MulTINCo: Multilingual Traditional Immersion and Native Corpus. Better-documented multiliteracy practices for more refined SLA studies.

Abstract

Whilst the links between learner corpus research (LCR) and Second Language Acquisition (SLA) have long been debated, McEnery et al. (2019, 17) claim that the promise made by Granger in 2009 that learner corpora would "soon be accepted as a bona fide data type in SLA research" has yet to be realized. This article aims to go one way towards realizing this promise by illustrating the benefits of collecting and analyzing data sets that better document multiliteracy practices.

We first contextualize our work within the field of LCR where calls for more multidimensional data sets have been made. We then present a new database called MulTINCo – Multilingual Traditional, Immersion, and Native Corpus – collected within the context of a large-scale project on *Content and Language Integrated Learning* (CLIL) in French-speaking Belgium. As our data set contains rich metadata and blends corpus data with other data types, we explain how it can help bridge the LCR/SLA gap. In Sections 3 and 4, we describe the data collected and the interface. In the last section of the paper, we wrap up the discussion on the methodological assets of such multidimensional data sets for SLA studies, and present outlooks on future research.

Keywords: Learner Corpus Research; SLA; multi-literacy; L2 Dutch; L2 English; L1 French; CLIL.

NOTE: [data availability statement – deleted]

1. Introduction

Over the last decades, Corpus Linguistics has used "large quantities of observational data compiled into data sets, called corpora, to provide evidence about language use by both first language (L1) and second language (L2) speakers" (McEnery et al. 2019, 74). It has proven particularly helpful for disentangling frequency patterns in natural language use (Biber 1998; Biber and Conrad 2019), as well as for showing the importance of lexis/grammar/syntax co-selection patterns like collocations (Granger and Bestgen 2014; Brezina 2018), colligations (Hoey 2005) or constructions (Yoon and Gries 2016; Hendrikx et al. 2019).

The field of Learner Corpus Research (LCR) started developing rapidly in the early 1990, when Granger launched the *International Corpus of Learner English* (ICLE) project (Granger 1988; Granger et al. 2002; 2009). Building on Sinclair's (1996) definition of corpora, Granger defines a learner corpus as follows:

> Computer Learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardised and homogeneous way and documented as to their origin and provenance. (Granger 2002, 7)

According to Granger (2002, 9), "the usefulness of a learner corpus is directly proportional to the care that has been exerted in controlling and encoding the variables". She recommends collecting both variables concerning the learners (such as the learning context, the learners' first language (L1), other L2's that the learners know and the level of L2 proficiency) as well as variables concerning the task (such as the time limit, whether the students had access to reference tools, and the audience/interlocutor) (Granger 2002, 9). A growing awareness of these factors is reflected in the numerous learner corpora collected over the world and the increasing number of publications

detailing methodological decisions during corpus building (e.g., Bell, Collins and Marsden, in press; Lozano and Mendikoetxea 2013). As illustrated by the 'Learner Corpora around the World'[1] webpage, managed by the Centre for English Corpus Linguistics (CECL) at the Université catholique de Louvain, an increasingly large number of learner corpora are being collected and the attention paid to the inclusion of detailed metadata is central in current LCR projects and studies (see Granger et al. 2015 for numerous illustrations).

Another development in the field of corpus linguistics involves the widespread use of learner corpora across disciplines, as observed by Callies and Paquot (2015):

> Learner corpus data are now being used in various disciplines including SLA research, Language Testing and Assessment and Pedagogical Lexicography. They also constitute the raw material for a wide range of Natural Language Processing (NLP) based tasks (e.g. training of part-of-speech taggers and parsers, automatic error detection and correction, automatic exercise generation) and are used in tools as varied as essay scoring systems and intelligent Computer Assisted Language Learning (CALL) programs. (Callies and Paquot 2015, 2).

Despite such evolutions in LCR, however, McEnery et al. (2019, 76) still mention an apparent disconnect between learner corpus studies and SLA and find "no clear sign that the situation is changing".

In this contribution, we aim to present ways to foster collaboration between LCR and SLA by presenting an innovative data set that not only contains corpus data and rich metadata but also blends corpus data with other data types (see Section 2) that document multiliteracy practices. The term multiliteracy is used as an umbrella term that encompasses : biliteracy (defined as the ability to read and write proficiently in two

[1] https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html#Learner_corpus-based_datasets

languages), oracy (defined as proficiency in oral expression and comprehension), and situated practice, originally formulated by the New London Group (1996) as one of the related components of multiliteracies pedagogy (immersion in meaningful practices within a community of learners, taking account of the affective and sociocultural needs of learners and their out-of-school communities and discourses, as an integral part of the learning experience). The MulTINCo database includes various types of L2 data (spoken and written L2 productions, longitudinal data) collected from French-speaking learners of Dutch and English as second languages in different educational settings (Content and Language Integrated Learning (CLIL[2]) and typical L2 classes). We use the term Second Language (L2) to refer to both Dutch and English taught to French-speaking learners in Belgium. These languages have a different status within French-speaking Belgium. Second languages (SL or L2) are typically used within the local context of a learner, whereas Foreign Languages (FL) are languages that are not used in the local context, or which are not commonly spoken by the community as a whole. According to this definition, Dutch would be a second language, whereas English would be a foreign language in French-speaking Belgium. However, these definitions are somewhat controversial and do not always reflect the true linguistic situation at stake. The students from French-speaking Belgium in the present sample reported less contact with Dutch than with English, for instance (Van Mensel et al. 2019). Therefore, we opt for the most commonly used term Second Language (L2) to refer to the acquisition of both Dutch and English in French-speaking Belgium.

[2] See section 2.6 for contextual details on CLIL (number of hours, types of content courses taught in the target language, etc.).

2. Situating MulTINCo in the field of learner corpus research

Calls have been made in LCR for more multidimensional corpora and data sets. This section presents some of the key features or data types that current learner corpora should aim to include. We also show how MulTINCo has tried to collect and include those features/data.

2.1. The value of longitudinal corpora

As argued by Ortega and Byrnes (2008, 18), "longitudinal designs can uniquely help researchers document the lengthy trajectories of adults who strive to become multicompetent and multicultural language users". It thus seems important to pursue efforts in collecting them. Longitudinal learner corpora are laborious to collect, which probably explains why they are relatively scarce. As Gilquin (2015, 28) observes, "there are more corpora containing cross-sectional than longitudinal data" since "multiple data collection from the same learners requires heavier logistics than one-off data collection". Moreover, existing longitudinal L2 data "tends to document learner development over a small number of months/years for a small number of learners" (Myles 2015, 316).

Author (2015) lists a number of rather large longitudinal learner corpora that follow the same set of participants over multiple data-gathering sessions. They include five subcorpora of the FLLOC (*French Learner Language Oral Corpora*) project, some of the subcorpora of the InterFra (Interlangue Française) corpus, the longitudinal subcorpus of the *Corpus Ecrit de Français de Langue Etrangère* (CEFLE), the *Barcelona English Language Corpus* (BELC), some subcorpora of the *Corpus of Learner German* (CLEG13), the *Telecollaborative Learner Corpus of English and German* (Telekorp), and the LONGDALE (*Longitudinal Database of Learner English*)

project. For more details and references on the corpora listed above, their design, the targeted language studied and the data type covered, see Author (2015).

MulTINCo contains written longitudinal data produced by over 400 learners in two L2's analyzed (viz. Dutch or English). Two data collection points were organized, with a 20-month gap between them (October 2015 and May 2017). MulTINCo also contains data produced by the learners in French, their mother tongue. The need for collecting L1 data has indeed been voiced by Author (2015) who recommends not only integrating – in learner corpora – information related to proficiency in the learners' L1 to enable researchers to be much more specific in their analyses and interpretations of bi- and multiliteracy practices. It is also important to collect L1 data of a type similar to the L2 data collected (same text type for instance) to enable an integrated comparison of the learners' proficiency levels in their L1 and L2.

2.2. The need for collecting data on/from less researched languages

A great number of English as a Foreign Language (EFL) learner corpora have been collected over the last decades (see https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html#Learner_corpus-based_datasets for concrete examples). Learner corpora with other L2s than English are however much scarcer. In fact, we find "more corpora of English than any other language" (Gilquin 2015, 28) and more than 70 per cent of the corpora are L2 English (Granger, Gilquin, and Meunier 2013). To our knowledge, only two learner corpora/data sets include L2 Dutch: the *European Science Foundation Second Language Database* (ESF database) (Klein and Perdue 1997, Perdue 1993) and the *Leerdercorpus Nederlands als Vreemde Taal* (Degand and Perrez 2004). This makes MulTINCo one of the few multilingual learner corpora including both a well-researched L2 (English) and a less researched one (Dutch). In addition to the importance of

including a less researched language, the impact of this inclusion in our research context (Belgium) is particularly important. These two L2s have a very different status in Belgium when it comes to language attitudes and motivation, and as our project also included socio-affective research objectives (see De Smet et al. in this special issue), it was central for the team to be able to work with two L2s with different statuses.

2.3. The plea for written and spoken data

More than 80 per cent of the existing learner corpora are written (Granger et al. 2013) and one of the main reasons for this underrepresentation of spoken corpora is the laborious task involved in transcribing spoken data (Gilquin 2015). However, spoken and written corpora both contribute to the analysis of learners' interlanguage in the sense that written data are well suited for the analysis of the learners' explicit knowledge (as learners usually have more time for processing, producing and potentially revising/editing the data) and spoken data for investigating the internal learner processes or implicit knowledge (as online processing leaves less time for adjustments in the production) (Myles 2015). Not only is more spoken data needed, but the audio files that go along with the transcriptions are also rarely provided, whilst these are necessary for phonetic research (Ballier and Martin 2015). Also, oral corpora should include more spoken data resulting from "natural speech as opposed to read speech" (Ballier and Martin 2015, 129).

MulTINCo contains both written and spoken data from the same learners. The spoken data was collected through a conversation exercise in pairs, which is closer to natural speech (see section 3.2. for the detailed description of the tasks). The audio files are also provided along with the transcripts.

2.4. The need for comparable learner and native corpora

Work on learner data makes it possible to identify features of interlanguage. The comparison with native data also allows researchers to assess the extent of the differences between target norms (Granger 2002), and – if it is part of their agenda – propose pedagogical applications to help learners reach the target that they are aiming at.

Still, it is crucial that the native speakers ideally perform the same task as the learners and be of similar age ranges (Myles 2015). Collecting similar types of productions from native speakers make the comparison stronger and, in case on native/non-native comparisons, offers a more realistic target (same text types in similar conditions).

MulTINCo includes learner data and native-speakers' data to serve as a benchmark of comparison. Both native and learner written data were collected at the same time (2015-2017) using the exact same task/prompt. All participants (learners and controls) were approximately the same age; the native speakers of English were slightly older.

It should also be added that, as argued by Littré (2014), statistical codes, task prompts or coding systems should be made available in order to favor replication studies and enhance the expertise in analyzing longitudinal data. In that respect, initiatives such as IRIS (Marsden et al. 2016) and its repository of instruments for research into L2s should be (re)commended. MulTINCo clearly documents all the tasks that were done by the learners (see also annexes in the present article for illustration).

2.5. Dearth of data from lower proficiency levels and need for more variety in registers/text types

The participants in learner corpora are often advanced learners, and mostly university students (Gilquin 2015). They are often composed of argumentative essays produced by students and collected during classroom tasks or exams. Yet, as argued by McEnery at al. (2019, 81) "users of a language are required to produce writing in many more registers than this. Short notes, emails, and positive and negative reviews are all forms of writing that we may suppose these learners of English may need to acquire. But these forms are absent from the existing learner corpora. So, there is a clear potential disjunction between the data current learner corpora provide and the research questions that SLA researchers may wish to address".

The learners in MulTINCo are younger (upper secondary school, mean age of 16.5 at the moment of the first data collection), which allows us to assess language proficiency in earlier stages of L2 learning (see Section 3.1) and the data type collected are emails and rather informal pair interactions.

2.6. The importance of comparing various instructed settings in second language teaching

Another valuable aspect of MulTINCo is its focus on two different educational settings, namely Content and Language Integrated Learning (CLIL) and typical second language learning classes where the L2 is the object of study (non-CLIL). CLIL is "an educational approach where curricular content is taught through the medium of a foreign language" (Dalton-Puffer 2011, 183). While Belgian secondary school learners in L2 learning classes are taught their L2 about 4 hours a week, CLIL students attend classes in their L2 between 8 and 13 hours per week (including the L2 class) (Hiligsmann et al. 2017). The database thus makes it possible to investigate the effect of the amount of L2 exposure on L2 proficiency in CLIL and non-CLIL learners. As such, MulTINCo contributes to

"[E]mpirical research into interrelations between learning contexts and language learners' performance [which] so far has mainly focused on a small number of learners and/or a very restricted set of context-related variables" (Mukherjee and Götz 2015, 439).

2.7. Rich metadata

According to Ädel (2015, 419), "learner corpus researchers need to integrate metadata to a much greater extent than has been the case to date". Examples of improvements include the inclusion of an "objective proficiency score" (Gilquin 2015, 30) and "information […] about incidental learning in everyday life through reading, entertainment, social networking, etc.", which would extend our understanding of the learners' amount of L2 input variable (Gilquin 2015, 30-31).

Besides the learners' L2 - and their L1 productions – and native corpora, MulTINCo also integrates numerous background variables in its database. It contains internal and external variables (native language, age, gender, non-verbal intelligence (Raven score), socio-economic status, curricular and extracurricular input, educational background, task description, etc.). These are discussed in more detail in Section 4.

To sum up this section, we believe that MulTINCo fills a gap in LCR as it answers the call for more multimodal corpora and data sets.

3. Participants and data collection

3.1 Participants

The database consists of e-mails written by 438 French-speaking learners of Dutch and English from nine secondary schools in Wallonia (French-speaking Belgium). The schools are diversified with regard to location (schools from all Walloon provinces),

socio-economic status[3], and educational network (official subsidized education, free subsidized education) (see also Hiligsmann et al. 2017, Van Mensel et al. 2019 for more details). The schools offer CLIL in Dutch and/or English, alongside the typical instruction mode in Wallonia (viz. French-medium instruction for all content courses and L2 classes taught in the target language). At the time of the first data collection, the learner participants' ages ranged from 15 to 18 and their mean age was 16.5. 207 (46.7%) of the learners are male and 231 (53.3%) are female.

Data from the control group of Dutch native speakers was gathered in the Netherlands and in Flanders (the Dutch-speaking community of Belgium) from 61 Dutch-speaking fifth grade secondary school students (average age: 16.7). 11 (18%) of the speakers are male and 50 (82%) are female. Data from the control group of English native speakers was gathered from 70 English-speaking American[4] students (average age: 19.4) in Florida in 2016. The slight age difference between the learner and the control group is a result of practical difficulties: collecting data from L1 English speakers under the legal age of maturity in the United Kingdom and in the United States turned out to be rather complex in terms of legal restrictions; therefore the control group of L1 English speakers consists of the youngest possible adults to whom we could have access, viz. freshmen at university. 11 (17%) of the native controls are male and 54 (83%) are female.

Table 1 summarizes the distribution of the participants according to their L2 (English/Dutch) and educational setting (CLIL/non-CLIL).

---

[3] This index, approved by the Government of the French community, is defined on the basis of an inter-university study which updates the calculation formula every five years. The five types of main variables used are: income per capita, level of education, unemployment rate, professional activities and housing comfort.

[4] Whilst British English is probably the most common variety in terms of instruction (textbooks), American English is often represented both in curricular and extra-curricular activities.

Table 1. Distribution of the participants according to the L2 and educational setting

	CLIL learners	Non-CLIL learners	Controls
English	96	97	70
Dutch	133	112	61
TOTAL	229	209	131

3.2 Data collection procedure and instruments

A first data collection (T1) took place in 2015 at the Université catholique de Louvain (in Louvain-la-Neuve, Belgium) when the learners were in their 5th year of secondary school (Grade 11). They performed a variety of computer-administered tasks, including two writing exercises. This was replicated in 2017 (T2), when they were in their 6th year of secondary school (Grade 12). The spoken data was collected in 2017 (a few months before T2) in three secondary schools from our sample (n=61).

The writing exercises consisted in the production of informal e-mails to a friend (min. 15 lines, in Lime Survey 200[5]) about everyday topics – either their last holidays or a party they attended (randomly assigned topics). The first e-mail had to be written in the learners' L2 (Dutch or English) and the second in their first language (French). Having the students write in their L1 first would have had an effect on their L2 writing as the topic would have been activated (which would have meant the L2 language use captured would not have been as spontaneous). They had maximum 25 minutes time to write each e-mail and had no access to online dictionaries or other reference tools. The

[5] *Lime Survey:* https://www.limesurvey.org/about-limesurvey/team-contributors

spell check facility on the computers used for the data collection was turned off. See appendices A to F for the guidelines that were given to the learners. The native speakers followed the same procedure as the learners but wrote only one e-mail in their first language[6]. Tables 2a and 2b provide the number of texts and number of words per (written) sub-corpora.

Table 2a: Number of texts and number of words per sub-corpus – L1 French and L1 and L2 English

	L1 French		L2 English Non-CLIL		L2 English CLIL		L1 English (controls)
	T1-2015	T2-2017	T1-2015	T2-2017	T1-2015	T2-2017	
Texts	176	166	90	77	90	86	57
Words	60,148	45,067	23,747	19,985	29,394	26,855	19,134
Average text length (in words)	342	271	264	260	327	312	336

Table 2b: Number of texts and number of words per sub-corpus - L1 French and L1 and L2 Dutch

	L1 French	L2 Dutch Non-CLIL	L2 Dutch CLIL	L1 Dutch (controls)

[6] The only difference is that the written data from the English native speakers was collected using the online software Qualtrics – the American ethical commission did not accept Lime Survey.

	T1-2015	T2-2017	T1-2015	T2-2017	T1-2015	T2-2017	
Texts	255	213	100	90	132	132	61
Words	80,237	64,535	19,399	18,648	37,209	35,712	13,791
Average text length (in words)	315	303	194	207	282	271	226

The oral task consisted in a conversation exercise in pair about the two same topics as for the written exercise (holidays or party). The task was first performed in the learners' L2, and then in French. See appendices G and H for the guidelines that were provided to the learners. Table 3 shows the number of recorded conversations.

Table 3: Number of conversations per sub-corpus – L2 English, L2 Dutch and L1 French

	TL English CLIL	TL English NON-CLIL	TL Dutch CLIL	TL Dutch NON-CLIL
Nb of productions	17	11	13	20
Nb of words	4,173	2,153	3,854	3,644
Mean length of production in seconds	245	196	296	182

The tasks were chosen to gather the most authentic natural written and spoken language possible, while maintaining a similar topic. In as far as essay writing is an authentic classroom activity, learner corpora of essay writing can be considered to be

authentic written data, and similarly a conversation in pairs can be considered to be authentic spoken data in a classroom context.

The total number of words (all sub-corpora combined) is 493,861. This means that the corpus size is rather small compared to for instance the *Corpus of Contemporary American English* (COCA) (560 million words), the *Corpus Hedendaags Nederlands* (more than 43 million words) or the *Frantext* (more than 271 million words). However, *MulTINCo* was not conceived to be representative for present day English, Dutch or French, but it was built specifically to observe different aspects of the language acquisition process, track the language development between and within individuals, allow for comparisons between learner and native data and examine the potential influence of different variables related to the language learning setting, the learner or the task.

4. The MulTINCo interface and tools

The interface is divided into four main windows: *Home*, *Corpus selection*, *Corpus download*, and *Concordances*. A demo version of the interface with a sample of the data collected is available at: http://corpora.uclouvain.be/multinco_demo[7]. The data set and the link with multiple data types is much larger than what is visible in the demo version which only offers a first glimpse into MulTINCo.

The home page (*Home*) provides basic information on the corpus, references to articles published within the CLIL project and updates on the availability of new versions.

[7] In the demo version transcriptions of the spoken data are available, but the audio files are not included. Readers interested in accessing the demo version should contact *author's email* to obtain a username and password.

The *Corpus selection* window enables users to filter the texts according to a number of variables. The 'Student' related variables include:

- *Gender*
- *Place of Birth*
- *Age*
- *Degree of bilingualism* (only French spoken outside school, sometimes another language spoken outside school, mostly another language spoken outside school)
- *Language(s) spoken at home*
- *Current informal contact with target language:* composite measure consisting of frequency of Internet use in the L2, frequency of productive and receptive L2 use outside school and frequency of contact with L1 speakers outside school (Cronbach's alpha .78)
- *Non-verbal intelligence*: score obtained by the learner on the Raven's Progressive Matrices test (Raven, Court and Raven 1998).
- *Level of education of the mother*[8]
- *Receptive vocabulary knowledge in L1 French*: score obtained from the learner on a standardized test of receptive vocabulary knowledge: EVIP[9]
- *Receptive vocabulary knowledge in L2 Dutch*: score obtained from the learner on a standardized test of receptive vocabulary knowledge: PPVT-III-NL[10]
- *Receptive vocabulary knowledge in L2 English*: score obtained from the learner on a standardized test of receptive vocabulary knowledge: PPVT-IV[11]
- *Knowledge of content subject (French)*: score obtained from the learner on the standardized achievement test for French as a content subject
- *Knowledge of content subject (history):* score obtained from the learner on the standardized achievement test for history as a content subject
- *Student ID*: to allow text filtering on the basis of the learners' individual (anonymized) identification code

The 'Instructed settings' variables include:

- *School:* represented by a number for the sake of anonymity
- *Class:* also represented by a number
- *Type of education:* CLIL or non-CLIL

---

[8] This may seem discriminating but the mother's level of education is very frequently used as a proxy for SES because it has a low rate of missing responses and is highly correlated with father's education (Entwisle and Astone 1994).

[9] Dunn, Dunn & Thériault-Whalen (1993).

[10] Dunn, Dunn & Schlichting (2005).

[11] Dunn & Dunn (2007).

- *CLIL choice* (sub-filter if "CLIL" selected before): corresponds to the person (the learner or the parents) who decided to opt for the CLIL track
- *CLIL since…* (sub-filter if "CLIL" selected before): number of years the learner has spent in CLIL
- *Number of weekly teaching hours provided in Dutch :* choice between four different options (0h, 2h, 4h, more than 4h per week)
- *Number of weekly teaching hours provided in English :* choice between four different options (0h, 2h, 4h, more than 4h per week)
- *Number of years spent learning English at school* (be in in CLIL or non-CLIL)
- *Number of years spent learning Dutch at school* (be it in CLIL or non-CLIL)
- *Number of school years repeated by the learner*: never, once, twice, three times or more

The 'task' related variables include:

- *Language of production*: English, Dutch or French
- *Topic*: party or holidays
- *Data collection points*: T1(2015) and T2 (2017)
- *Text ID*: option used to type in the code of one specific text in order to access this text directly

An additional 'Manual selection' window offers the opportunity to fine-tune the selection by visualizing the metadata and, potentially, deselecting one or several texts.
Once the selection of the corpus has been done on the basis of all the variables, the texts can be downloaded (both as single files and as one aggregated file) thanks to the 'Corpus download' window. The downloaded data can be processed by any software tool.

The concordance function ('Concordances') allows searching of the entire corpus or the selected texts/transcriptions for concordances (e.g. only the texts produced by female non-CLIL learners in Dutch at data collection time T1). This means that a word form, lemma and/or a part-of-speech tag is shown in its immediate context. If one is interested in a particular concordance line, it is possible – by clicking on the small arrow on the left – to gain access to a 'concordance details' page, which summarizes all the

variables related to the text in which the specific entry was found, but also to the entire text and part-of-speech information (based on the TreeTagger, see https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/).

5. Methodological assets of new generation learner corpora: Better documented multi-literacy practices for more refined SLA studies

Over the last decades, researchers have developed several models that aim to analyze transfer effects in interlanguage. Jarvis (2000, 245) claims that "a unified framework" could greatly clarify the functioning of transfer and cross-linguistic influence on interlanguage. From the 1990s onwards, various attempts have been made to capture the impact of cross-linguistic similarities and differences in models, such as Granger's (1996) *Integrated Contrastive Model* – which was revised in Granger (2015); Jarvis's (2000) model; Gilquin's (2008) Detecting-Explanation-Evaluation (DEE) transfer model; and Hiligsmann and Rasier's (2011) *Integrated Contrastive Research Design*. All these models aim to investigate transfer by using a combination of Contrastive Analysis (CA) (comparing different languages) and Contrastive Interlanguage Analysis (CIA) (comparing different language varieties, among which the L2 produced by learners). The specific comparisons included in the CA and in the CIA components differ from model to model according to their specific aims but the models coincide in assessing the cross-linguistic influence on interlanguage by comparing (i) the learners' L1 and the target language; (ii) the learners' L1 and their interlanguage (L2); (iii) the target language and the learners' interlanguage; and (iv) the interlanguage of learners with different L1's.

MulTINCo allows us to conduct a type of CIA including the learners' L1 versus targeted native-language comparisons, two different interlanguages of learners (L2 Dutch and English) with the same L1. It also allows for a comparison of different production modes (written and spoken data produced by the same learners). Finally, the variables

included in the corpus make it possible to assess their specific impact on L2 learning. Paquot et al. (accepted) explain that having access to numerous variables is a must for multi-factorial analyses. The inclusion of predictors such as proficiency level, essay prompt/topic, various text/linguistic variables make it possible to better explain degrees of variance in a dataset by correlating certain predictors.

All in all, MulTINCo allows for multiple types of investigations. For example:

- the comparison of L1-speakers' productions (French, Dutch, English) to identify cross-linguistic differences and possible difficulties for the learners;
- the comparison of learners' productions (L2 Dutch/English) to L1 Dutch/English productions (from the control group) to assess the typical properties of the learner language (in terms of variation, collocational preferences, etc.) in comparison to the L1's;
- the comparison of learners' L1 (French) and their L2 productions to help us identify possible transfer effects from their L1;
- the comparison of the learners' productions over time to identify developments after 1,5 years of more L2 instruction;
- the comparison of the learners' L2 productions in CLIL and in Non-CLIL contexts;
- the comparison of the results of the CLIL group in the two L2 conditions, to check whether CLIL may have different outcomes in different L2 contexts.

While we have a rich data set, MulTINCo also has its limitations and does not contain all the data types needed to conduct a study according to the CIA models presented earlier. However, the database enables us to conduct studies following a research design that is original as it takes into account two different L2s, two types of educational settings, a

longitudinal dimension, spoken and written data from the same learners, alongside a rich set of control and background variables.

It is important to note that the majority of studies using CIA largely focus on the learners' L1 as a variable and tend to overestimate its effects (Callies 2015), which is problematic as L2 acquisition is affected by other variables than the L1, and its acquisition follows an underlying developmental trajectory. However, we believe that "LCR has a lot to gain from embracing a variationist perspective, taking full advantage of the recorded metadata that pertain to both learner and task variables" (Callies 2015, 49). In fact, other important variables should be taken into account, besides the learners' L1, when comparing learners' with native-speakers' productions, as explained by Paquot et al. (2019) in their plea for the inclusion of a large number of predictors in future LCR/SLA studies. Learner variability has been somewhat neglected in the methodological practices that have been used so far in LCR, hence the importance of integrating metadata to a greater extent (Ädel 2015). The learner variables made available in the database are numerous; they involve the learners' main characteristics (age, gender), their linguistic background (L1, home language, years of formal instruction, frequency of informal extracurricular input, etc.), information about their cognition (non-verbal intelligence), and their educational background (home situation, socio-economic status, etc.). The inclusion of these variables will hopefully allow for more learner corpus research looking at other factors than the mother tongue.

Corpus comparability issues are also often an issue in LCR as the available reference corpora are larger than the learner corpora and they also differ in text type or task setting (see for example Breeze 2007 on the influence of task variables; Granger and Paquot 2009, and Gentil and Meunier 2018 on register and genre differences). Controlling these variables is essential for the validity of the results. Our learner and native-speakers'

data are relatively comparable since they all performed the exact same writing task provided with the same task instructions and in similar task settings.

In this contribution, we have presented a new data set and interface which integrates various L1 and L2 data types, together with rich metadata that well-documented multi-literacy practices. We hope that MulTINCo – and similar data sets – will help overcome one of the limitations mentioned by McEnery et al. (2019, 80) that there "are currently few well-structured data sets that permit the study of variables relevant to SLA, such as proficiency level, sociolinguistic variables (e.g., age, educational attainment, social class), contextual features (including different registers and modes of communication), and L1 background". Blended data sets like ours make it possible to carry out more refined SLA studies and come up with more fine-grained interpretation of the results.

The MulTINCo database has already been fruitfully exploited in several studies. Bulon et al. (2017) look at the global written proficiency of CLIL and Non-CLIL learners in the two L2's (English/Dutch) and in their L1 (French) using a number of lexical and syntactic measures. In a follow-up study, Van Mensel et al. (accepted) examine the influence of different types of input (CLIL, formal instruction and informal contact) on the learners' variability in writing. Bulon (forthcoming) compares the CLIL and non-CLIL learners' written phraseological language in L2 English in terms of frequency, variety and accuracy. Other analyses related to phraseology involve a comparison of the learners' phraseological language in the two communications modes (written versus spoken), as well as an investigation of their written phraseological development over time (see Bulon and Meunier in this special issue and Bulon 2019). Hendrikx, Van Goethem and Wulff (2019) and Hendrikx (2019) look at the cross-linguistic influence and

(longitudinal) impact of CLIL on the acquisition of intensifying constructions using the written data collected from the control groups (native speakers) and the learners (L1 and L2 Dutch or English). Studies carried out by university students in the framework of their Master's dissertation or seminar papers involved comparisons of the use of word order (Dumont 2018), comparative constructions (Broisson and Van Goethem 2018) and intensification in speech (Buntinx and Van Goethem 2018) in the productions of CLIL and non-CLIL learners in particular subsets of the database.

For the time being, as the project is still going on, the corpus is only accessible to the members of the research team and their collaborators. In the future, we aim to extend access to the data set to the research community.

We acknowledge that one of the limitations of MulTINCo is that it is relatively small compared to other existing corpora that contain millions of words. However, MulTINCo stands out from these big corpora by its variety in the data type and the metadata available. A second limitation concerns the longitudinal learner data: only two waves of data collection (T1 and T2) were carried out, which may be not ideal to track linguistic development in detail. A third limitation, inherent to all corpora – even very large ones – is that researcher interested in low frequency features will need to use different data types and methods to answer their questions.

These limitations notwithstanding, we sincerely believe that MulTINCo – through its rich and blended data set - is a highly valuable database for current and future LCR and SLA research; it provides useful data types and numerous variables that allow for a better and methodologically stronger investigation of second/multilingual acquisition and practices.

Acknowledgements

<deleted>

References

Author

Authors

Ädel, A. 2015. "Variability in Learner Corpora". In *The Cambridge Handbook of Learner Corpus Research*, edited by S. Granger, G. Gilquin, and F. Meunier, 401-421. Cambridge: Cambridge University Press.

Ballier, N. and P. Martin. 2015. "Speech Annotation of Learner Corpora". In *The Cambridge Handbook of Learner Corpus Research*, edited by S. Granger, G. Gilquin, and F. Meunier, 107-134. Cambridge: Cambridge University Press.

Bell, P., L. Collins, and E. Marsden. In press. "Building an Oral and Written Learner Corpus of a School Programme: Methodological Issue". In *Learner Corpora and Second Language Acquisition Research*, edited by B. LeBruyn and M. Paquot. Cambridge: Cambridge University Press.

Biber, D. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, D. and S. Conrad. 2019. *Register, Genre, and Style* (Cambridge Textbooks in Linguistics). Second edition. Cambridge: Cambridge University Press. doi:10.1017/9781108686136

Breeze, R. 2007. "How Personal is this Text? Researching Writer and Reader Presence in Student Writing Using WordSmith Tools". *CORELL: Computer Resources for Language Learning* 1: 14-21. https://www.researchgate.net/profile/Ruth_Breeze/publication/266491163_How_Personal_is_this_Text_Researching_Writer_and_Reader_Presence_in_Student_Writing_Using_Wordsmith_Tools/links/5448d1ea0cf2d62c3052c94b.pdf

Brezina, V. 2018. “Collocation graphs and networks: Selected applications”. In *Lexical collocation analysis: Advances and applications*, edited by P. Cantos-Gómez and M. Almela-Sánchez, 59–83. Cham: Springer International.

Broisson, Z. and K. Van Goethem. 2018. “Comparative constructions in French-speaking Belgian learners of English: A contrastive approach”. Poster presented at the *Using Corpora in Contrastive and Translation Studies Conference* (5th edition), 12-14 Sept. 2018, Louvain-la-Neuve, Belgium.

Bulon, A. Forthcoming. “Comparing the ‘Phrasicon’ of Teenagers in Immersive and Non-Immersive Settings: Does Input Quantity Impact Range and Accuracy?” *Journal of Immersion and Content-Based Language Education*.

Bulon, A. 2019. *The Acquisition of Phraseological Units by French-Speaking Learners of English and Dutch in CLIL and Non-CLIL Settings: Exposure Effects on Range and Accuracy*. PhD dissertation. Louvain-la-Neuve: Université catholique de Louvain.

Bulon, A., I. Hendrikx, F. Meunier, and K. Van Goethem. 2017. “Using Global Complexity Measures to Assess Second Language Proficiency: Comparing CLIL and Non-CLIL Learners of English and Dutch in French-Speaking Belgium.” *Travaux du CBL* 11 (1): 1-25. https://sites.uclouvain.be/bkl-cbl/wp-content/uploads/2017/04/Bulon_et_al_2017.pdf

Buntinx, N. and K. Van Goethem. 2018. “Cross-linguistic perspectives on intensification in speech: A comparison of L1 French and L2 English and Dutch”. Poster presented at the *Using Corpora in Contrastive and Translation Studies Conference* (5th edition). 12-14 Sept. 2018. Louvain-la-Neuve, Belgium

Callies, M. 2015. “Variability in Learner Corpora.” In *The Cambridge Handbook of Learner Corpus Research*, edited by S. Granger, G. Gilquin, and F. Meunier, 401-421. Cambridge: Cambridge University Press.

Callies, M. and M. Paquot. 2015. “Learner Corpus Research: An Interdisciplinary Field on the Move.” *International Journal of Learner Corpus Research*, 1 (1): 1-6. doi: 10.1075/ijlcr.1.1.00edi issn 2215–1478 / e-issn 2215–1486

Dalton-Puffer, C. 2011. “Content-and-Language Integrated Learning: From Practice to Principles?” *Annual Review of Applied Linguistics*, 31: 182-204. doi: 10.1017/S0267190511000092

Degand, L. and J. Perrez. 2004. “Causale connectieven in het leerdercorpus Nederlands.” *N/F. Tijdschrift van de Association des néerlandistes de Belgique francophone*, 4: 115-128. https://orbi.uliege.be/bitstream/2268/134576/1/degand&perrez(2004)-nf.pdf

Dumont, M. 2018. “Woordvolgorde bij CLIL- en niet-CLIL-leerders van het Nederlands: één pot nat? Vergelijkend onderzoek naar de beheersing van de PV2- en V-eind-regels bij (niet-)CLIL-leerders en bij moedertaalsprekers”. Master’s dissertation. Université catholique de Louvain.

Dunn, L. and L.M. Dunn. 2007. *PPVT-4: Peabody Picture Vocabulary Test*. Pearson Assessments.

Dunn, L., L.M. Dunn, and J. E. P. T. Schlichting. 2005. *Peabody Picture Vocabulary Test-III-NL*. Harcourt Test Publishers.

Dunn, L., L.M. Dunn, and C. Theriault-Whalen. 1993. *EVIP: Echelle de Vocabulaire en Images Peabody*. Pearson Canada Assessment.

Entwisle, D.R. and N.M. Astone. 1994. “Some practical guidelines for measuring youth’s race/ ethnicity and socioeconomic status.” *Child Development*, 65: 1521-1540. doi: 10.1111/j.1467-8624.1994.tb00833.x

Gentil, G. and F. Meunier. 2018. “A systemic functional linguistic approach to usage-based research and instruction. The case of nominalization in L2 academic writing.” In *Usage-inspired L2 instruction. Researched pedagogy*, edited by A. E. Tyler, L. Ortega, M. Uno, and H.I. Park, 267-289. Amsterdam and Philadelphia: John Benjamins.

Gilquin, G. 2008. “Combining Contrastive and Interlanguage Analysis to Apprehend Transfer: Detection, Explanation, Evaluation.” In *Linking up Contrastive and Learner Corpus Research*, edited by G. Gilquin, S. Papp, and M.B. Díez-Bedmar, 3-33. Amsterdam: Rodopi.

Gilquin, G. 2015. “From Design to Collection of Learner Corpora.” In *The Cambridge Handbook of Learner Corpus Research*, edited by S. Granger, G. Gilquin, and F. Meunier, 9-34. Cambridge: Cambridge University Press.

Granger, S. 1996. “From CA to CIA and back: An Integrated Approach to Computerized Bilingual and Learner Corpora.” In *Languages in Contrast. Text-Based Cross-Linguistic Studies*, edited by K. Aijmer, 37-51, Lund: Lund University Press.

Granger, S., ed. 1998. *Learner English on Computer*. London: Addison Wesley Longman.

Granger, S. 2002. “A Bird’s-Eye View of Learner Corpus Research.” In *Language Learning and Language Teaching Volume 6: Computer Learner Corpora Second Language Acquisition and Foreign Language Teaching*, edited by S. Granger, J. Hung, and S. Petch-Tyson, 3-33. Amsterdam: John Benjamins Publishing Company.

Granger, S. 2009. “The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation”. In *Corpora and language teaching*, edited by K. Aijmer, 13–32. Amsterdam, The Netherlands: John Benjamins.

Granger, S. 2015. “Contrastive Interlanguage Analysis: A Reappraisal.” *International Journal of Learner Corpus Research,* 1 (1): 7-24. doi: 10.1075/ijlcr.1.1.01gra

Granger, S. and Y. Bestgen. 2014. “The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study”. *International Review of Applied Linguistics in Language Teaching*, 52(3), 229–252. doi:10.1515/iral-2014-0011

Granger, S., E. Dagneaux, and F. Meunier. 2002. *The International Corpus of Learner English. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.

Granger, S., F. Meunier, and M. Paquot. 2009. *The International Corpus of Learner English – version 2. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.

Granger, S. and M. Paquot. 2009. “Lexical Verbs in Academic Discourse: A Corpus-Driven Study of Learner Use.” In *Academic Writing. At the Interface of Corpus and Discourse*, edited by M. Charles, D. Pecorari, and S. Hunston, 193-214. London: Continuum.

Granger, S., G. Gilquin, and F. Meunier, eds. 2013. *Twenty Years of Learner Corpus Research: Looking Back, Moving Ahead*. Louvain-la-Neuve: Presses universitaires de Louvain.

Granger, S., G. Gilquin, and F. Meunier, eds. 2015. *The Cambridge Handbook of Learner Corpus Research.* Cambridge: Cambridge University Press.

Hendrikx, I., K. Van Goethem, and S. Wulff. 2019. “Intensifying Constructions in

French-Speaking L2 Learners of English and Dutch: Cross-Linguistic Influence and Exposure Effects." *International Journal of Learner Corpus Research*, 5 (1): 63-103. doi: 10.1075/ijlcr.18002.hen

Hendrikx, I. 2019. *The acquisition of intensifying constructions in Dutch and English by French-speaking CLIL and non-CLIL students: Cross-linguistic influence and exposure effects*. PhD dissertation, Université catholique de Louvain.

Hiligsmann, P. and L. Rasier. 2011. "Het geïntegreerde contrastieve onderzoeksdesign onder de loep." In *Nederlands in het perspectief van uitspraakverwerving en contrastieve taalkunde Lage Landen Studies*, edited by L. Rasier, V. van Heuven, B. Defrancq, and P. Hiligsmann, 183-194. Gent: Academia Press.

Hiligsmann, P., L. Van Mensel, B. Galand, L. Mettewie, F. Meunier, A. Sczmalec, K. Van Goethem, A. Bulon, A. De Smet, I. Hendrikx, and M. Simonis. 2017. "Assessing Content and Language Integrated Learning (CLIL) in the French-speaking Community of Belgium: linguistic, cognitive and educational perspectives". *Cahiers du GIRSEF*, 109, 1-24.

Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Jarvis, S. 2000. "Methodological Rigor in the Study of Transfer: Identifying L1 Influence in the Interlanguage Lexicon." *Language Learning*, 50 (2), 245-309. doi: doi.org/10.1111/0023-8333.00118

Klein, W. and C. Perdue. 1997. "The Basic Variety (Or: Couldn't Natural Languages Be Much Simpler?)." *Second Language Research,* 13 (4): 301-347. doi: 10.1191/026765897666879396

Littré, D. 2014. *A Cognitive, Longitudinal Study of the Use of the English Present Progressive by Intermediate and Advanced French-Speaking Learners*. PhD

dissertation, Université catholique de Louvain.

Lozano, C. and A. Mendikoetxea. 2013. Learner corpora and Second Language Acquisition: The design and collection of CEDEL2. In *Automatic Treatment and Analysis of Learner Corpus Data*, edited by A. Díaz-Negrillo, N. Ballier, and P. Thompson, 65-100. Amsterdam: John Benjamins.

Marsden, E., A.Mackey, and L. Plonsky. 2016. "The IRIS Repository: Advancing research practice and methodology." In *Advancing methodology and practice: The IRIS repository of instruments for research into second languages,* edited by A. Mackey and E. Marsden, 1-21. New-York: Routledge.

McEnery, T., V. Brezina, D. Gablasova, and J. Banerjee. 2019. "Corpus Linguistics, Learner Corpora, and SLA: Employing Technology to Analyze Language Use". *Annual Review of Applied Linguistics*, 39, 74-92. doi:10.1017/S0267190519000096

Myles, F. 2015. "SLA Theory and Learner Corpus Research." In *The Cambridge Handbook of Learner Corpus Research*, edited by S. Granger, G. Gilquin, and F. Meunier, 309-331. Cambridge: Cambridge University Press.

Mukherjee, J. and S. Götz. 2015. "Learner Corpora and Learning Context." In *The Cambridge Handbook of Learner Corpus Research*, edited by S. Granger, G. Gilquin, and F. Meunier, 423-442. Cambridge: Cambridge University Press.

Ortega, L. and H. Byrnes. 2018. "Theorizing Advancedness, Setting up the Longitudinal Research Agenda." In *The Longitudinal Study of Advanced L2 Capacities*, edited by L. Ortega and H.Byrnes, 281-300. New-York: Routledge.

Paquot, M., H. Naets, and S.Th. Gries. Accepted. "Using syntactic co-occurrences to trace phraseological complexity development in learner writing: verb + object structures in LONGDALE." In *Second Language Acquisition and Learner*

Corpora, edited by B. Le Bruyn and M. Paquot. Cambridge: Cambridge University Press.

Perdue, C., ed. 1993. *Adult Language Acquisition: Volume 2, the Results: Cross-Linguistic Perspectives*. Cambridge: Cambridge University Press.

Raven, J.C., J.H. Court, and J. Raven. 1998. *Progressive Coloured Matrices.* Oxford: Oxford Psychologists Press.

Sinclair, J. 1996. EAGLES. Preliminary Recommendations on Corpus Typology. *EAGLES Document EAG TCWG-CTYP/P* http://www.ilc.cnr.it/EAGLES/corpustyp/corpustyp.html

The New London Group. 1996. "A Pedagogy of Multiliteracies: Designing Social Futures". *Harvard Educational Review* 66 (1): 60–93. doi:10.17763/haer.66.1.17370n67v22j160u.

Van Mensel, L., A. Bulon, I. Hendrikx, F. Meunier, and K. Van Goethem. Accepted. "Effects of Input on L2 Writing Skills in English and Dutch: CLIL and Non-CLIL Learners in French-Speaking Belgium." *Journal of Immersion and Content-Based Language Education*.

Van Mensel, L., P. Hiligsmann, L. Mettewie, and B. Galand. 2019. "CLIL, an Elitist Language Learning Approach? A Background Analysis of English and Dutch CLIL Pupils in French-Speaking Belgium." *Language, Culture and Curriculum*. DOI: 10.1080/07908318.2019.1571078.

Yoon, J. and Gries, S.Th., eds. 2016. *Corpus-based Approaches to Construction Grammar*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Appendixes

A. Writing task instructions (TL Dutch)/Topic: holidays

Schrijf een e-mail in het Nederlands (15 à 25 regels) aan een vriend(in) en vertel hem of haar over je vakantie.

Beschrijf wat je het allerleukst, het mooist en het grappigst vond. Vertel ook wat je juist helemaal niet leuk vond.

Als je niet zoveel inspiratie hebt, kun je bijvoorbeeld schrijven over de mensen die je ontmoet hebt, het weer, de problemen die je had, de sfeer in het algemeen, het eten, ... Je kunt je eventueel ook baseren op de vakantiefoto's hieronder.

Van: eenvoorbeeld@email.com

Voor: eenandervoorbeeld@email.com

Onderwerp: Vakantie

Bericht:

B. Writing task instructions (TL Dutch)/ Topic: party

Schrijf een e-mail in het Nederlands aan een vriend(in) en vertel hem of haar over een feestje waar je gisteren bent geweest.

Beschrijf wat je het allerleukst, het grappigst en het vreemdst vond. Vertel ook wat je juist helemaal niet leuk vond.

Als je niet zoveel inspiratie hebt, kun je bijvoorbeeld schrijven over de muziek, de sfeer, de gasten, het eten; de cadeaus, de acts/performances,… Je kunt je eventueel ook baseren op de foto's hieronder.

Van: eenvoorbeeld@email.com

Voor: eenandervoorbeeld@email.com

Onderwerp: Feestje

Bericht:

C. Writing task instructions (TL English)/Topic: holidays

Write an e-mail in English (15 to 25 lines) to a friend to tell him/her about your holidays. Describe what you liked most, and tell about the things that were beautiful or funny. Also, tell him/her about things that happened and that you did not like at all.

You may refer to the holiday pictures below for inspiration.

From: anexample@email.com

To: anotherexample@email.com

Object: Holidays

Message:

D. Writing task instructions (TL English)/Topic: party

Write an e-mail in English (15 to 25 lines) to a friend to tell him/her about the party you attended last night.

Describe what you liked most, and speak about some funny and crazy things that happened last night. Also, tell him/her about things that happened and that you did not like at all.

You may want to talk about the music, the atmosphere, the guests, the food, the presents, the performances, etc. You may also refer to the pictures below for inspiration.

From: anexample@email.com

To: anotherexample@email.com

Object: Party

Message:

E. Writing task instructions (L1 French)/Topic: holidays

Ecris un e-mail en français (15 à 25 lignes) à un(e) ami(e) et raconte-lui comment étaient tes vacances.

Raconte-lui ce qui a été le plus sympa, le plus beau et le plus drôle, et raconte-lui aussi des éléments de tes vacances que tu n'as pas du tout aimés.

Si tu es en manque d'inspiration, tu peux éventuellement parler des personnes que tu as rencontrées, la météo, les problèmes, l'ambiance, la nourriture, etc. Tu peux aussi t'inspirer des photos de vacances ci-dessous.

De: unexemple@email.com

A: unautreexemple@email.com

Objet: Vacances

Message:

F. Writing task instructions (L1 French)/Topic: party

Ecris un e-mail en français (15 à 25 lignes) à un(e) ami(e) et raconte-lui comment était la fête à laquelle tu as participé hier soir.

Raconte-lui ce qui a été le plus sympa, le plus bizarre et le plus drôle, et raconte-lui aussi des éléments de la fête que tu n'as pas du tout aimés.

Si tu es en manque d'inspiration, tu peux éventuellement parler de la musique, l'ambiance, les invités, la nourriture, les cadeaux, les performances, etc. Tu peux aussi t'inspirer des photos ci-dessous.

De: unexemple@email.com

A: unautreexemple@email.com

Objet: Fête

Message:

G. Oral task instructions/ Topic: party

Tu vas participer à un petit exercice de quelques minutes durant lequel tu vas travailler par groupe de deux. Tu vas recevoir une carte avec des images qui sont différentes de celles de ton partenaire. A aucun moment tu ne pourras observer la carte de ton interlocuteur.

Contexte : La carte est une invitation pour 2 personnes à une fête. Ton ami(e) en a reçu une aussi, mais ce n'est pas la même. Imagine que vous vous êtes donné rendez-vous pour en discuter et décider à quelle fête vous irez.

Tâche : Commence par poser des questions à ton partenaire afin d'en savoir plus sur la fête à laquelle il/elle est invité(e). Ensuite, essaye de le/la convaincre d'aller à la fête décrite sur ta carte. Vous devez arriver à un compromis.

Durant cet exercice, tu utiliseras l'anglais/néerlandais. Veille à répondre clairement aux questions de ton interlocuteur et à lui demander le plus d'informations possible. Si tu ne comprends pas ses questions, n'hésite pas à lui faire répéter.

H. Oral task instructions/ Topic: holidays

Tu vas participer à un petit exercice de quelques minutes durant lequel tu vas travailler par groupe de deux. Tu vas recevoir une carte avec des images qui sont différentes de celles de ton partenaire. A aucun moment tu ne pourras observer la carte de ton interlocuteur.

Contexte : Tu as gagné un voyage pour 2 personnes en Italie. Ton ami(e) en a gagné aussi, mais pas dans le même pays. Imagine que vous vous êtes donné rendez-vous pour en discuter et décider de l'endroit où vous irez.

Tâche : Commence par poser des questions à ton partenaire afin d'en savoir plus sur le type de voyage qu'il/elle a gagné. Ensuite, essaye de convaincre ton partenaire d'aller à l'endroit décrit sur ta carte. Vous devez arriver à un compromis.

Durant cet exercice, tu utiliseras l'anglais/néerlandais. Veille à répondre clairement aux questions de ton interlocuteur et à lui demander le plus d'informations possible. Si tu ne comprends pas ses questions, n'hésite pas à lui faire répéter.

I. Sample questions from questionnaire

Frequency of informal contact with the second language outside school, on a scale from 1 to 5 (never – rarely – sometimes – often – very often), based on the following items:

[Original in French]

1. Sur internet, à quelle fréquence utilises-tu l'anglais / le néerlandais?
2. A quelle fréquence parles-tu l'anglais / le néerlandais en dehors de l'école?
3. A quelle fréquence entends-tu l'anglais / le néerlandais en dehors de l'école?

4. As-tu des contacts avec des anglophones / des néerlandophones en dehors de l’école ou de la maison?

[English translation]

1. How often do you use English / Dutch on the Internet?
2. How often do you speak English / Dutch outside school?
3. How often do you hear English / Dutch outside school?
4. Do you have any contact with English-speakers / Dutch-speakers outside school or home?