and didactic intervention.

Contextualization

The objective of this stage is to discover the participants' interests and multimodal consumption in order to guide and inform the corpus crea-

tion and, later, the materials that are used in the study. A Google Forms survey is carried out to collect such data. The instrument has three sections: 1) personal background, 2) perceptions of the teaching-learning process, and 3) learners' multimodal consumption and interests.

The first section delves into their experiences previous to learning English as a foreign language during their education and the ways in which this had taken place. It aims to identify possible correlations between the years of study and the participants' perceptions regarding their own foreign language learning process.

In the second section, participant's language teaching and learning beliefs are explored. Concretely, based on Singer et al.'s (2018) study, the questions elicit the participants' perceptions regarding a good or bad English class, teacher and learner. In addition, questions involving what elements should be incorporated in or removed from an English class to make it more significant or relevant are included. The data collected from these items provide vital input to critically assess the feasibility of the proposal within a selected context. Thus, tasks within the proposal are adjusted to meet students' perceptions of a good lesson to avoid tensions between the learners' educational culture-related beliefs and foster the adoption of the DDL proposal on part of the learners.

Finally, the third section provides significant information for corpus and material elaboration. The questions in this section of the survey gather sources of multimodal consumption, particularly television series on different platforms, such as Netflix, Amazon, YouTube, etc. Additionally, the modality of such consumption, e.g. subtitled, dubbed in Spanish, etc, is also surveyed. The section concludes by investigating the participants' perceptions of the usefulness of the television series in their own English learning process. The survey can be found in Appendix 1.

Corpus and unit design

The corpus for this study is compiled from online TV series that are part of the participants' regular multimodal consumption. Students' responses to the survey detailed in Section 3.1 produce a list of 54 series. Series that have been broadcast from 2013 are selected to build the corpus, resulting in a total of 37 series. Transcripts of all seasons and episodes are collected from the Springfield website (Springfield, 2020)

and saved as plain text (.txt). These are later loaded to the *AntConc* software (Anthony, 2019), which is a freeware concordancer that allows users to create and analyse corpora. This results in a total of 1,253 files, which generates a corpus of 5,592,571 words. The list of the 37 series can be found in Appendix 2.

This corpus is analyzed to study advanced features of reported speech, which is chosen as the most suitable language phenomenon to examine using a DDL-TB approach. Additional vocabulary, idioms and expressions are selected from the coursebooks suggested in the outline, namely *Ready for Advanced 3rd Edition* (Norris & French, 2016), *Upstream Advanced* (Evans et al., 2014) and *Upstream Proficiency* (Evans & Dooley, 2015). After the selection of the components, eight onsite sessions are designed according to Singer's (2016) guidelines, as shown in Table 1.

Table 1

DDL-TB didactic unit planner for the English Language 7 course unit

Session	Software management	Design stage	Linguistic skills
1	Introduction to DDL principles; Developing search questions	Initiation	Reading comprehension
2	Installation of the software with the files; general searches	Initiation	Reading comprehension Oral and written production
3	General searches; finding suitable examples	Formation	Reading comprehension Written production
4	Flexible search tools; Frequency plotting	Formation	Reading comprehension Written production
5	Developing questions; Searches with own ideas	Formation/ Automatization	Oral production Reading comprehension
6	Searches with own ideas; reporting findings	Formation/ Automatization	Oral production Reading comprehension
7	Collaborative assignment final project	Automatization	Oral production
8	Oral presentations and feedback of the final project	Automatization	Oral production

Source: Self-made.

The planner considers the transition from a teacher-centered to a learner-centered stance which, according to Boulton (2009), could foster the effectiveness of the approach in learners. In addition, a gradual introduction to the concordancing software across language-related tasks to avoid frustration due to inability to use the concordancer appropriately (Kennedy & Miceli, 2001). Linguistic skills are also evenly distributed throughout the unit to maximize language skill development during the DDL-TB didactic intervention.

The first two sessions contain several teacher-guided activities with the support of basic search tools. This is due to the fact that initial survey revealed that students regard teachers as the leading figure in the classroom whose responsibility consisted in generating learning opportunities. To avoid immediately challenging these conceptualizations of the language classroom, these sessions rely heavily on teacher guidance and support. Figure 1 corresponds to an extract from the second session.

USING THE CONCORDANCER:

In **Session 1** we learned how to install the concordancer software following these simple steps:

- Open Google.
- Search **AntConc**.
- The first option is <http://www.laurenceanthony.net/software.html>
- Once you are in there, click on Windows [3.5.3]
- After that, click on Ejecutar. You have installed the software successfully.
- Pin the program to the taskbar for further use.

After that, we learned how to load the corpus onto the software:

- On the upper left corner, click on **File**.
- Click on **Open Dir...**
- Search for the folder **Corpus** in the flash drive provided by your teacher.
- Open the folder and click on each of the folders indicated by your teacher.
- Click on **Acceptar**. You have loaded the corpus correctly.
- Compare your list with that of your classmates.
- **Remember:** The concordancer **AntConc** works with **.txt** documents.

We also used the software to find some words and experimented with some of its functions:

1. In **Search Term** type **who**. How many hits did you get? _____
2. Now click on **Words** to select or de-select the option. How many hits did you get? _____
3. Now click on **Case** to select or de-select the option. How many hits did you get? _____
4. Now, in **Search Term** type **whom**. How many hits did you get? _____

Figure 1. Sample of an activity taken from the materials created at initiation level

Source: Self-made.

As sessions progress in their complexity, tasks allow students to familiarize themselves with the concordancer software tools and prompt them to find meaningful examples as those shown in Figures 2 and 3.

Here students engage in more dynamic tasks as they explore the corpus for answers that are reasonably predictable for the teachers.

1. What do you think these two pictures mean?



Figure 1 *Figure 2*

2. To find the answer, we need to explore the corpus to find the most frequent word and the 200th most frequent one.

To do so, follow the following steps:

1. Click on the 'Wordlist' tab on the upper right side of the screen.
2. Click on the Start button. **DO NOT TYPE ANYTHING IN.**
3. Look at the results and answer the questions below?

Which is the most frequent word in the entire corpus? _____

Which is the 200th most frequent word in the corpus? _____

Figure 2. Task involving concordancer plotting
Source: Self-made.

2. REFUSE and DENY

a) Compare the verbs *refuse* and *deny*. How many hits did you get for each? _____

Which of them is the most frequent? In which series? Look at the tab **Concordance Plot** to find that.

b) Inflect the words and add more context as done in the previous exercises. How many hits did you get?
Add more combinations in the boxes. Try with capital and lowercase letter. Remember to click on Words or Case to select or de-select the option.

c) What is the most recurrent pattern? Provide examples in context.

Figure 3. Example collection and analysis task
Source: Self-made.

The graduation of tasks from session four onwards offers the participants the possibility of exploiting the corpus in a much more autonomous way to deepen their knowledge about verb patterns. In addition, they are asked to compare the vocabulary examples provided in the course books with the corpus results. Grammar practice is also fostered throughout the sessions, which included instances of spoken free practice. This involves finding relevant examples, reporting them using suitable structures and then sharing them in groups to foster peer feedback. Figure 3 shows a grammar practice task in Session 5.

SOME FREE PRACTICE

After completing the previous activities, you will have the chance to focus on your favorite series from the available corpus. You will perform the following tasks.

1. Choose a series you like.
2. Find 5 examples of **Direct Speech**.
3. How would you **report** them? Use different **reporting verbs and patterns**.

For example:

Game of Thrones, season 1, episode 4

DS - Lord Baelish: "Distrusting me was the wisest thing you've done since you've climbed off your horse".

RS Alt 1: Lord Baelish warned Ned Stark not to trust him.

RS Alt 2: Lord Baelish warned Ned Stark against trusting him.

4. Find 5 examples of **Reported Speech**
5. Give two alternatives of **Direct Speech**.

Game of Thrones, season 1, episode 7

RS - Lord Baelish: I did warn you not to trust me.

DS Alt 1: If I were you, I would not trust me.

DS Alt 2: I don't think you should trust me.

Figure 4. Student-driven grammar practice

Source: Self-made.

The unit concludes with a collaborative, integrative corpus exploration assignment that aims to integrate DDL-related skills and the development of English language proficiency. The participants have to generate a question, explore the corpus, document their findings and present them to their class. The final oral presentation fosters the participants' engagement and furthers the collective discussion of the findings. Figure 4 shows the instructions for the final task of the didactic unit.

PRESENTATION OF RESULTS AND CONCLUSIONS

1. You will present your analysis and conclusions to the class with the help of a PowerPoint presentation (or similar).
2. Your presentation must include:
 - An introduction in which you present motivations for having selected those particular language items, your research questions, and your hypothesis.
 - A clear definition of the items selected, citing sources correctly.
 - An analysis of the chosen items, with examples taken from the corpus.
 - Graphs and figures that summarize findings.
 - A conclusion, in which a summary of your results is presented, as well as an answer to the research question, confirming or rejecting your hypothesis.
3. Your presentation will be evaluated with a rubric especially designed for this activity, which will be sent to you in the near future.
4. Your presentation should last 12 to 15 minutes.

Figure 5. DDL-TB unit final task
Source: Self-made.

All the materials generated for the intervention are calibrated and tested by the authors and two other English language professionals. Length of tasks have been adjusted and some exercises graduated in sessions four and five to make the transition to student-led searches smoother. After these final adjustments of the materials, the implementation of the didactic unit is carried out. The materials used in the proposal are freely available to the community and can be provided upon request.

Didactic intervention

The DDL-TB didactic unit is implemented the Translation Studies undergraduate program at Universidad de Santiago de Chile, specifically in an English Language 7 course unit, which is set at C1 level of the CEFR. The participants of this study correspond to an intentional sample of 24 fourth-year learners enrolled in the course unit. They are all between the ages of 20 and 25 years old. At the end of the course, the participants complete an online questionnaire about their perceptions of the DDL didactic unit, which consisted of 10 closed and 3 open-ended questions. Unlike one-to-one interviews or focus groups in which the lecturers' role as an interviewer could influence the participants'

responses to the questions, the questionnaire has been chosen because it allows students to comment on their experiences anonymously and voluntarily.

The sessions of the didactic unit are conducted systematically over the course of four weeks. The total duration is 16 hours: 12 on-site hours and 4 at-home autonomous work hours to prepare the final task of the project. The sessions take place in a computer laboratory with Internet access and at the same time frame during the whole intervention process. There are no external interferences during the implementation of the didactic unit.

The application of this unit is approved by the Director of the Translation Studies program and the Coordinator of the English language course units of the program. All procedures are carried out according to the regulations of the Ethics Committee of Universidad de Santiago de Chile.

Students' perceptions about the DDL unit

The questionnaire to gather the participants' perceptions explores five dimensions: 1) overall impressions, 2) corpus suitability, 3) unit curricular design, 4) learner autonomy, and 5) transferable skills. The final questions of this survey can be found in Appendix 3.

On the whole, 11 out of the 24 participants, i.e. 45%, voluntarily completed the final questionnaire that focused on five aspects aforementioned. Quotations from students' answers that encapsulate positive or negative perceptions are presented to support the interpretation of the analysis. These have been translated from Spanish by the authors. The analysis of perceptions is conducted using Grounded Theory (Thornberg & Charmaz, 2014) on a question-by-question basis, coding the most relevant aspects and categorizing students' experience into positive or negative. In cases of ambivalence, answerers are divided into the positive and negative aspects of the response. Codes are later organized according to their frequency to determine the most significant components of their perceptions.

First, in relation to the participants' overall impressions, nine participants (81%) highlight the innovation that the proposal presented compared to traditional strategies normally used in class. Further-

more, they value the possibility to study how language is used in different contexts within a framework of tasks that make the learning process more dynamic.

I think the more opportunities to learn by using different tools, the [concordancer] software in this case, the more enjoyable and interesting the process becomes, and the knowledge acquired is more meaningful. (S6)

I think it was interesting and different from regular classes because it helps us get closer to English used in daily life. (S4)

These findings are consistent with reports from the literature (see Leńko-Szymańska & Boulton, 2015). It appears that DDL adds to the repertoire of classroom tasks creating new opportunities for learning which students are able to identify.

Regarding corpus suitability, ten participants express that the corpus is appropriate for learning language uses, as it shows current language in use, unlike traditional textbooks commonly used in class. In addition, it presents a variety of Englishes, topics and contexts, which provide a functional vision of the language. More critically, the participants comment very positively on the use of online series to constitute a corpus of study:

[Working with the software] was interesting because it was a new modality and using series that most of us know or have watched made it even more interesting. (S11)

Some participants, however, warn that the corpus is more oriented towards American-based series. This presents challenges to establish comparisons between different accents, word choices or expressions.

The corpus was unbalanced. It had very few British series and studying language differences was difficult. (S4)

It appears that the disparity in the amount of British series conflicts with the English variety students have studied for most part of their program. Thus, when performing searches, the examples could not match their expectations due to the particularities and differences of each type.

Similarly, the participants are ambivalent towards the teachers' role during the implementation of the unit. While some regard that

lecturers' guidance is sufficient and useful, others claim that, in spite of their support, the software is not user-friendly, which makes its use particularly challenging. Similarly, when asked about comparing the software to more conventional classes, the participants are divided into two distinct positions. On the one hand, six participants state that this approach offers affordances, such as peer feedback or self-driven discovery, which allows them to go beyond the course books. On the other hand, the other five participants comment on the tediousness of the process, particularly finding examples:

I think the only use of the tool was to verify what we had learned in class and then searching and finding a couple of examples became very tedious. (S7)

These experiences could be related to the fact that the tasks are sequenced in a TB fashion which is normally used in the setting, which means that some students might not be able to differentiate the limits of DDL from their regular learning experiences. Thus, they cannot decide if it has been more effective for their learning process.

In spite of these issues, 81% agree that the final task has been particularly useful to foster their understanding of language in real contexts.

This experience made me reflect on the resources I should check when I study. Sticking to the traditional is not always the most appropriate. (S1)

...because we use series it is easier to see how certain words are used in different contexts and different cultures. (S6)

Those who disagree recommend not to introduce these sorts of tools at this level of their training:

...we are at a point that we have answers to almost all our questions and [in] these research tasks [we] only pretend that we don't know the answer to please the lecturer and submit a research project that we knew the answer to beforehand. (S7)

This answer illustrates that some of students' actions could well have been influenced by the expectations constructed around the didactic intervention. In other words, teachers' introduction to a new approach to the classroom could have predisposed learners to some extent in their participation and engagement, which could have been

forced at times. This is particularly relevant, as this could potentially undermine their motivation and interest in the approach in and outside the classroom.

In spite of some of the participants' criticism, they all considered that the use of concordancing software promotes their learner autonomy, particularly the ability to design their own research questions for the corpus:

... We were free to decide what to analyze, so we had to decide how to approach [the question] and decide whether it was feasible or not. (S9)

... With these research projects, we now have guidelines to follow in our own future research. (S6)

Two participants add that further practice and familiarization with the software is still needed, as they feel that at times there is not enough time to carry out the search, find the examples, reflect on them and then present them to the class.

The majority of the participants (81%) state that they are likely to use the software again in the future. They affirm that they would use it to find word and expression frequency, as well as to determine common or suitable structures for a particular context. More importantly, the participants declare that they are likely to use it in the future to carry out their own independent research:

I will use it to solve my own questions in the future, like what structure is more natural when writing or speaking. (S4)

...In fact, my classmate and I are thinking of incorporating [the concordancer software] into our thesis. (S11)

When asked to make connections between the affordances the software has as future professional translators, the participants suggest that the concordancer would help them determine patterns, frequency, registers and other elements in language:

[I'd use it] to determine what is the best context to use X word. In addition, [I'd use it] to know which speaker uses more than one word depending on the origin of the series. Also, it'd be useful to know different ways [words] can be pronounced and further phonetic knowledge. (S5)

In addition, some students state that the corpus could serve as a model for honing their language skills, such as writing:

As a translator, it is very important for me to write well. This software and corpus could help me improve [my writing skills]. (S3)

These perceptions suggest that, in spite of the challenges and difficulties concerning the implementation of DDL, the participants are able to identify the value DDL holds in their language learning process beyond the classroom.

The participants propose three recommendations as to how future interventions could be improved. The first one is the expansion of the corpus. They suggest including a wider variety of countries and series to the corpus, while some others propose using other types of materials, such as documentaries. Although they note it is already large, there are occasions in which examples are not available or found.

The second suggestion is to apply this approach to lower levels of the program or to teach other languages. In other words, the participants recognize the value that this approach has in terms of autonomy, which is why they propose that lecturers could try to implement it as soon as learners start the program.

Lastly, they proposed incorporating more specialized texts to compile a technical corpus to be used as a reference for translating purposes. This suggestion presupposes that the participants are able to grasp the potential DDL offers for the development of their TC.

In Summary, although the participants' perceptions of the overall experience are positive in all of the five dimensions explored in the final questionnaire, their perceptions differ in three critical aspects. First, students claim that more British series are required to determine possible contrasts with American ones. This could be due to the influence that British RP has on students in higher education, most of whom feel that it is the standard that they all should aim to achieve. Consequently, the corpus, being rich in American English, differs from what some participants deem as necessary in an English language class. Second, teachers are made responsible for the implementation of the approach, and at times, blamed for the user interphase of the concordancer software. This might be related to the participants' beliefs that result from their educational experiences throughout their lives, where teachers are the ones

that lead the whole learning process and thoroughly prepare all materials. Lastly, some participants' level of English (C1) might play a role in their negative perception of the software, as concordancer enquiries are regarded as an unnecessary exercise.

Teachers' reflections on the implementation

The implementation of the DDL-TB unit presented opportunities and challenges for students as well as the lecturers who carried it out. The participants' perceptions of the proposal have contributed to a reflective, metacognitive analysis on part of the authors who planned, designed and implemented the sessions.

The first reflections are related to the corpus construction and design, for which its size was a critical issue. While literature suggests that it should be 'small' (Aston, 1997; 2002; Gavioli & Aston, 2001; Tribble, 1997), it became apparent that a corpus smaller than 1,000,000 words was insufficient to account for the language phenomena that this study aimed to examine. This was initially reported when first software inductions were conducted, but as students were able to explore the entire corpus, sufficient examples were obtained. Another consideration echoes the remarks made by students concerning the balance of types of Englishes constituting the corpus: most TV series used American English and very few included British English. This conflicted with previous language rules and grammar awareness promoted by the British course books and materials used in previous courses. It is believed that this needs to be considered next time when designing or expanding the corpus.

Second, the balance between DDL and language tasks is another issue to caution in future interventions. As stated in 4, vocabulary tasks were adapted and taken from the course books to provide variety to the sessions, as proposed by Allan (2006). However, this made lexical tasks seem at times rather disconnected from the corpus. Further projects may benefit by articulating tasks coherently with the corpora and moving entirely from the textbook and decreasing the use of textbooks.

Third, it might be necessary to extend the amount of time students engage with DDL tools and tasks. Although Bernardini (2001) suggests that students are eventually able to understand and use DDL concor-

dancing tools, it could well be that a minimum time is needed before this point is achieved. Aston (1996) affirms that hours of practice are needed before students can carry out independent search. Thus, future units might need to consider a longer period of transition and graduation between teacher and student-led searches.

Fourth, the fact that the language phenomenon chosen for the DDL-TB unit was grammar-oriented, i.e. reported speech and verb patterns, made it difficult to provide the necessary guidance for final projects when the participants moved to independent lexically-oriented research. Future implementations could consider thematic and lexical criteria for the creation of corpora.

A final point to consider is the amount of time dedicated to the design and preparation of the sessions. The preparation of the sessions was time-consuming, as tasks needed to be piloted in order to verify the examples that students were likely to find, and also to minimize self-doubts (Kaltenböck & Mehlmauer-Larcher, 2002). Thus, it is important to acknowledge the time involved in the planning and design of similar didactic proposals in the future.

Conclusions

This study implemented a DDL-TB didactic unit for an English language 7 course unit based on Singer's (2016) pedagogical proposal. The activities were designed to graduate learner autonomy and concordance software knowledge. The eight onsite sessions involved working with verb patterns at level C1 and vocabulary and expressions from the different course books suggested in the course outline.

After the application of the course unit, the participants believe the approach to be suitable for translation training programs, as suggested by Singer (2016). Concretely, they highlight its value in terms of authentic language in context, expression frequency and registers. They also commented positively on the corpus content, i.e. their own multimodal consumption. This works as a motivational element throughout the unit. In addition, the participants are able to visualize potential future uses of the approach and concordancer software in their short-term future, mostly as a reference for language use. Furthermore, some students conceive it as a reference to improve their writing skills.

A minority of the participants comment on some issues of the proposal, particularly the meaningfulness of the tool at their current level of proficiency. In their view, the whole procedure becomes irrelevant in their language development. This is strongly related to the feeling of tediousness they experienced while carrying out the searches (Whistle, 1999). Additionally, they also comment that a more balanced variety of series in terms of their origin would be necessary, i.e. more British series. Lastly, they suggest more time would be needed to be fully familiarized with the concordancer software. These perceptions are an important feedback to hone the current didactic proposal and make it more meaningful for students.

Some reflections following the implementation of the DDL-TB unit and students' survey encompass the size and balance of the corpus, task coherence, DDL practice and preparation time. First, although literature suggests a relatively small corpus, it is believed that at C1 level language discussions, a large corpus of no less than 1,000,000 words might be necessary to find sufficient data to work with. The corpus in this DDL-TB unit was almost 6,000,000 and proved to be sufficient. Second, DDL and linguistic tasks should be coherent in terms of topics or genres. This could provide a sense of cohesion to the unit which could foster the integration of the DDL approach. Third, more sessions are needed to introduce the approach and graduate learner autonomy. Future units could potentially last the entire course unit. Finally, it is important to acknowledge the substantial amount of time required to plan and design a DDL-TB approach beyond a single independent unit. It is thought that collaborative work between a team of teachers could lower the workload impact a proposal like this could have on individual teachers.

As to future research, the authors are considering expanding the multimodal consumption to written sources, such as novels, biographies and comics, to cater for a more balanced corpus. This would additionally increase its size, thus amplifying the chances of obtaining examples of language use.

To the authors' knowledge, this exploratory-descriptive study constitutes the one of the first corpus study involving language teaching in the context translation education in Chile, which paves the way for future research in this particular field and settings.

References

- Agencia de Calidad de la Educación. (2018). *Informe de resultados estudio nacional de inglés III medio 2017*. http://archivos.agenciaeducacion.cl/Informe_Estudio_Nacional_Ingles_III.pdf
- Allan, R. (2006). Data-driven learning and vocabulary: Investigating the use of concordances with advanced learners of English. *Centre for Language and Communication Studies, Occasional Paper*, 66.
- Anthony, L. (04/01/2019). *AntConc*. Laurence Anthony web. <https://www.laurenceanthony.net/>
- Aston, G. (1996). The British National Corpus as a language learner resource. In S. Botley, J. Glass, T. McEnery, & A. Wilson (Eds.), *Proceedings of TALC 1996. UCREL Technical Papers* (Vol. 9, pp. 178-191). UCREL.
- Aston, G. (1997). Small and large corpora in language learning. In B. Lewandowska-Tomaszczyk & P. J. Melia (Eds.), *Proceedings of the First International Conference on Practical Applications in Language Corpora* (pp. 51-62). Łódź University Press.
- Aston, G. (2002). The learner as corpus designer. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 9-25). Rodopi.
- Bernardini, S. (2001). Corpora in the classroom: An overview and some reflections on future developments. In J. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 15-36). John Benjamins Publishing Company.
- Boulton, A. (2009). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, 35(1), 81-106.
- Boulton, A. (2011). Bringing corpora to the masses: Free and easy tools for interdisciplinary language studies. In N. Kübler (Ed.), *Corpora, Language, Teaching, and Resources: From Theory to Practice* (pp. 69-96). Peter Lang.
- Boulton, A. & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348-393. <https://doi.org/10.1111/lang.12224>
- Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17(1), 47-64. <https://doi.org/10.1017/s0958344005000510>

- Bygate, M. (2015). *Domains and Directions in the Development of TBLT: A Decade of Plenaries from the International Conference*. John Benjamins Publishing Company.
- Chen, M. & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. *International Journal of Corpus Linguistics*, 23(3), 335-369. <https://doi.org/10.1075/ijcl.16130.che>
- Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment—Companion volume*. Council of Europe Publishing. www.coe.int/lang-cefr
- Cresswell, A. (2007). Getting to 'know' connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo, Q. Luis, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 267-287). Rodopi.
- Education First, (EF). (2020). *EF English Proficiency Index. Latin America*. EF. <https://www.ef.co.uk/epi/regions/latin-america/>
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- European Master's in Translation. (2017). *EMT Competence Framework 2017*. https://ec.europa.eu/info/sites/info/files/emt_competence_fw_2017_en_web.pdf
- Evans, V. & Dooley, J. (2015). *Upstream Proficiency C2. Student's book* (4th ed.). Express Publishing.
- Evans, V., Dooley, J., & Edwards, L. (2014). *Upstream Advanced C1. Student's book* (3rd ed.). Express Publishing.
- Flowerdew, L. (2015). Data-driven learning and language learning theories: Whither the twain shall meet. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15-36). John Benjamins Publishing Company.
- Gavioli, L. & Aston, G. (2001). Enriching reality: Language corpora in language pedagogy. *ELT Journal*, 55(3), 238-246. <https://doi.org/10.1093/elt/55.3.238>
- Hurtado Albir, A. (2011). *Traducción y traductología. Introducción a la traductología*. Cátedra.
- Johns, T. F. (1991). Should you be persuaded—two examples of data-driven learning. *ELR Journal*, 4, 1-16.

- Johns, T. F. (2000). Data-driven learning: The perpetual challenge. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (Vol. 42, pp. 107-117). Brill/Rodopi.
- Jones, C. & Waller, D. (2015). *Corpus linguistics for grammar a guide for research*. Routledge.
- Kaltenböck, G. & Mehlmauer-Larcher, B. (2002). Teaching ESP: How text corpora can help. In A. Pulverness (Ed.), *IATEFL 2002: York conference selections* (pp. 31-33). IATEFL.
- Kennedy, C. & Miceli, T. (2001). An evaluation of intermediate learners' approaches to corpus investigation. *Language Learning & Technology*, 5(3), 77-90. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44567/1/05_03_kennedy.pdf
- Kress, G. (2010). *Multimodality. A social semiotic approach to contemporary communication*. Routledge.
- Leńko-Szymańska, A. & Boulton, A. (Eds.). (2015). *Multiple affordances of language corpora for data-driven learning*. John Benjamins Publishing Company.
- Long, M. (2015). *Second Language Acquisition and Task-Based Language Teaching*. John Wiley & Sons.
- Meunier, F. (2002). The pedagogical value of native and learner corpora in EFL grammar teaching. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (Vol. 6, pp. 119-141). John Benjamins Publishing Company.
- Norris, R. & French, A. (2016). *Ready for Advanced Coursebook* (3rd ed.). Macmillan.
- Robinson, D. (1991). *The translator's turn*. The John Hopkins University Press.
- Sinclair, J. (2003). *Reading concordancers: An introduction*. Pearson.
- Singer, N. (2016). A proposal for language teaching in translator training programmes using data-driven learning in a task-based approach. *International Journal of English Language & Translation Studies*, 4(2), 155-167.
- Singer, N. & Basaure, R. (2018, June 21). *Lo que no se quiere discutir: ¿Quiénes son los profesores de lengua inglesa en programas de formación de traductores en Chile?* [Conference]. IV Congreso Internacional sobre investigación en Didáctica de la traducción, Universitat Autònoma de Barcelona,

- Barcelona, España. https://grupsderecerca.uab.cat/pacte/sites/grupsderecerca.uab.cat/pacte/files/cuaderno_congreso_20_06_2018.pdf
- Singer, N., Rubio, M., & Rubio, R. (2018). Representaciones de estudiantes de traducción en el aprendizaje de una lengua extranjera. *Onomázein*, (39), 245-269. <https://doi.org/10.7764/onomazein.39.10>
- Springfield. (2020). *Springfield! Springfield! TV & Movie Scripts*. <https://www.springfieldspringfield.co.uk>
- Thornberg, R. & Charmaz, K. (2014). Grounded theory and theoretical coding. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 153-169). Sage.
- Tribble, C. (1997). Improvising corpora for ELT: Quick-and-dirty ways of developing corpora for language teaching. *Proceedings of the First International Conference. Practical Applications in Language Corpora*, University of Lodz, Poland. https://www.tribble.co.uk/text/Tribble_C_Palc_97.pdf
- Tribble, C. (2012). Corpora in the language teaching classroom. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (p. wbeal0226). Blackwell.
- Whistle, J. (1999). Concordancing with learners using an 'off the web' corpus. *ReCALL*, 11(2), 74-80. <https://doi.org/10.1017/s0958344000004985>
- Widdowson, H. G. (1978). *Teaching language as communication*. Oxford University Press.

Appendix 1

Online survey

Personal background

- 1) Age
 - 2) Section (Class)
 - 3) Years studying English
-

Perceptions about the teaching-learning process

- 1) How would you describe a good English language teacher?
 - 2) How would you describe a good English language learner?
 - 3) What would a very good English class be like?
 - 4) What would a bad English class be like?
 - 5) What elements should be included in a language class in order for it to be meaningful or relevant to you?
 - 6) What elements should be removed from a language class in order for it to be meaningful or relevant to you?
 - 7) Do you think it is necessary to include technology in a language class? Why?
-

Interests

- 7) What series, which are originally in English, do you watch?
 - 8) Do you watch the series subtitled or dubbed?
 - 9) On which platform do you watch the series?
 - 10) In which device do you watch the series?
 - 11) How often do you watch series?
 - 12) Do you pay attention to any idiomatic aspects when you watch series? Which one(s)?
 - 13) In which way do you think watching series specifically helps you hone your English language skills?
-

Source: Self-made.

Appendix 2

Corpus series

13 Reasons Why	Rick and Morty
A Discovery of Witches	Riverdale
Bates Motel	Safe
BoJack Horseman	Shooter
Brooklyn 99	The 100
Chilling Adventures of Sabrina	The Alienist
DaVinci's Demons	The Crown
Disjointed	The Haunting of Hill House
Hannibal	The OA
House of Cards	The Punisher
How to Get Away with Murder	The Resident
Killing Eve	The Sinner
Love Death + Robots	The Umbrella Academy
Lucifer	Tidying Up with Marie Kondo
My Mad Fat Diary	Titans
One Day at a Time	True Detective
Orphan Black	Vikings
Peaky Blinders	You
Queer Eye	

Source: Self-made.

Appendix 3

Final survey

- 1) Did you find it interesting to work with a piece of software in class?
Why?
 - 2) Do you think the corpus was appropriate to be used in class? Why?
 - 3) Do you think instructions given by the lecturers about the concordancer software tools were sufficient to carry out a good work?
 - 4) Do you think you learned more about English grammar and language use than you would have in a more traditional class?
 - 5) Do you think the final project helped you understand English language in a more real context?
 - 6) Do you think the final project fostered the development of your linguistic and researcher skills?
 - 7) Did you have enough autonomy to carry out your own searches and create your own research questions?
 - 8) After this semester, do you feel capable of carrying out searches autonomously?
 - 9) How do you plan to use the software and corpus in the future?
 - 10) How do you think the use of this software and corpus helped you develop your translator competence skills?
 - 11) Have you got any further comments or suggestions regarding the software and corpus?
-

Source: Self-made.