Automating task design: bridging the gap between second language research and L2 instruction

Joan CASTELLVÍ VIVES (contact author for correspondence)
Universitat de Barcelona (España)
joan.castellvi@ub.edu
https://orcid.org/0000-0003-4676-9643

Roger GILABERT GUERRERO
Universitat de Barcelona (España)
rogergilabert@ub.edu
https://orcid.org/0000-0001-6464-4161

Elisabet COMELLES PUJADAS
Universitat de Barcelona(España)
elicomelles@ub.edu
https://orcid.org/0000-0002-4753-2712

Abstract: The present study addresses the issue of the transfer from second language acquisition (SLA) research to second and foreign language (L2) instruction, particularly within the context of Task-Based Language Teaching (TBLT). The main goal of this article is to show how technology, particularly the web-based tool *taskGen* (https://taskgen.eu), which primarily serves to create teaching materials, also functions as a mediator for transfer and provides valuable information on teaching practices, teachers' needs, pedagogical strategies, and L2 task design. This article is organised as follows. Firstly, a web-based tool that assists and trains teachers in L2 task design is presented, along with the instantiation and integration of natural language processing (NLP) resources in the tool; later, we show how quantitative information obtained from the massive use of the tool through data analytics can be combined with qualitative methods such as interviews, observations, think aloud protocols, questionnaires, case studies, and transcripts to generate research. Automatic (quantitative) data collection gathers information of the teaching contexts, task aims, task flow structures, focus on form options, and the access and time on help files. Not only does the collected information feed back to the tool, but it also contributes to understanding teachers' design choices and patterns, and to exploring the intersection between SLA and TBLT theory and task design practises. Qualitative analysis is being conducted and oriented toward teachers' experiences during task design. Finally, we present the future lines of research that we anticipate will be carried out with the further use and development of *taskGen*.

Keywords: educational technology; computational linguistics; task-based language teaching; second language task design

Català:

Automatització del disseny de tasques: creant ponts entre la recerca en segones llengües i la seva instrucció

Resum: *El present estudi planteja la qüestió de la transferència de la recerca en l'adquisició de segones llengües (SLA) a la instrucció de segones llengües i llengües estrangeres (L2), més concretament en el marc de l'Ensenyament de Llengües Basat en Tasques (TBLT). L'objectiu principal d'aquest article és mostrar com la tecnologia, i en particular l'eina web taskGen (https://taskgen.eu), l'ús primer de la qual és la creació de materials didàctics, també funciona com a mitjà per a la transferència i proporciona informació valuosa sobre les pràctiques d'ensenyament, necessitats dels professors, estratègies pedagògiques i el disseny de tasques en L2. Aquest article està organitzat de la següent manera. En primer lloc, es presenta taskGen com a una eina web que ajuda i forma els professors en el disseny de tasques L2 i que integra recursos de processament de llenguatge natural (NLP); a continuació es mostra com la informació quantitativa obtinguda a través de l'anàlisi de dades a partir l'ús massiu de l'eina es pot combinar amb mètodes qualitatius com entrevistes, observacions, protocols d'opinió de l'usuari en directe, qüestionaris, estudis de casos i transcripcions per generar recerca. La recopilació automàtica (quantitativa) de dades recull informació dels contextos d'ensenyament, objectius de la tasca, estructures del flux de tasques, opcions d'atenció en la forma, i l'accés i el temps als fitxers d'ajuda. La informació recopilada, a més de retroalimentar l'eina, contribueix a comprendre les opcions i patrons en els dissenys dels professors, i a explorar la intersecció entre la teoria SLA i TBLT, d'una banda, i les pràctiques de disseny de tasques, de l'altra. L'anàlisi qualitatiu que es du a terme s'orienta cap a les experiències dels professors durant el disseny de tasques. Finalment, es presenten les línies de recerca futures que preveiem que aniran dunt a terme a mesura que creixi l'ús de taskGen i es vagin implementant més funcionalitat de l'eina.*

Paraules clau: *tecnologia educativa; lingüística computacional; ensenyament basat en tasques; disseny de tasques per a segones llengües*

Spanish:

Automatización del diseño de tareas: creando puentes entre la investigación en segundas lenguas y su instrucción

Resumen: *El presente estudio aborda la cuestión de la transferencia de la investigación en la adquisición de segundas lenguas (SLA) a la instrucción de segundas lenguas y lenguas extranjeras (L2), más concretamente en el marco del Enseñanza de Lenguas Basada en Tareas (TBLT). El objetivo principal de este artículo es mostrar cómo la tecnología, y en particular la herramienta web taskGen (https://taskgen.eu), cuyo primer uso es la creación de materiales didácticos, también funciona como medio para la transferencia y proporciona información valiosa sobre las prácticas de enseñanza, necesidades de los profesores, estrategias pedagógicas y el diseño de tareas en L2. Este artículo está organizado de la siguiente manera. En primer lugar, se presenta taskGen como una herramienta web que ayuda y forma a los profesores en el diseño de tareas L2 e integra recursos de procesamiento de lenguaje natural (NLP); a continuación, se muestra cómo la información cuantitativa obtenida a través del análisis de datos a partir del uso masivo de la herramienta se puede combinar con métodos cualitativos como entrevistas, observaciones, protocolos de opinión en vivo, cuestionarios, estudios de casos y transcripciones para generar investigación. La recopilación automática (cuantitativa) de datos recoge información de los contextos de enseñanza, objetivos de las tareas, estructuras del flujo de tareas, opciones de atención en la forma, y el acceso y el tiempo a los archivos de ayuda. La información recopilada, además de retroalimentar la herramienta, contribuye a comprender las opciones y patrones en los diseños de los profesores, y a explorar la intersección entre la teoría SLA y TBLT, por un lado, y las prácticas de diseño de tareas, por otro. El análisis cualitativo que se lleva a cabo se orienta hacia las experiencias de los profesores durante el diseño de tareas. Finalmente, se presentan las líneas de investigación futuras que prevemos que se llevarán a cabo a medida que crezca el uso de taskGen y se vayan implementando nuevas funcionalidades en la herramienta.*

Palabras clave: *tecnología educativa; lingüística computacional; enseñanza basada en tareas; diseño de tareas para segundas lenguas*

Introduction

It is not uncommon in the field of applied linguistics to hear scholars expressing their concerns regarding the general lack of transfer from second language acquisition (SLA) concepts and constructs to second and foreign language (L2) instruction (Sato & Lowen, 2019). The connections between SLA theory and L2 instruction have been described as indirect and complex, and have been said to range from tenuous to non-existent (Gregg, 2001). From a task-based language teaching (TBLT) perspective specifically, the contribution of SLA research to language teaching practice has been recurrently discussed over the years (Nunan, 1991; Ellis, 1997; Markee, 1997; MacDonald *et al.*, 2001; Malicka

et al., 2019; Gilabert, 2023). TBLT as a teaching approach has seen an exponential growth in the last three decades, and it is progressively becoming mainstream in language education to the point of being part of national language policy in countries like Thailand or China. One of the central tenets of TBLT as a teaching approach is its coherence with SLA findings and its capacity to incorporate practical applications of advances in SLA research.

In the paper that follows our main claim is that technology can serve not only as a potential mediator for such transfer but also as a powerful source of research into teacher choices and practices during task design. After a brief description of the *taskGen* tool that assists, partially automates, and trains teachers in L2 task design, the goal of this paper is to gauge the research potential of such a tool. As we describe the tool, we point out the potential for research of the tool's features. Research is conducted by combining quantitative information obtained from the massive use of the tool through *data analytics* and qualitative methods such as interviews, observations, and think aloud protocols.

1. The automated task design generator: *taskGen*

1.1 Interface architecture and organisation

Based on the principles of TBLT models and frameworks (Long, 2015; Ellis, Skehan, Shintani & Lambert, 2019), the automated task design generator, *taskGen (*https://taskgen.eu*)*, is a semi-automated, flexible, and adaptive web-based tool designed to assist L2 task designers in creating tasks for L2 learning. It guides teachers in the design of communicative tasks, including their associated pre-tasks and post-tasks. The tool came as a reaction to claims in the literature (Ellis *et al.*, 2019; Gilabert, 2023) that constructs coming from second language acquisition in general, and TBLT in particular, were not being transferred sufficiently to teacher design practices. As widely accepted in the field, a task can be defined as a process with a communicative goal and outcome. Learners must follow a series of steps and draw on their cognitive and communicative resources in order to complete the task, which is typically assessed in terms of task completion.

In the spirit of that definition, *taskGen* guides teachers/designers to decide, firstly, on the communicative goal of the task (see Figures 1 and 2) as well as the main task organization, before they are guided to create the pre-tasks that precede and support the main tasks, or the post-tasks that consolidate the concepts and language involved in the pre-task and main task. That way the tool guarantees that the task is driven by a communicative objective rather than by a linguistic one, as it would be the case in traditional language teaching.

Figure 1

Task structure includes a main task preceded by a pre-task phase and post-task phase.
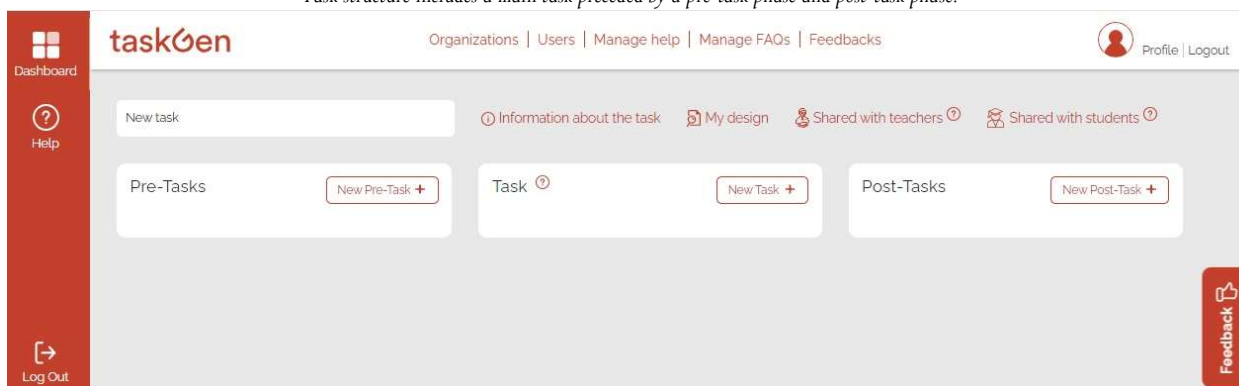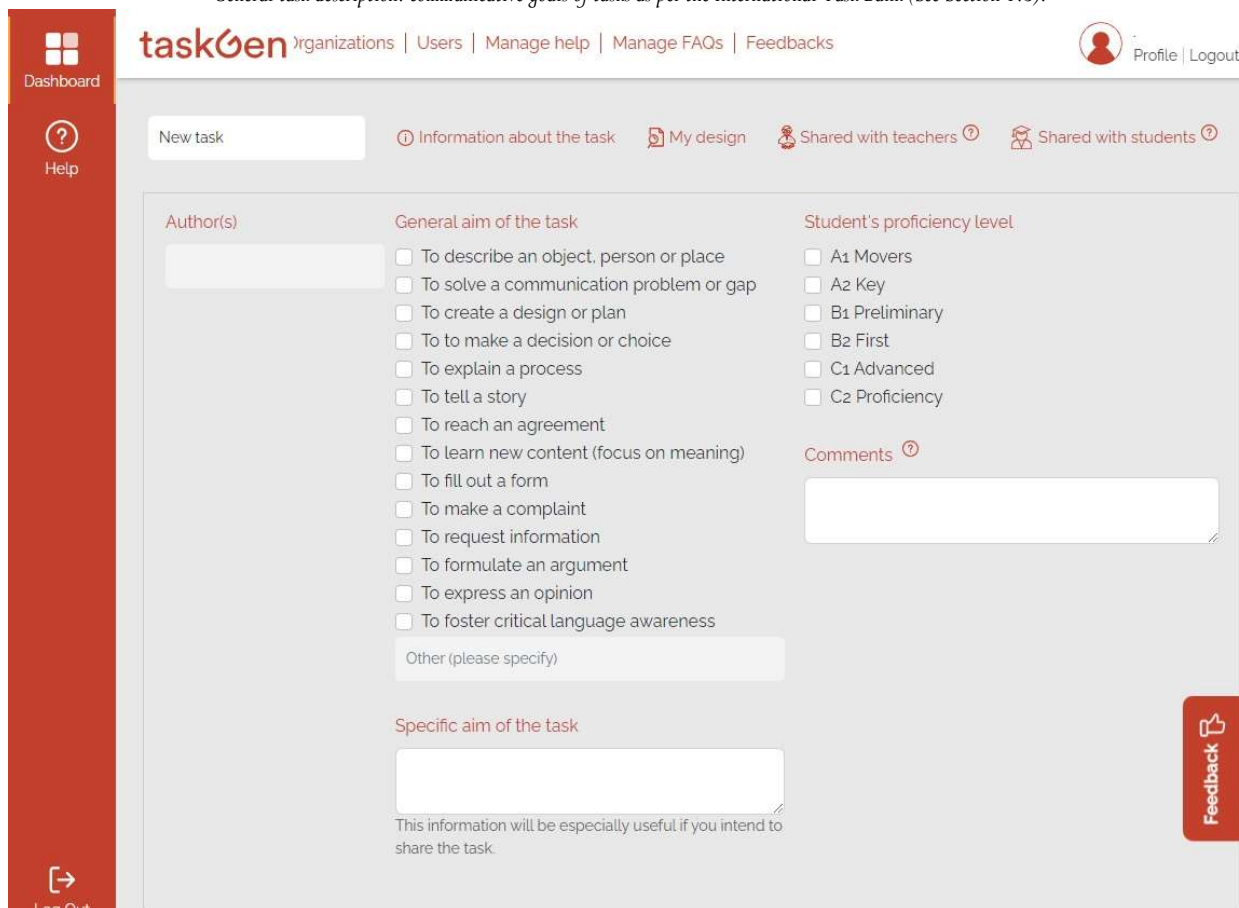

Figure 2

General task description: communicative goals of tasks as per the International Task Bank (See Section 1.3).



The main task interface allows the configuration of any type of task (i.e. problem-solving tasks, decision-making tasks, narrative tasks…). Teachers/designers can choose the following parameters: the number groups within

a class, the number of participants per group, their roles in the task (e.g. the hotel receptionist and two hotel guest roles), and prompts or materials received by the participants (see Figure 3). This will be important in order to assess through *data analytics* the impact of the concept of interaction on task design.

Figure 3
The main task interface: organisation of learners into groups and assignment of roles and prompts (materials).
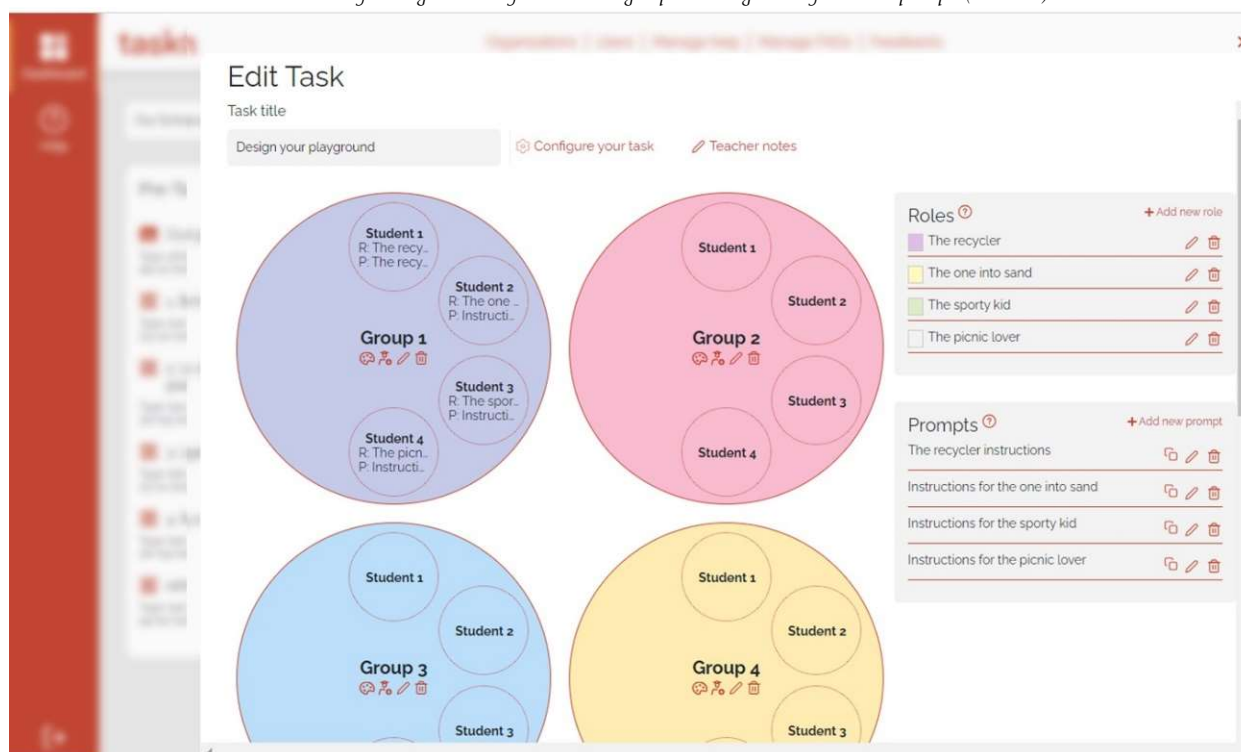
As teachers make those decisions, they can use help options (help menu and FAQs - see Figure 4) to obtain information about interactive design factors and their consequence for language use and learning (i.e. whether if information is split or shared among participants; whether information flows in one way, two ways, or in multiple ways; if tasks are made open or closed as so whether many solutions are possible or only a single one is; whether learning has convergent and divergent goals within the task). They can also see examples of how other designers used those variables in their designs, as well as the original references. In terms of research, and also through both qualitative interviews and observation, as well as through analytics, we will be allowed to measure the degree of support that teachers use during their design (see Section 2 below).

Figure 4
Help menus and FAQs can be used to aid design choices.
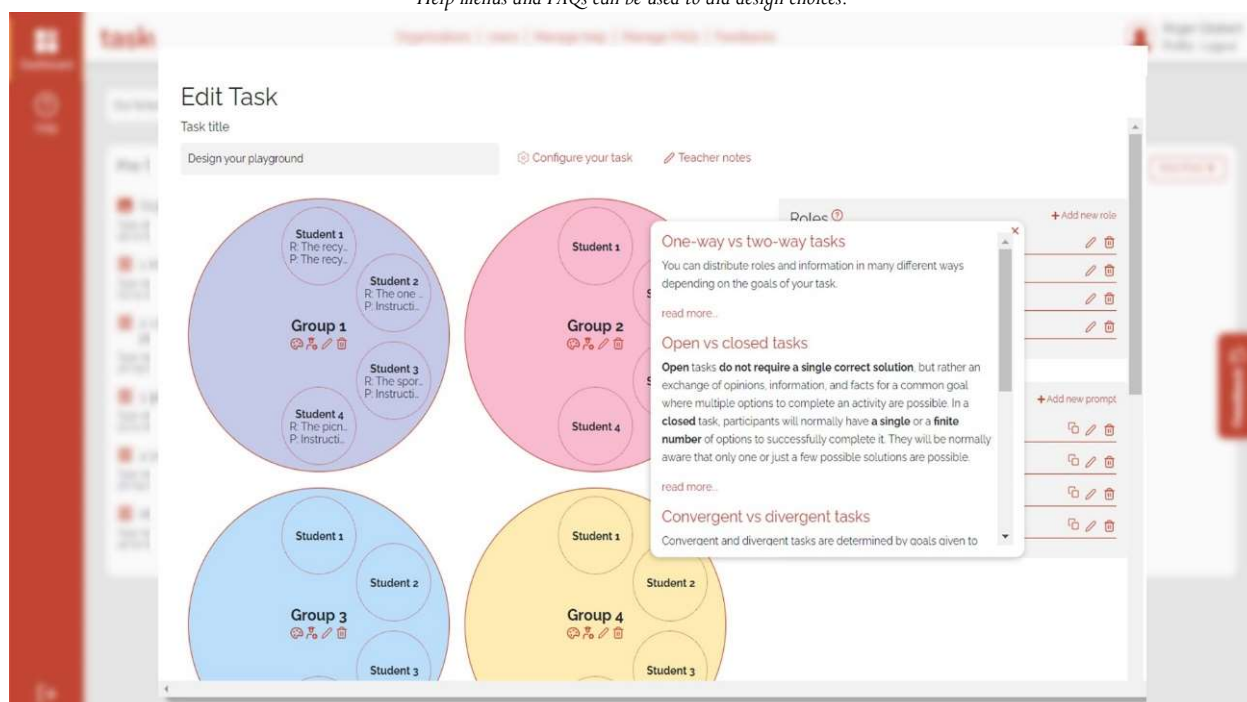
When texts are introduced in the tool (for example, in a pre-task), *taskGen* incorporates an input analysis module with natural language processing (NLP) resources for the English language. Such functionality detects and provides information related to morphological, syntactic features, multi-word units (phrasal verbs, collocations) and pragmatic expressions for the textual input provided in the task prompts, as well as in any other text in the pre-tasks and post-tasks. This allows for the application of focus on form techniques (Long, 2015), such as input enhancement (see Section 1.2). Again, while teachers are encouraged by the literature and in training courses to use focus-on-form techniques during task design, it is an empirical question whether they use them or not, and how exactly they use them (e.g. what linguistic dimensions do they focus on? How do they draw attention to certain features through input enhancement?) Data analytics can provide accurate and precise information about how focus on form happens during design.

Figure 5

The taskGen dashboard: user's profile, creation, edition, sharing and deletion of tasks, and access to the general help menu.
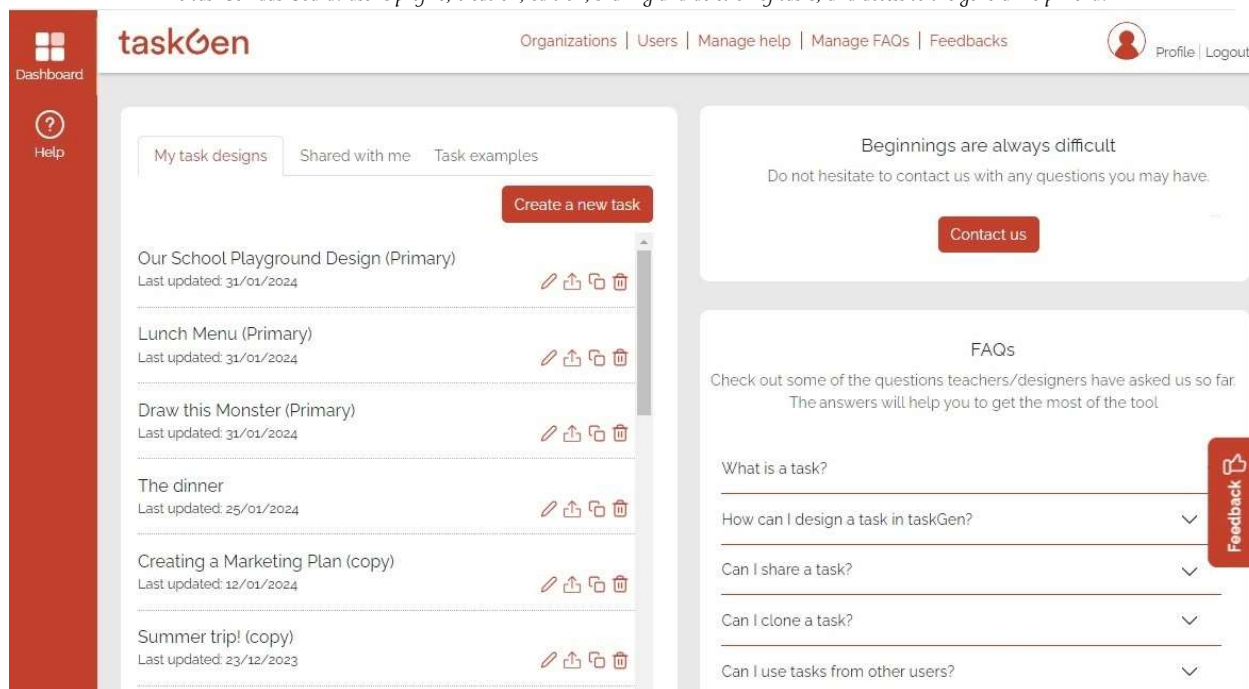
In addition to the dashboard, the main task module, and the pre- and post-task modules, all teaching materials created in the task, pre-task, and post-task interfaces can be viewed and managed in the output module. This module allows the possibility of cloning and sharing the tasks created with other colleagues and *taskGen* users (teacher site), and it also generates an output with all the task elements that can be presented to students in class or online (student site). The interface allows for the selection and sharing of the whole task or particular task elements (see Figures 5 and 6). This way teachers can clone and adapt their tasks to the diverse student populations in their classes, by creating slightly different versions of tasks for different students (with different proficiency levels, or with different needs) within a group or across groups. In sum, teachers can personalise their teaching materials, which is interesting in all contexts, but particularly so in the context of primary and secondary schools with highly diverse groups in our local teaching context. Quantitative analysis will show us whether teachers clone their own tasks to create different versions of the same tasks for different students, and so we will gauge the degree of adaptation and personalisation that teachers apply to their designs.

Figure 6

The taskGen task design interface: pre-task, during-task and post-task; access to the output module: 'My design', 'Shared with teachers'
and 'Shared with students' mini-sites
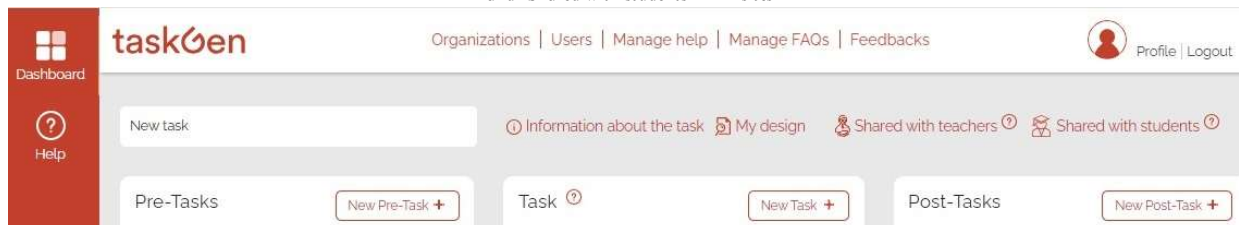
As a main contribution of this project, the solution to the problem of transfer from SLA research to actual task design is achieved by integrating SLA, task design, and focus on form concepts directly in the tool, and by collecting data about actual task design and what surrounds it. It must be noted that our learning curve has become exponential. Through both qualitative and quantitative methods described in section 2, we have started to gain understanding of how much knowledge about SLA and task design concepts teachers bring into their task design, how they may use the available options provided by the tool in their task designs, and their perception of what task design can do for them and their learners. In the building of the tool, we also detected uses and effects that we had not originally anticipated and that went beyond the original idea of the tool, such as its potential for task co-construction, for personalization and adaptation, for community building, or its training dimension. A major conclusion from our current project so far is that, to guarantee a comprehensive picture of task design, it is not enough to draw on knowledge coming from SLA and TBLT. An interdisciplinary effort is required instead. In our case such knowledge has come from the fields of computational linguistics and natural language processing, learning design, and teacher cognition.

The *taskGen* tool has been developed within a three-year state funded project and it is currently operational worldwide, serving the global community of teachers and task designers and being piloted by hundreds of users. Pilot testing of the tool is being leveraged to conduct a controlled experiment with a twofold purpose: on the one hand, to assess the functionality of the tool by teachers and, on the other, to collect information about the teachers' design choices that teachers make through both qualitative (interviews, observations) and quantitative methods (data analytics) (see Section 2).
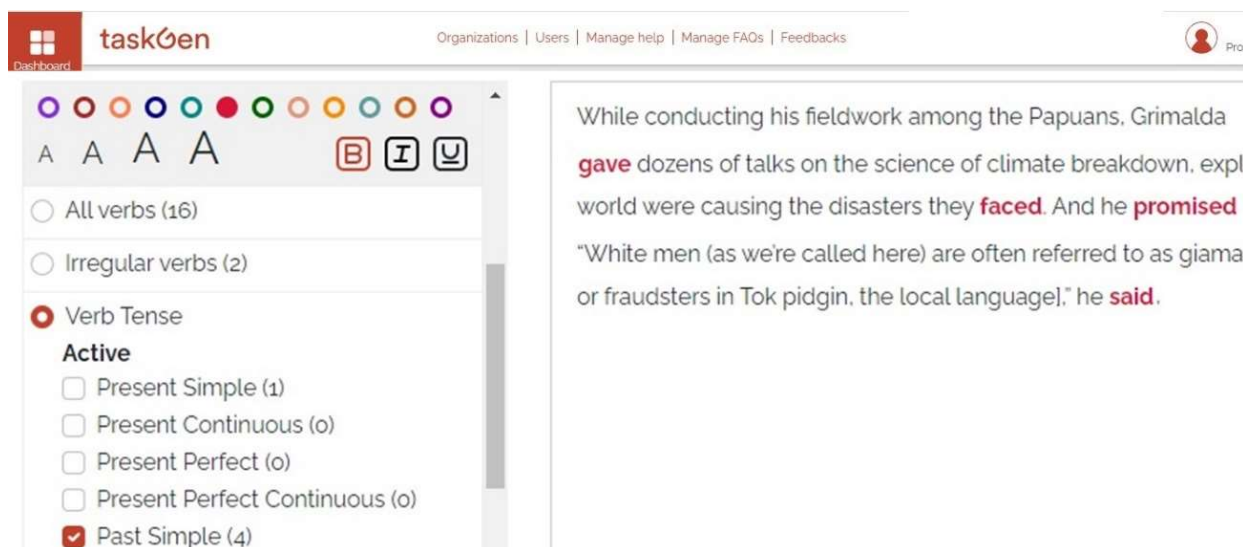
1.2 The NLP module

Natural Language Processing (NLP) is a subfield of Artificial Intelligence (AI), which aims to enable computers to understand, interpret, and generate human language. Within NLP there are very well-known tasks, such as text understanding, text generation, machine translation, and speech recognition. NLP has practical applications in various contexts covering healthcare (analysis of clinical documentation), customer service (chatbots), etc. It has also been applied to the teaching and learning of languages in several ways that may include automated writing evaluation (AWE)

tools to automatically assess L2 students' writings and give them feedback (Koltovskaia, 2020; Wang *et al.*, 2020; Guo & Wang, 2023); generative AI as a (multimodal) writing aid (Warschauer *et al.*, 2023; Kang & Yi, 2023); machine translation as a writing aid (Lee, 2020; Lee, 2021; Stapleton, 2021); dialogue systems to allow learners to interact with conversational agents in a meaningful way (Sydorenko, 2015; Divekar *et al.*, 2021; Bibauw *et al.*, 2022a; Bibauw *et al.*, 2022b; Zhai & Wibowo, 2023); readability and text level identification systems (Filighera *et al.*, 2019; Martinc *et al.*, 2021), or text simplification tools (Alva-Manchego *et al.*, 2020; Grabar & Saggion, 2022).

In the *taskGen* project, the focus of the NLP module is specifically directed towards automating the processing of written texts. The NLP module aids teachers in swiftly identifying and extracting pertinent linguistic information so that they can easily apply focus-on-form techniques automatically, such as input enhancement (through colour and font size highlighting) (see Figure 7).

Figure 7

Example of automatic detection and input enhancement of linguistic items through colouring through the "analyze" function of taskGen.


The NLP module in *taskGen* automatically analyses a piece of written text entered by the teacher and identifies linguistic traits in the text at different linguistic levels (i.e. morphology, syntax, and pragmatics). This automatic analysis is performed thanks to already existing tools, such as Stanza NLP package (Qi *et al.*, 2020), a neural pipeline with several processors for tokenization, lemmatization, PoS-tagging and dependency parsing, as well as other resources which have been especially developed for the *taskGen* project (e.g. pragmatic formulae dictionary).

Although the tool is still under construction, we can currently identify the following traits: frequency of a word within the text, word class (e.g. nouns, adjectives, prepositions), comparative and superlative adjectives, irregular nouns and irregular verbs, verb tense (e.g. past simple, present continuous), modal verbs, phrasal verbs and pragmatic formulae (e.g. 'Why don't you […]?', 'Good morning', 'Could you please […]?'). Moreover, the module

also computes the number of occurrences of the above-mentioned linguistic traits in the text, which may help the teacher decide if a text is appropriate (i.e. including certain structures) for their teaching goal. That is, if a teacher wants their students to carry out a storytelling task, he/she may like to include a text rich in verb tenses used in narratives as part of the pre-task. The NLP module will help by automatically providing the number of such verb tenses in the text, allowing the teacher to decide if the text is appropriate for that pre-task (see Figure 8).

Figure 8
Number of Past tenses in the text
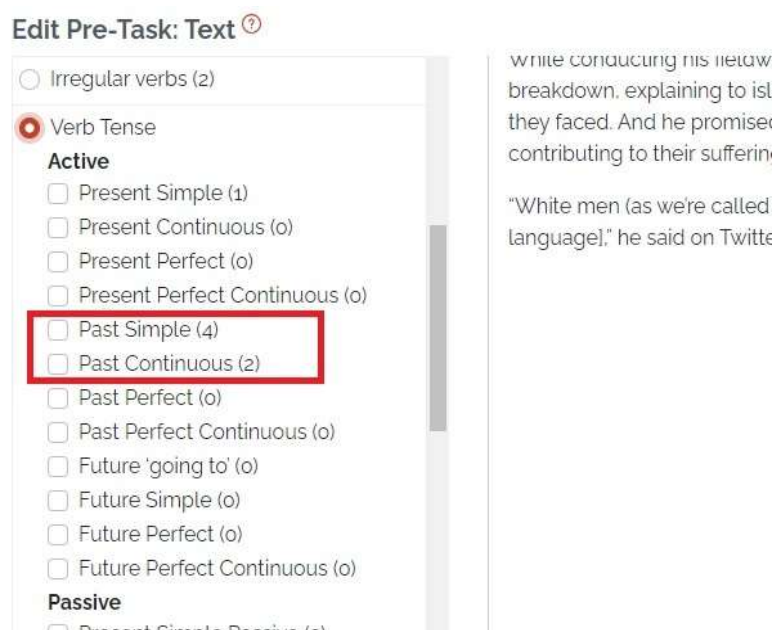
1.3 Automatic data collection

In this section we provide information about the kind of information that is automatically collected by the tool and that users consent to sharing. This includes information about the teaching context, the aim of the designed tasks, the task structure and the type of activities included in pre-tasks, tasks, and post-tasks, as the theoretical aspects task designers are interested in when they look at aids and FAQs.

- **The teaching context**

  The system asks for information about where user usually teaches (university, secondary school, primary school, private tutoring, language school, or pre-service training), their teaching modality (in-person instruction, fully online or blended instruction), their general teaching framework (TBLT, content-and-language integrated learning, project-based learning, language for specific Purposes, language for academic purposes, special needs), their teaching experience, as well as the student's proficiency level (Common

European Framework of Reference, levels A1 to C2). Through quantitative analysis, all these categories can be used to observe emerging behaviours in the different contexts and from different perspectives.

- **The aim of tasks**

General aims are in line with the aims suggested by the *International Task Bank* (https://tblt.indiana.edu/index.html). They include: describing an object, person or place; solving a communication problem or gap; creating a design or plan; making a decision or choice; explaining a process; telling a story; to reach an agreement; learning new content (focus on meaning); filling out a form; making a complaint; requesting information; formulating an argument; expressing an opinion; and fostering critical language awareness, among others. The tool also provides a space for the specific aim of the task (e.g. to provide appropriate, polite and communicatively effective answers to requests via email; learn how to buy groceries in the L2, and find the best price for the items on a shopping list, among others. Task goals may also be associated with different design options, and our quantitative analysis of teachers' choice may give us information about how different task types may generate different designs.

- **The task flow structure**

Information is automatically obtained about the structural complexity of the task phase, that is, how teachers organise the class individual work, work in pairs, small group work or whole class groups. The system also keeps track of pre-task (e.g. audio, video, text…) and post-task choices (e.g. task repetition, oral-to-written/written-to-oral, among others). Such information also includes what task designers begin with when creating a new task, the proportion of skills in their designs (e.g. number of texts, video and audio files, etc.), and the type of interaction they choose (e.g. one-way and two-way tasks, open and closed tasks, or convergent and divergent tasks). To our knowledge, research into task design has never shown what teachers' design may look like in terms of structure. As we will address below, we are already able to obtain data from system use of task structures in our quantitative analysis.

- **Focus on form**

When learners use the text analyser included in pre-tasks, tasks, and post-tasks, information is gathered about the linguistic dimensions that teachers choose to draw their learners' attention to. Information can be obtained about the choices of pragmatic, morphosyntactic, and syntactic dimensions as well as the focus on form techniques that teachers make (see Section 3 below for a detailed description). This is of crucial importance since focus on form has been assumed rather than empirically checked, and data analytics will provide us with accurate information about how teachers draw attention to form in their designs.

- **The help files**

  The help menu is also an important potential source of information. As it is organised as a digest of information related to task design with references to the relevant literature of the concepts and constructs, the count of clicks and the time they spend on aids and FAQs can provide valuable information about how many times designers request assistance and read about task design, and for how long. To our knowledge, we have never empirically checked the degree with which teachers rely on the existing definitions, concepts, and literature in TBLT. Our data analytics approach will provide us with such information.

2. Generating qualitative and quantitative research

The *taskGen* project was originally conceived as an innovation project in which we set out to resolve the issue of the transfer from concepts and theories from SLA and TBLT to the actual design practice by teachers, and the problem of lack of training and resources of the design of communicative tasks for second language classrooms, both locally and universally. This has been resolved by developing the *taskGen* tool by means of digital innovation by integrating theory-based web design and natural language processing into pedagogical design. We have also initiated our understanding of teachers' choices during task design through qualitative and quantitative research.

Two research studies associated with the *taskGen* tool combine qualitative and quantitative methods: Arnold's (2023) study on task design perception by freelance and private language school teachers' and Trager's current PhD thesis on both teachers' perceptions of task design in general and *taskGen* in particular, and teachers' actual use of options during task design. Qualitative methods such as interviews, think aloud protocols or task observation tap into teachers' mental processes and the rationale involved in task design. Quantitative methods provide massive information about users' choices during task design in a way that, to our knowledge, has not been explored before. What are teachers' thought processes during task design? What are their needs and gaps? What are their choices and how are they connected with what we know about SLA concepts and constructs?

2.1 Qualitative research

Arnold's (2023) set out to discover how EFL/ESL freelance and private language school teachers (1) understand and define pedagogic tasks in the language learning classroom, (2) how they integrate pedagogically-oriented technological innovation in lesson planning for the language learning classroom, and (3) how do EFL/ESL freelance teachers engage in decision-making during the process of lesson planning and task design. Eighteen independent freelance teachers or private language school teachers with at least 1 year teaching experience participated in the study. Only 44.4% reported familiarity with TBLT, with 55.5% reporting either uncertainty or lack of familiarity. In order to answer the three research questions, Arnold used semi-structured in-person or virtual interviews with 18 teachers, and a think aloud protocol with two of the teachers. The protocol integrated questions about teacher practices (e.g. "how do you begin

TEISEL

Tecnologías para la investigación en segundas lenguas

ISSN 2696-676X

UNIVERSITAT DE BARCELONA

Facultat de Filología
i Comunicació

in lesson planning?" or "how were you trained to plan your lessons?") beliefs (e.g. "are you familiar with TBLT?" or "how often do you use technology in your classes") with a virtual walk-through of the *taskGen* program (e.g. "where would you intuitively click next?" or "what type of technical support would you like to see here?"). Audio recordings, automatic transcription software and screen recording were used to collect the answers and behaviours from participants. With two of the participants Arnold used think aloud protocols in which participants were asked to verbalise their thoughts as they used a functional version of the *taskGen* tool. Interviews yielded 1,009 minutes of screenshot/audio recorded data with the think aloud protocols adding 50 minutes of video and audio-recorded data. Transcripts were analysed through the process of labelling and coding emergent themes through *Atlas.ti* (https://atlasti.com), a qualitative data analysis software. Results revealed very different levels of training among teachers, with most teachers having received little or no training in lesson planning. The starting point of a lesson ranged from student interests, a miscellaneous topic, a textbook, a learning goal, something entertaining or engaging, a specific speaking skill, a video, and pre-made online lesson plans. As for teachers' definitions of tasks varied, teachers' perceptions of a task generally revolved around the concept of an action, which is in line with definitions of tasks in the TBLT literature. Other findings related to teachers' belief revealed that they were relatively unconvinced by their ability or availability to design tasks. The main concern for many teachers was access to a wider community, where their tasks would benefit others, and they would benefit in return. Finally, teachers reported that reusing by cloning a previously researched task could be modified to various complexity levels and applied to their students' needs. Rather than starting from scratch, using a pre-existing and pre-vetted task design made the adaptation far less daunting.

What we would like to highlight from Arnold study is the kind of data that came from semi-structured interview which directly tapped into teachers' beliefs (e.g. "*I feel like a task is using tactile learning in addition to audiovisual-type learning.*") and practices (e.g. "*I actually adore them [tasks]. But it's always the problem for me that they're very time-consuming, that's why, for example, last year I didn't organise anything of that kind because I didn't feel like…I had enough energy for that*"). Most importantly, the findings by Arnold helped us shape and adapt our tool design to the needs and concerns of teachers. A clear and direct example of that is teachers' interest in sharing their design and learning from other teachers' designs in a wider community of teachers (e.g. "*…When you do put the time into creating something, the return on investment often feels like…if you share it with other people…you're really happy that other people are using it. It's nice when you create something and you're like…I want someone else to have access to this. Look how great this is. Please use it*"). Such a crucial finding led us to the creation of a sharing function of task designs whereby *taskGen* users can share their task designs with other *taskGen* users.

Trager's ongoing PhD study has made use of both a qualitative and quantitative approach. With the aims of exploring teacher needs for lesson planning, understanding the appropriation of the tool by teachers, and improving the tool's usability, Trager has already conducted three qualitative studies employing semi-structured interviews,

observations and think aloud protocols. In study 1, seven primary school EFL teachers and two pre-service teachers participated in semi-structured interviews similar to Arnold's with the non-functional prototype. In study 2, think aloud protocols while using the *taskGen* tool were employed with thirty pre-service teachers and six primary school teachers. In the third study, thirty-three pre-service and inservice teachers answered an online usability survey. Results from Trager's study suggest that task construction with task cycle format (pre-tasks, tasks, post-tasks) is working satisfactorily. Pre-service teachers particularly valued the tool suggestions of various pre- and post-tasks, and highlighted how the tool brings structure to their designs. In general, however, teachers underutilised the aids and recommendations. Pictures, videos, and real objects are used in pre- and post-tasks, and pre-tasks were designed for whole class use, while most main tasks happen in small group work. 84.1% of the participants in the survey agreed that *taskGen* was easy to use and 80% found it useful.
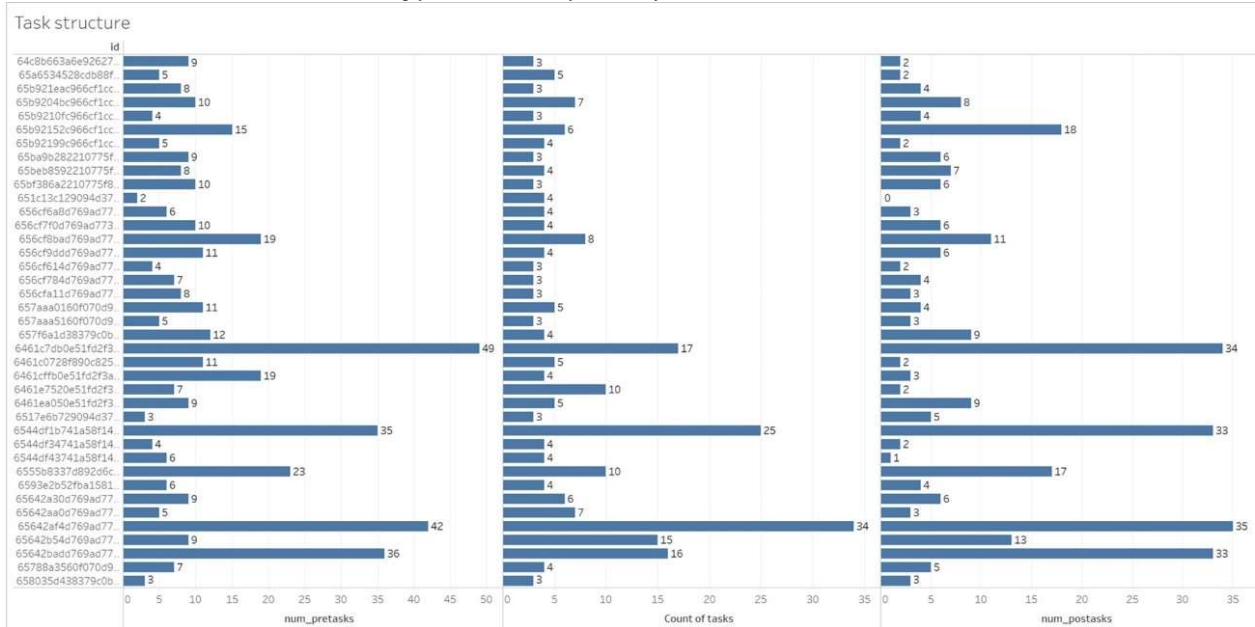
2.1 Quantitative research

In terms of quantitative research, in an exploratory study Trager aims at collecting information about the options selected by primary school teachers during task design. Trager will collect massive data about: the profile of the participants; their teaching experience; the time they spend creating tasks; the relative amount of time devoted to pre-tasks, tasks, and post-tasks; the use of and time on aids; the type of pre-task and post-task options they choose; the length of the texts they choose; and the kind of focus on form options they employ in order to draw learners' attention to language (e.g. did they make use of input enhancement? on what forms and how many words?). This is made possible thanks to the use of timestamps (start/finish time) for each activity which is then collected by means of *Tableau* software for data analytics (https://www.tableau.com). *Tableau* allows for the collection, analysis, and presentation of data from the *taskGen* database. *Tableau* lets users cross examine any of the variables and labels included in the web site's infrastructure. In the following example (Figure 9 below) information was extracted about the number of pre-tasks, tasks, and post-tasks created by individual users of the tool. This can potentially show different patterns of behavior in different contexts and from different perspectives.

TEISEL
Tecnologías para la investigación en segundas lenguas

UNIVERSITAT DE
BARCELONA
Facultat de Filologia
i Comunicació

Figure 9

Number of pre-task, tasks and post-tasks by users as extracted with Tableau.

Data can be grouped by any of the categories presented in Section 3. In the example below (see Figure 10), we extracted the number of tasks designed so far with our tool from different teaching frameworks, which revealed that most people have designed tasks from a TBLT perspective (a bias of our research, by which we have contacted TBLT programs first), with tasks designed in a Content and Language Integrated Learning (CLIL) context following, and then from a traditional teaching perspective, and from a project-based one, with only a few having been designed in Language for Specific Purposes (LSP) contexts.

Figure 10

Example of the number of tasks by teaching perspective: TBLT, CLIL, PBL, traditional teaching, and LSP.
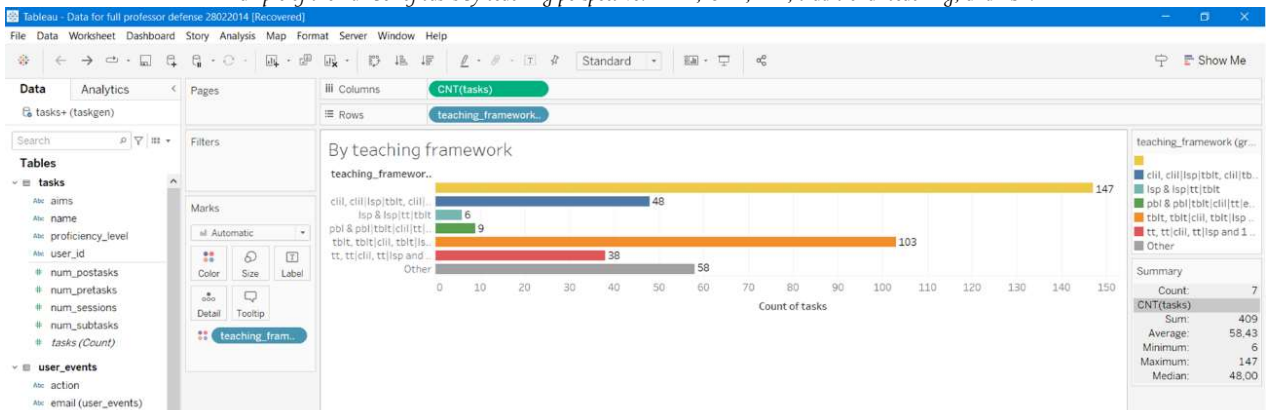
Data can be exported to Excel and SPSS or any other statistical packages in order to, for example, establish the proportion of pre-tasks and post-tasks to main tasks. Teachers at different educational levels may be approaching

task design in different ways, and Tableau can help us research the emerging patterns from the data. Data can be also classified and contrasted by country, education approach, years of experience, or any other variable included in the system (see section 1.3 above).

Motivated by the overall research question of how SLA concept in general and task design factors in particular transfer to actual task design, in the *taskGen* project we are interested in exploring the emerging patterns of task design in relation to profile data (see table 1.). By means of *Tableau*, we can do so by classifying data by any profile categories in order to capture any emerging and contrasting patterns of behavior among different groups.

Table 1
Research questions and SLA concepts motivating our quantitative research into task design in relation to profile variables

SLA and task design concepts	Profile variables
- choice of task goals and task types - task structures (how many pre-tasks, tasks, post-tasks) - interactional variables (split vs shared, open vs closed, convergent vs divergent) - input types (text, video, audio, multimodal…), and their length - oral and written modes - choice (or lack thereof) of linguistic focus (what dimensions? In relation to what goals?) - focus on form techniques (input enhancement by highlighting, bolding, font size, underlining) - their level of engagement with task design (number of tasks, total design time for each task, design time per section, total number of returns to the tasks; use of aids and training, number of aids consulted, total time of aid use)	- levels of education (primary, secondary, university, private lessons, language schools…). - teaching modalities (face-to-face, online, hybrid…) - teaching perspectives (TBLT, CLIL, Project-based learning, traditional teaching, language for specific purposes, language for academic purposes, special needs, EMI. - years of experience - levels of training in TBLT? - proficiency levels of students - countries or cities

Additionally, we plan on calculating any returns by teachers to tasks and task sections, which would tap into design and re-design practices. To our knowledge, no previous studies have had this privileged view at how task design actually happens.

3. Future research

If further funding is granted by the state agency, other intervention studies in schools are also planned for the second phase of the project that will let us go into schools and work with teachers, and have observations and follow ups of the tool use. Such data will allow the triangulation of qualitative and quantitative methods that will give us a much more complete picture of task design among L2 teachers.

Additionally, we plan to shift our focus to the learners with four intervention studies. Study 1 will look at the effects of *taskGen* on primary school students' development of English oral language and literacy abilities from grades 3 to 4 in CLIL & EFL learning contexts. Research questions include how prior teacher training affects CLIL vs. EFL teachers' methodology and students' development of English oral language and literacy abilities over a two-year period (grades 3 and 4). A mixed-method approach with both qualitative tools (interviews and surveys) and a quantitative analysis of Complexity, Accuracy and Fluency measures to see student gains in CLIL vs. EFL classrooms will be used. Correlations between teacher training variables, student test scores and the use of mixed-effects models is part of the design. Study 2 will take place in a primary school educational context, and it will look at the effects of *taskGen* on pragmatic aspects on children's language development and teachers' perceptions. Learner's pragmatic development will be measured in terms of exposure and use of speech acts (requests/suggestions). Teachers' perceptions and practices will be captured through case studies and classroom observations. Study 3 will be conducted in a secondary school context, and will look into the effects of task design on pragmatic development. In interlanguage pragmatics we have seen that even if all studies are saying that 1) teaching pragmatics is beneficial for learners, and 2) that tasks enhance pragmatic learning, pragmatics is still largely at the periphery of the language curriculum. The most recent studies have shown that one of the main problems is teachers' insufficient knowledge on pragmatics and a lack of teaching resources. With *taskGen* the second problem can be addressed in a three-step fashion: training teachers on how to use the tool for the purpose of pragmatics, have them design tasks, and implement and assess their impact on pragmatic development. Study 4, in a tertiary content, will investigate the use of *taskGen* by TBLT and CLIL teachers and students in a fully asynchronous online teaching context. This extends the use of the tool to a cohort of teaching professionals who will use it in an exclusively online authentic teaching context. The main objectives that lead this investigation are (1) to implement and integrate the tool with full instructions in an asynchronous context, (2) to track the patterns of use of the tool in this population by using data analytics, (3) to gain teachers and students' perceptions of the use of the tool by means of self-assessment, recall and interviews, (4) to compare the tasks designed with the tool with those designed in other, traditional ways.

Through both the use of NLP resources and Large Language Models (LLMs) we plan on developing options in relation to input elaboration, a focus-on-form technique that facilitates understanding of difficult words in text without eliminating them. Another function will be the automatic detection of text levels (A1-C2), as well as the simplification and complexification of texts. Different types of learners require different approaches to language instruction. In this regard, the tool can help explore the linguistic aspects used to create materials addressed to different learning profiles, which opens the door to adaptation and personalization.

Finally, in the future we are planning on expanding the number of languages covered by the NLP module to Catalan, Spanish and Russian. This will definitely allow contrastive studies, such as comparing the linguistic aspects used by teachers at different levels and in different languages.

Conclusions

In this paper we have presented an example of how technology may mediate the transfer from SLA theories to actual task design but L2 teacher. More especially, we have presented a web-based tool that assists and trains L2 teachers in task design by following our accrued knowledge of three decades of TBLT. We have described the main features of the tool and its possibilities for the creation, storage, and management of task designs. We have shown its additional functionalities for focusing form and highlighting linguistic features, for training teachers as they design their tasks, and for sharing tasks with other teachers and with learners. We have also zoomed in on the natural language processing capabilities, and their contribution to focus on form through automatic text analysis and application of focus on form techniques (input enhancement). Later, we have highlighted its enormous potential to investigate teachers' choices during task design through both qualitative and quantitative research by methods such as semi-structured interviews or think aloud protocols that tap into the thought processes that teachers engage during task design. The outcomes of such research have had an enormous impact on the design of our tool and will keep informing us about teachers' choices and patterns of task design.

We should at this point emphasise the fact that a number of lessons we had not predicted have emerged from users of the tool and participants in our studies so far. These include its potential for task design co-construction, for personalization and adaptation, for community building, or its training dimension. Firstly, teachers have the potential to engage into co-constructed designs by which two or more teachers share a task design back and forth until they agree upon the final version of a task. While most people current users seem to be designing tasks individually, we have discovered the potential for co-construction that we had not contemplated originally. Secondly, the cloning function in the dashboard facilitates teachers' generation of slightly different versions of the same task that may be used within the same group or across groups. This has already caught the attention of teachers and we wonder how much personalization and integration of individual differences and special needs will take place. Thirdly, the tool has already shown its potential for community building. This was detected during our qualitative analysis and with data analytics we will be able to see how much sharing takes place among teachers using *taskGen*. In the fourth place, teacher trainers have stood out because of their interest in using the *taskGen* tool for training purposes since it contains design tips that connect design options to the existing literature on task design.

For the second phase of the *taskGen* project, which is dependent on state funding, we plan on adding new NLP resources and artificial intelligence features for the automatic detection of the levels of texts (from A1 to C2), and for

the simplification or complexification of texts. This latter functionality should help teachers reduce or increase the difficulty of texts by automatically increasing/decreasing word frequency, and by breaking down or complexifying sentence structures. We have planned to adapt the tool and its NLP module to Spanish, Catalan, and Russian, hence showing the impact of the tool on the design of morphologically simple languages (English), with more morphologically complex ones (Spanish and Catalan), and highly morphologically complex languages (Russian). Ultimately, we hope to continue assisting L2 teachers with their task design, exploring the enormous possibilities as a window into actual task design by teachers, as well as its potential for research purpose, and to increase our understanding of L2 task design by teachers both locally and universally.

Acknowledgements

References

Alva-Manchego, Fernando; Scarton, Carolina; Specia, Lucia (2020). Data-Driven Sentence Simplification: Survey and Benchmark. *Computational Linguistics, 46*(1),135-187. https://aclanthology.org/2020.cl-1.4.pdf

Arnold, Gina (2023). Teacher perceptions on task design & technological integration in lesson planning for the EFL/ESL classroom. Unpublished MA thesis at the University of Barcelona.

Bibauw, Serge; François, Thomas; Desmet, Piet (2022a). Dialogue systems for language learning: Chatbots and beyond. In N. Ziegler & M. González-Lloret (Eds.), *The Routledge Handbook of Second Language Acquisition and Technology* (pp. 121-134). Routledge. https://doi.org/10.4324/9781351117586-12

Bibauw, Serge; Van den Noortgate, Wim; François, Thomas; Desmet, Piet (2022b). Dialogue systems for language learning: A meta-analysis. *Language Learning & Technology, 26(1)*, 1-24. https://hdl.handle.net/10125/73488

Divekar, Rahul; Drozdal, Jaimie; Chabot, Samuel; Zhou, Yalun; Su, Hui; Chen, Yue; Zhu, Houming; Hendler, James; Braasch, Jonas (2022). Foreign language acquisition via artificial intelligence and extended reality: design and evaluation, *Computer Assisted Language Learning*, *35*(9), 2332-2360. https://doi.org/10.1080/09588221.2021.1879162

Ellis, Rod (1997). SLA and Language Pedagogy: An Educational Perspective. *Studies in Second Language Acquisition, 19*(1), 69-92. https://www.jstor.org/stable/44488668

Ellis, Rod; Skehan, Peter; Li, Shaofeng; Shintani, Natsuko; Lambert, Craig (2019). *Task-based language teaching: Theory and practice.* Cambridge University Press. https://doi.org/10.1017/9781108643689

Filighera, Anna; Steuer, Tim; Rensing, Christopher (2019). Automatic Text Difficulty Estimation Using Embeddings and Neural Networks. In Scheffel, M; Broisin, J; Pammer-Schindler, V; Ioannou, A; Schneider, J. (Eds) *Transforming Learning with Meaningful Technologies*. EC-TEL 2019. *Lecture Notes in Computer Science,* 11722. Springer International Publishing. https://doi.org/10.1007/978-3-030-29736-7_25

Grabar, Natalia; Saggion, Horacio (2022). Evaluation of Automatic Text Simplification: Where are we now, where should we go from here. In *Actes de la 29e Conférence sur le Traitement Automatique des Langues Naturelles. Vol. 1: conférence principale* (pp. 453-463). ATALA. https://aclanthology.org/2022.jeptalnrecital-taln.47/

Gregg, Kevin R. (2001) Learnability and SLA theory. In P. Robinson (Ed.) *Cognition and Second Language Instruction*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524780

Gilabert, Roger (2023). Where SLA and Language Education Meet: The Transfer from Task-Based Needs Analysis to Task Design. *Education Sciences*, *13*(10), 1015. https://doi.org/10.3390/educsci13101015

Guo, Kai; Wang, Deliang (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-12146-0

Kang, Joohoon; Yi, Youngjoo (2023). Beyond ChatGPT: Multimodal generative AI for L2 writers. *Journal of Second Language Writing, 62*. https://doi.org/10.1016/j.jslw.2023.101070

Koltovskaia, Svetlana (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing, 44*, 100450. https://doi.org/10.1016/j.asw.2020.100450

Lee, Sangmin-Michelle (2020). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*, *33*(3), 157-175. https://doi.org/10.1080/09588221.2018.1553186

Lee, You-Jean (2021). Still taboo? Using machine translation for low-level EFL writers. *ELT Journal*, *75*(4), 432-441. https://doi.org/10.1093/elt/ccab018

Long, Mike H. (2015). *Second Language Acquisition and Task-Based Language Teaching*. Wiley.

MacDonald, Malcolm; Badger, Richard; White, Goddith (2001). Changing values: what use are theories of language learning and teaching? *Teaching and Teacher Education*, *17*(8), 949-963. http://dx.doi.org/10.1016/S0742-051X(01)00042-7

Malicka, Aleksandra; Gilabert, Roger; Norris, John (2019). From needs analysis to task design: Insights from an English for specific purposes context Language Teaching Research. *Language Teaching Research*, *23*(1), 78-106. https://doi.org/10.1177/1362168817714278

Markee, Numa (1997). Second Language Acquisition Research: A Resource for Changing Teachers' Professional Cultures? *The Modern Language Journal. Special Issue: How Language Teaching is Constructed*, *81*(1)*,* 80-93. https://doi.org/10.1111/j.1540-4781.1997.tb01628.x

Martinc, Matej; Pollak, Senja; Robnik-Šikonja, Marco (2021). Supervised and Unsupervised Neural Approaches to Text Readability. *Computational Linguistics*, *47*(1),141-179. https://aclanthology.org/2021.cl-1.6.pdf

Nunan, David (1991). Methods in Second Language Classroom-Oriented Research: A Critical Review. *Studies in Second Language Acquisition, 13*(2)*, Methods in Second Language Research*, 249-274. https://www.jstor.org/stable/44488593

Qi, Peng; Zhang, Yuhao; Zhang, Yuhui; Bolton, Jason; Manning, Christopher (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In *Association for Computational Linguistics (ACL) System Demonstrations*. https://aclanthology.org/2020.acl-demos.14.pdf

Tecnologías para la investigación en segundas lenguas

Sato, Masatoshi; Loewen, Shawn (2019). Do teachers care about research? The research-pedagogy dialogue. *ELT Journal, 73*(1). https://doi.org/10.1093/elt/ccy048

Stapleton, Paul; Leung Ka Kin, Becky (2019). Assessing the accuracy and teachers' impressions of Google Translate: A study of primary L2 writers in Hong Kong, *English for Specific Purposes*, *56*, 18-34. https://doi.org/10.1016/j.esp.2019.07.001

Sydorenko, Tetyana (2015). The use of computer-delivered structured tasks in pragmatic instruction: An exploratory study. *Intercultural Pragmatics, 12(3)*, 333-362. https://doi.org/10.1515/ip-2015-0017

Wang, Elain; Matsumura, Lindsay. C; Correnti, Richard; Litman, Diane; Zhang, Haoran; Howe, Emily; Quintana, Rafael (2020). eRevis(ing): Students' revision of text evidence use in an automated writing evaluation system. *Assessing Writing, 44*, 100449. https://doi.org/10.1016/j.asw.2020.100449

Warschauer, Mark; Tseng, Waverly; Yim, Soobin; Webster, Thomas; Jacob, Sharin; Du, Qian; Tate, Tamara (2023). The affordances and contradictions of AI-generated text for writers of english as a second or foreign language. *Journal of Second Language Writing*, *62*. https://doi.org/10.1016/j.jslw.2023.101071

Zhai, Chunpeng; Wibowo, Santoso. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university, *Computers and Education: Artificial Intelligence*, *4*. https://doi.org/10.1016/j.caeai.2023.100134.