The potential of publicly shared longitudinal learner corpora in SLA research

Nicole Tracy-Ventura & Amanda Huensch

University of South Florida

Department of World Languages, University of South Florida, 4202 E. Fowler Avenue CPR 107,

Tampa, FL, USA 33620, Tel.: +1 813 974 2548, Email address: huensch@usf.edu

Abstract

Most second language acquisition researchers would agree that longitudinal data can potentially yield the most valuable insights into second language development, yet few longitudinal studies exist. It is also rare to find longitudinal studies with data collected beyond the one year mark. In this chapter we argue for the need to adopt more longitudinal research agendas in SLA and the potential of publicly shared learner corpora to help broaden the impact of the data collected. We use as an example work from the longitudinal project, LANGSNAP, which now includes a new set of data collected in 2016, expanding the project to 5 years and with a new research question, that of the long-term evolution of foreign language proficiency.

Key words: second language acquisition; longitudinal research; learner corpus research; Spanish as a second language; French as a second language; foreign language attrition; study abroad

Introduction

Most second language acquisition (SLA) researchers would agree that longitudinal data can potentially yield the most valuable insights into second language (L2) development by allowing us to examine L2 learning as it is taking place over time (Myles, 2008; Ortega & Byrnes, 2008; Ortega & Iberri-Shea, 2005). However, despite this understanding of the benefits of longitudinal research, the field continues to favor cross-sectional research designs for a variety of reasons. Harklau (2008) discussed logistical problems of time and resources, related in many ways to the structure of academia. In North America, for example, the tenure clock (and promotion to full professor) places immense pressure on researchers to produce short-term projects with the goal of a specific quantity of publications, ideally in top-tier journals. Additionally, well-designed longitudinal studies with more than a few participants can be expensive and may require financial support which is often difficult to secure in our field. Nonetheless, several examples of longitudinal SLA studies exist, some of which date back to the early days of SLA (e.g. Leopold, 1939; Huebner, 1983; Perdue, 1993; Schumann, 1978) and others which are more recent (Derwing & Munro, 2013; Jia & Aaronson, 2003; Meunier & Littre, 2013; Mitchell, Tracy-Ventura, & McManus, 2017; Skiba, Dittmar, & Bressem, 2008; Vyatkina, 2013). Dedicating time and resources to collecting longitudinal data is necessary if we are to better understand the pattern of L2 development from the early days of language learning to advanced abilities (Ortega & Byrnes, 2008).

In this chapter we expand upon this argument and discuss in detail many of the practical challenges of planning and conducting longitudinal research, and in our case, building a longitudinal learner corpus that is shared publicly. We follow the suggestions of Myles (2005, 2008), Meunier (2015), and MacWhinney (2017) who call for sharing and making publicly

available learner corpora which are formatted with agreed-upon conventions for transcribing, storing, and analyzing data. The longitudinal learner corpus we describe in this chapter is the LANGSNAP corpus, initiated by Mitchell et al. (2017) which includes data collected six times during May 2011-February 2013 from 56 university learners of French and Spanish. Recently, we have continued this project and during May-June 2016 collected data from a subsample of the original participants (*n*=33), those who volunteered to participate again. The follow-up project has been titled LANGSNAP 3.0 to indicate that the new data were collected approximately 3 years after the final data collection wave of the original project (February 2013), increasing the length of the study to 5 years. In what follows we discuss the importance of longitudinal data for SLA, some recent longitudinal studies, and the potential of publicly shared electronic learner corpora to help broaden the impact of valuable longitudinal data. Additionally, we provide a critical reflection on issues related to developing and implementing our longitudinal SLA project and its resulting learner corpus with the goal of informing future work in this area.

Literature Review

Longitudinal research in SLA

As described above, longitudinal research has long been a part of SLA because it is arguably the most reliable data for describing the development of L2 abilities. Yet even today, longitudinal research struggles to become a priority. In their 2008 edited collection focusing on longitudinal research and advanced L2 capacities, Ortega and Byrnes claimed that,

> The lack of an explicit or sustained focus on longitudinal questions has meant that, after some 40 years of disciplinary history, we know little about the longitudinal pace and pattern of development in second language and literacy, much less when development is

understood to span the lifetime of multilingual and multicultural people who set out to function in several languages including an L2 (p. 3).

Nearly ten years later, the situation remains barely unchanged with a few notable exceptions such as Derwing and Munro (2013), a 7-year study which is described later. The length of longitudinal studies varies, and many times coincide with institutional schedules (e.g., a semester or an academic year) (Ortega & Iberri-Shea, 2005). A year is a popular length in much longitudinal SLA research (Baba & Nitta, 2014; Polat & Kim, 2014; Young-Scholten & Langer, 2015; Zheng, 2012), a benchmark used by the *Inventory of longitudinal studies in the social sciences* (Young, Savola, & Phelps, 1991, as cited in Harklau, 2008). Ortega and Iberri-Shea (2005) suggested that when researchers are planning the length of a longitudinal study, major events in the social or institutional context should inform the length of observation.

Due to the amount of time and resources needed for longitudinal research most studies tend to be qualitative case-studies of a few participants with the addition of some quantitative data on language development (e.g., Kinginger & Blattner, 2008; Spenader, 2008; Zheng, 2012). Rees and Klapper (2008), in their discussion of longitudinal study abroad research, argued for combining both quantitative and qualitative data, as well as improving the statistical rigor used in longitudinal research. For example, they highlight the lack of instrument reliability which gets reported in longitudinal studies and the limited discussion of how missing values are treated when using inferential statistics.

Since the publication of Ortega and Byrnes' (2008) edited collection, one area in which more longitudinal research has appeared is in studies motivated by Dynamic Systems Theory (DST: Larsen-Freeman & Cameron, 2008). In DST, longitudinal studies are necessary to "capture variability at various levels and timescales, from the general shape of the development process over a long period of time to the short-term variability that takes place between data collection

intervals, to the within-session variability that inevitably arises" (Larsen-Freeman & Cameron, 2008, p. 208). Baba and Nitta (2014) used DST to explore two EFL university students' writing fluency in 30 samples collected weekly throughout an academic year. Polat and Kim (2014) examined one untutored ESL learner's complexity, accuracy, and lexical diversity using DST in data from oral interviews conducted once every two weeks for a year. Such studies typically include few participants but data are collected at multiple, frequent intervals throughout the study which aids in the identification of critical moments in L2 development.

Other examples of longitudinal studies exist which focus specifically on different linguistic features. For example, Derwing and Munro (2013) report on a 7-year longitudinal study examining the development of oral skills (comprehensibility, fluency, and accentedness based on ratings of perceived fluency) in adult immigrant learners in Canada. Unlike most longitudinal studies, their participants were from two different L1 groups: Mandarin (*n*=11) and Slavic language speakers (Russian=7; Ukrainian=4). Several earlier publications focused on data from the first 1-2 years of this study (e.g., Derwing, Munro, & Thomson, 2008; Derwing, Munro, Thomson, & Rossiter, 2009; Munro & Derwing, 2008). Results from Derwing and Munro (2013) focused on the change from the 2-year to the 7-year points, demonstrating differences between the two L1 groups. The Slavic speakers improved over those five years in comprehensibility and fluency (although not accentedness), whereas the Mandarin speakers did not. Derwing and Munro also compared the scores at year 7 with other variables such as years of prior English study and age of arrival. No significant correlations were found for years of prior English study but significant correlations were found between age of arrival and comprehensibility, as well as age of arrival and accentedness.

Crossley, Salsbury, and McNamara (2010) used corpus-based tools to investigate polysemy, or "the use of words with multiple senses and their growth in use with L2 learners" (p. 578). Their data came from six L2 English false-beginner participants who were interviewed every two weeks throughout one year. Interviews lasted 30-45 minutes and were later transcribed and analyzed quantitatively using computational tools based on lexical databases such as WordNet (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990) and CELEX word frequency data from the COBUILD corpus (Baayen, Piepenbrock, & Gulikers, 1996). Results demonstrated that the six learners showed quick gains within the first four months but then plateaued in the frequency of polysemous words produced. However, further qualitative analysis provided evidence that they were able to use more varied senses of a lexical item after four months, even though the actual frequency of overall use of that lexical item did not change. This study is a good example of how corpus-based tools can be used to investigate longitudinal change of linguistic foci that would be difficult to assess without the use of automatic analysis programs. It also demonstrates the important contribution of qualitative analysis to the interpretation of quantitative findings (i.e., frequency counts), a key analytic characteristic of corpus linguistics (see Biber, Conrad, & Reppen, 1998). Learner corpus research (LCR) involves the use of automatic and interactive computer programs to analyze L2 data, but as described by Myles (2015), this area of research has not typically been informed by SLA research and SLA research has also been generally slow to adopt computerized tools used in LCR. In the following section we follow the arguments of Myles (2008) and Meunier (2015), and discuss the benefit of longitudinal SLA-informed learner corpora and in particular, those which are made publicly available.

The potential of publicly shared longitudinal learner corpora

Once two quite separate fields, recently SLA and LCR have begun to come together in the joint effort to understand L2 development. As evidence of this rapprochement, a special issue of *The Modern Language Journal* was dedicated to exploring the relevance of LCR for SLA research, focusing specifically on the potential of LCR in supporting a more longitudinal orientation in SLA. Articles in the special issue described various L1-L2 pairs such as English-German (Vyatkina, 2013), English-Spanish (Yuldashev, Fernandez, & Thorne, 2013), and French-English (Meunier & Littre, 2013), among others. These articles demonstrated how questions of central concern in SLA could be analyzed using corpus-based tools. For example, Vyatkina (2013) focused on the development of syntactic complexity in writing over four semesters of university German classes. Her two learners were beginners at the start of the study and produced 4-5 writing samples per semester. The resulting learner corpus was analyzed using a combination of automatic, semi-automatic, and interactive tagging procedures which allowed her to compare frequency values of the complexity measures investigated in the study.

Myles (2008) was one of the first to discuss the benefits of electronic longitudinal learner corpora for SLA. She argued that longitudinal datasets are the most suitable for documenting and analyzing learner development, but also acknowledged the logistical challenges of studying the same participants over long periods of time. Because such data can be difficult to attain, she argued for making longitudinal data publicly available to the research community as electronic corpora. Such a move would allow researchers from different points of view to work on the data and to help the field identify other kinds of datasets that are needed. In order for datasets to be shared in an efficient way, Myles highlighted the importance of using agreed-upon conventions

for transcribing, storing, and analyzing data so that other researchers can also work with the

corpora, and take advantage of readily available computerized tools. In her work with Rosamond

Mitchell and colleagues on the French Learner Language Oral Corpora project (FLLOC: Myles,

2005; Myles & Mitchell, 2004) and the Spanish counterpart, SPLLOC (Domínguez, Tracy-

Ventura, Arche, Mitchell, & Myles, 2013; Mitchell, Domínguez, Arche, Myles, & Marsden,

2008),[1] tools from the CHILDES system have been adopted (MacWhinney, 2000). The

CHILDES team pioneered the principle of publicly accessible learner corpora first for research

on first language acquisition but now include SLA corpora as well. They continue to support and

develop their Talkbank resources and analysis tools, which are shared freely with the

international research community. MacWhinney (2017) describes the SLABank and BilingBank

Corpora which are currently part of the CHILDES database, in addition to the tools which are

available as part of the CLAN program. While the SLA corpora continue to grow, MacWhinney

states that "The biggest gap in current data on SLA is that we have no openly available densely

collected (Maslen, Theakston, Lieven, & Tomasello, 2004) longitudinal SLA data" (p. 7). The

LANGSNAP corpus, with its now 5-year longitudinal data, hopefully will begin to address this

gap.

LANGSNAP and LANGSNAP 3.0

Our current project contributes new data to the Language and Social Networks Abroad

Project (LANGSNAP: Mitchell et al., 2017) which began in 2011 and is described next. The new

data allows us to investigate the long-term evolution of foreign language (FL) proficiency post-

---

[1] These learner corpora are both accessible online and free to download from their individual websites (http://www.flloc.soton.ac.uk; http://www.splloc.soton.ac.uk) or via the CHILDES SLA resource at www.talkbank.org.

instruction and study/residence abroad, which for our participants includes the possible outcomes of attrition, maintenance, or development. The long-term evolution of FL proficiency is an important and understudied area of research relevant to SLA. We acknowledge that the term "foreign language" may be problematic for some, but we chose to use it following Schmid and Mehotcheva (2012) who argued that a distinction should be made regarding attrition of languages learned naturalistically (e.g., immigrants learning the language of their new country) and those learned via explicit instruction. They refer to the former as L2 attrition and the latter as FL attrition. The LANGSNAP participants were first instructed language learners, and for that reason we will use the term FL throughout the paper.

As a previously funded project by the UK Economic and Social Research Council (May 2011-October 2013), LANGSNAP tracked 56 university language majors (the majority were English L1 speakers) over the course of two years: once prior to residence abroad, three times during their 9-month stay abroad program, and two times after return to their home university. One group of students spent the academic year in France (n=29), one group in Spain (n=18) and another in Mexico (n=9). The purpose of LANGSNAP was to explore the sources of individual variation found in the SLA research on the effects of residence/study abroad by focusing on social, contextual, and individual differences (e.g., social networks, placement type, and personality). The project's resulting oral and written longitudinal learner corpus (over 700,000 words) has been carefully transcribed and made publicly available online for other researchers' use on the project's website: http://langsnap.soton.ac.uk/ and will soon be available on talkbank.org. Major findings of this study demonstrated that both L2 groups made significant gains in their oral skills but few in their written skills. Changes in oral fluency are described in more detail in Huensch and Tracy-Ventura (2017b), and more general results of complexity,

accuracy, and fluency in both speech and writing are described in Mitchell et al. (2017). More intense L2-speaking social networks were beneficial for fluency development but not accuracy development. One personality trait that helped explain linguistic gains was flexibility, as measured on the Multicultural Personality Questionnaire (van der Zee & van Oudenhoven, 2000). Those participants who were more flexible tended to make more gains in fluency and accuracy in oral tasks, and in the measure of oral proficiency (McManus, Tracy-Ventura, & Mitchell, 2016).

Building on LANGSNAP, the follow-up project, known as LANGSNAP 3.0, involved collecting new data from the same participants to investigate the detailed processes of attrition/maintenance/development after return from study/residence abroad and post target-language instruction. The limited research to date suggests that some aspects of linguistic gains made while abroad may be fragile and subject to attrition after return home (Howard, 2012). Since graduation many of the LANGSNAP participants have been living in a mainly English-speaking environment and working in jobs that require little or no use of French or Spanish. Therefore, we anticipated that many of these participants would be experiencing FL attrition to some degree. However, others have spent further time abroad, and some are known to be employed in L2-using professions (e.g., language teaching). The longitudinal nature of LANGSNAP and the fact that a variety of language skills were analyzed each time and arguably at the time of peak attainment (the end of their year abroad; see Bardovi-Harlig & Stringer, 2010) provide a means by which to investigate the long-term evolution of FL proficiency. In particular, the specific aims of the follow-up project were to (a) examine the extent to which LANGSNAP participants have experienced attrition/maintenance/development in French and Spanish in the following areas: general proficiency; vocabulary knowledge; accuracy, fluency,

lexical diversity, and syntactic complexity in oral and written production; (b) examine the extent to which personal and external factors (e.g., proficiency, self-reported proficiency, motivation, amount of L2 use) predict participants' differential levels of attrition/ maintenance/ development; and (c) further develop the publicly available longitudinal corpus for the research community.

To accomplish these aims, we traveled to England in the summer of 2016 to reconnect with and collect new data from as many of the original LANGSNAP participants who were willing to participate once again. We received three grants to support the project: one from the *Language Learning* Small Research Grant program, and two internal grants from our university. These grants paid for travel and equipment, as well as support for two graduate research assistants to help with the transcription and coding of the data which has since been made publicly available, following the tradition of LANGSNAP and other learner corpus projects initiated at the University of Southampton (e.g., FLLOC & SPLLOC). In the sections that follow, we provide a critical reflection on issues related to developing and implementing this longitudinal SLA project and its resulting learner corpus. We discuss challenges that have surfaced with regard to participant attrition, project planning, data collection, and data preparation, and provide insights based on the compromises we have made with the goal of informing future work in this area.

Considerations for mitigating participant attrition

For longitudinal research, mitigating participant attrition is a particularly important concern because once the first wave of data collection has occurred, the participant sample is limited to those participants. Therefore, planning ahead and implementing strategies from the outset to reduce participant attrition is critical. One strategy that has been beneficial in our work

is a commitment to building rapport and a strong personal connection with participants which began during LANGSNAP when they were visited three times by a member of the research team during their 9-month stay abroad. Additionally, a private Facebook group was established during LANGSNAP to allow the project researchers and participants to stay connected. Using this group, we were able to post announcements about upcoming visits, and also check in with the participants after the project ended and post information about our findings. When we began to plan the follow-up study, this group became a valuable resource. For example, we were able to see that most of the participants were working in the UK and only a few were working abroad. To get a sense of how many would be willing to participate in a follow-up study, in October 2015 we posted a description of our intention to collect new data and asked who might be interested in participating. Within 24 hours, 50% of the participants (*n*=26) had enthusiastically responded. We believe that such a response demonstrated the strength of the relationship that was built between the participants and the LANGSNAP research team. In fact, over the course of the original project there was a very low rate of participant attrition, only three of 59 participants over approximately two years. Therefore, to maintain this relationship so that we could continue our long-term investigation of FL proficiency beyond this first follow-up study, we felt it was necessary to again visit the participants in person to conduct the interviews and language assessments. Furthermore, collecting face-to-face data also meant that we would have the highest possible sound quality of the oral data, which is critical for acoustic analyses in Praat (Boersma & Weenink, 2015). Ultimately, establishing a strong relationship with participants and creating a space for continued contact have been two key components in the success of the current longitudinal project.

Another aspect of longitudinal research that could influence participant attrition is personnel turnover. Because of this, having participants grow accustomed to interacting with multiple members of the team can be worthwhile. For example, in LANGSNAP, the participants met several different members of the research team so the addition of Amanda Huensch, who joined the research group after 2013, went smoothly. As co-investigator on LANGSNAP 3.0, she was heavily involved with all aspects of the project and occasionally met participants on her own. Initial contact was made by Nicole Tracy-Ventura, whom the participants already knew, but later Amanda interacted solely with several participants. Of course, it is not possible to completely avoid participant attrition in longitudinal research; however, by implementing several simple strategies this issue can be mitigated.

Considerations for study design: Using the same instruments or different?

Choice of instruments and their frequency of use is one of the most important considerations when designing longitudinal SLA research. As mentioned previously, the original LANGSNAP study included six data collection waves and explored linguistic development and its relationship to social networks and language use. The breadth of data collected in LANGSNAP positioned us well to initiate our current project on the long-term evolution of FL proficiency. Data included both spoken and written samples of learner language that ranged from oral picture narratives to observational data. For two of the tasks (oral picture narrative and written argumentative essay), three sets of comparable prompts were selected, piloted, and implemented in LANGSNAP. In this way, a balance was struck between task repetition effects (a year passed between the use of each prompt) and unwanted variation as a result of the specific

task (Weigle, 2002). For more details about the types of data collected and the tasks used in LANGSNAP see Tracy-Ventura, Mitchell, and McManus (2016).

Upon initiation of the follow-up project, we were faced with data collection decisions in terms of which of the previously used instruments to administer and whether to adopt or develop anything new. To maintain continuity with the existing corpus, we chose to use the narrative and essay prompts that would have come 'next' in the LANGSNAP sequence. Given that it had been 3 years since the last data collection wave, we were less concerned with practice/familiarity effects. However, this meant including task prompts that were potentially less timely (e.g., argumentative writing prompts about gay marriage and adoption – we will return to this issue). Because the focus of LANGSNAP 3.0 was different and centered on the long-term evolution of FL proficiency, we needed to consider whether any new materials were necessary. Because interviews in both the participants' L1 and target language had been previously employed, with questions varying at each data collection wave, we decided to use these tasks to ask questions related to how participants had maintained contact, or not, with French and/or Spanish. We also decided to create a background questionnaire to gather additional information such as the kinds of jobs they have had since graduation, if they have had any romantic partners with whom they use their L2, how motivated they are to use their L2, etc. This questionnaire was modeled after the ones by Keijzer (2007) and Mehotcheva (2010) which we found on Monika Schmid's language attrition website (https://languageattrition.org/). To get a sense of how often the participants did various activities in their L2s, we included the Language Engagement Questionnaire that was developed for LANGSNAP (McManus, Mitchell, & Tracy-Ventura, 2014). We also administered the same proficiency test, the elicited imitation test (Ortega, Iwashita, Norris, & Rabie, 2002; Tracy-Ventura, McManus, Norris, & Ortega, 2014), and a test

of vocabulary knowledge, the Swansea X-lex test (Meara & Milton, 2003). Ultimately, the study design we chose for the 2016 data collection wave included many of the same instruments from LANGSNAP, allowing us to continue researching complexity, accuracy, fluency, lexis (CAFL), and proficiency, important constructs in SLA that had already been a part of the original LANGSNAP design.

Considerations for data collection

Longitudinal research also poses challenges related to data collection, particularly when the subjects are young multilingual adults who are interested in international travel. Therefore, for us there was a strong likelihood that many of our participants would be in different locations around the world. Because we wanted to collect data from all of the participants willing to take part, we accepted the fact that some data would need to be collected via Skype if a face-to-face meeting was not possible. In the end we were able to meet with 33 participants (59% of the original group), 28 of which were face-to-face meetings mostly spread across the south of England. For five of the participants it was necessary to conduct the interviews over Skype because they were living abroad and our budget would not cover such travel (e.g., Australia, France, Spain). Skype is a convenient tool though not ideal because the sound quality of the recordings can be variable depending on the internet speeds of both parties. Finally, it requires having online tasks that can be turned on and off, only allowing participants access while they are completing the tasks. Despite these potential challenges, using Skype turned out to be a viable option for us and provided a means for us to connect with five participants (15%) who we would have otherwise not been able to meet.

Of course, planning the visits was not without issues either. It takes a lot of coordination to plan 30+ visits and to try and stay within a modest travel budget. In addition, the participants were now young professionals as opposed to university students, so most of the time their schedules required meeting in the evenings or on weekends. We were grateful to many of the participants who were flexible with dates and willing to meet us in convenient locations where the background noise would be minimal, sometimes in their own flats even. Unfortunately that was not always possible and so we have some recordings that were made in coffee shops, restaurants, and even outside. There were a few participants who were unable to meet us at the last minute due to illnesses or lack of time. We also have examples of incomplete data (thankfully not many) mostly due to human error or technological failure. In these cases, when some data were collected and others not, we had to make a decision about whether to repeat tasks. We chose not to have participants complete the tasks again primarily to avoid differences across data collection but also to respect participants' time. In situations like these, the use of multiple tasks and data collection methods allows for at least partial inclusion of these participants' data.

Overall, previously established rapport, thorough planning, and the ability to be flexible and respect the time of our participants allowed us to connect with a majority of the LANGSNAP participants for this new project. We were also lucky to have the support of colleagues Rosamond Mitchell and Kevin McManus who were in the UK during this time and helped in a number of ways. Additionally, we hired a research assistant who was a L1 French speaker to assist with some of the French data collection.

Considerations for data preparation: *What to annotate for a publicly shared learner corpus?*

When preparing a learner corpus that will be shared publicly, a variety of considerations need to be kept in mind. These are questions such as what format/program to use for transcription, what to annotate in the transcripts, and what data to share publicly. As suggested by Myles (2008), using agreed-upon conventions for transcribing, storing, and analyzing data makes it easier for other researchers to work with the corpora and take advantage of readily available computerized tools. Thus, our first step in data preparation was the transcription of oral data following CHAT conventions (see MacWhinney, 2000), as was done in LANGSNAP and previous projects initiated by Myles, Mitchell, and colleagues at the University of Southampton who were some of the first to argue for the potential of data sharing in SLA. Today, as evidenced by a growing number of corpora available via CHILDES, many other SLA researchers format their data in CHAT and use CLAN for analysis as well. CHAT conventions allow a variety of metadata to be indicated within each file, which facilitates analysis in corpus research. For example, each transcript file includes a metadata header (marked with the @ symbol) to provide information about the corpus, the speakers, the data files, and other optional information that could be relevant for other researchers.

Regarding what to annotate in the transcript, we took an approach to make the data as broadly useful as possible, but of course, any transcription involves making decisions. For example, one seemingly non-controversial aspect of transcription would be to anonymize interviews to protect the identity of participants. In our case, this involved anonymizing names, places, institutions, etc. with the simple codes of 'Name', 'City'. However, this raises an issue of the usefulness of data for those wanting to conduct qualitative analysis where individual names could be important for understanding social relationships for example. Another consideration for

SLA corpora was whether and how to mark grammatical errors. For our purposes, the only

'errors' that are marked in our transcripts are mispronunciations or slight deviations in form that

made lexical items appear to be unknown words in the CLAN lexicon, and therefore

unanalyzable when the part-of-speech tagger was used (e.g., see line 19 in Figure 1 where

construiaba is uttered instead of *construía*). If there were other grammatical errors, such as errors

in grammatical gender or subject-verb agreement, they were not marked in the transcripts.

Another aspect of transcribing in CLAN is that one must decide how to break up

utterances. To facilitate automatic part of speech tagging and syntactic analysis, longer turns are

separated into a series of shorter turns, each of which begins a new main tier in the transcript.

This additional step means that even if the data is monologic, there will be multiple new lines

that begin with the same speaker ID. In our oral data, we followed the recommendations of

Foster, Tonkyn, and Wigglesworth (2000) and separated utterances into Analysis of Speech

Units (ASU). This unit is composed of any "independent clause, or subclausal unit, together with

any subordinate clause(s) associated with either" (p. 365). Figure 1 is an example of a transcript

from the picture-based narrative data collected for LANGSNAP 3.0. All the lines that begin with

@ are part of the header. Each line of the transcript that begins with *161, the speaker ID,

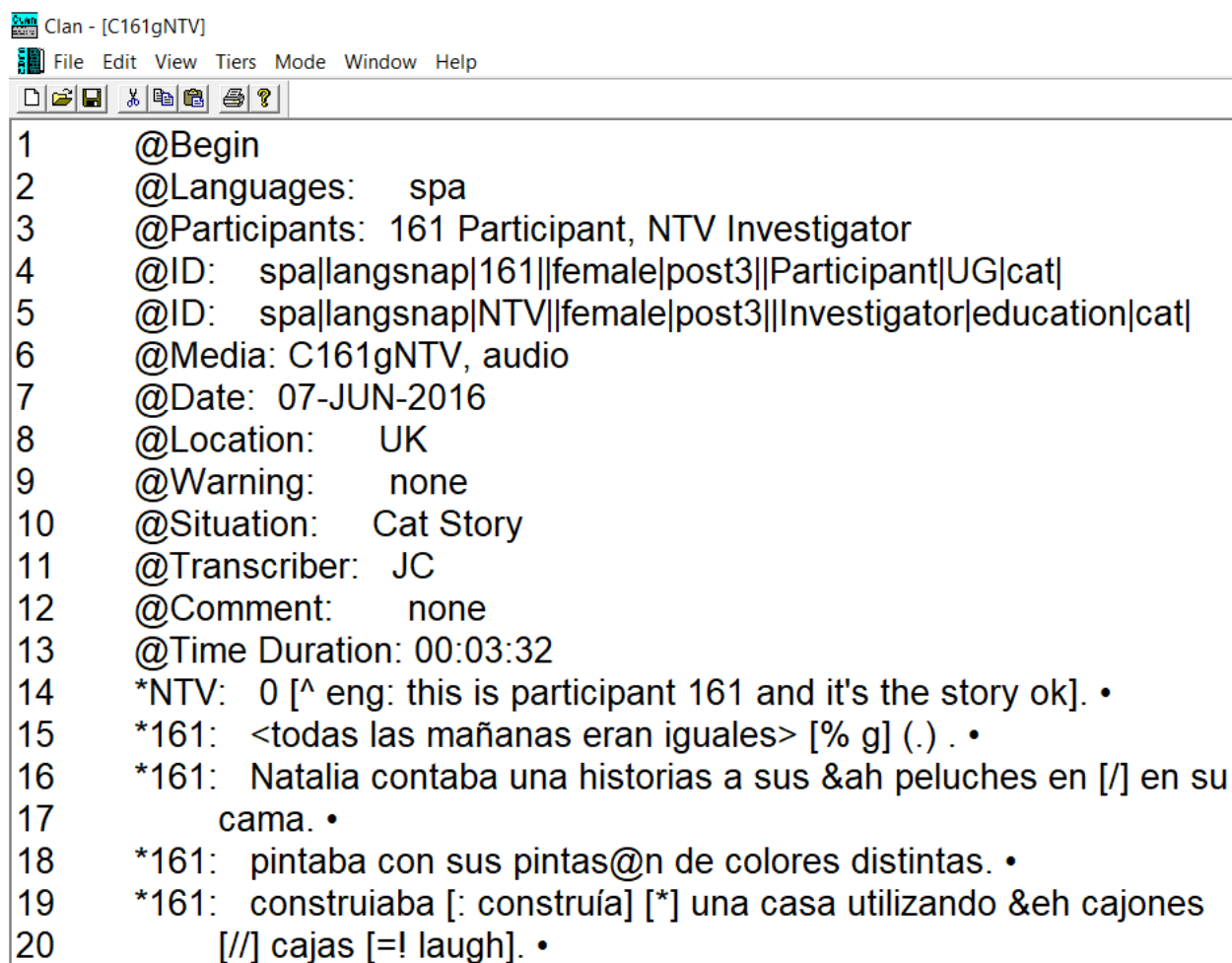represents an ASU produced by the participant.

Figure 1. Example start of a transcript

At this point we should acknowledge that coding ASUs can be complicated, in part due to the nature of spoken language. One aspect of ambiguity relates to the separating of coordinated verb phrases. Foster et al. address this issue with the following statement: "In cases where coordination of verb phrases occurs, the coordinated phrases will normally be considered to belong to the same AS-unit, unless the first phrase is marked by falling or rising intonation and is followed by a pause of at least 0.5 seconds" (p. 367). An example of this from our data comes from the following example in English:

She painted (0.8) and played with her friends on their bikes and went to the park.

 In this case, there is a pause of greater than 0.5 seconds after *painted*; however, coders found it difficult to agree upon the existence of clear falling or rising intonation. Given this, and the fact that Foster et al. characterize the ASU as "a mainly syntactic unit", we chose to include all coordinated verb phrases in the same ASU (p. 365).

Another potentially problematic aspect of coding ASUs is related to doing so crosslinguistically as Foster et al.'s discussion of ASUs is based on English data. Our data included French and Spanish (and English – see Huensch & Tracy-Ventura, 2017a). One major difference between French and Spanish is that Spanish is a pro-drop language (i.e., it is not necessary to include the subject pronoun overtly). Thus, we had to decide how to code instances without explicit subjects. Take the following two examples:

(1) French transcript

24	*106:	*et Pompon (.) il reste dans le cuisine.*
		'and Pompon he stays in the kitchen.'
25	*106:	*il joue dans le:s arbres.*
		'he plays in the trees.'
26	*106:	*il joue avec les [/] les petits <inse(ctes)> [/] insectes.*
		'he plays with the [/] the little <inse(cts)> [/] insects.'
27	*106:	*et il reste <dans le soleil> [//] en soleil.*
		'and he stays <in the sun> [//] in the sun.'

(2) Spanish transcript

22	*178:	*y Pancho dormía hasta muy tarde.*
		'and Pancho would sleep until very late.'
23	*178:	*salía de la casa.*
		'he would leave the house.'
24	*178:	*&ehm subía su árbol favorito.*
		'&ehm he would climb his favorite tree.'
25	*178:	*jugaba en el jardín con las mariposas.*
		'he would play in the garden with the butterflies.'
26	*178:	*y dormía.*
		'and he would sleep.'
27	*178:	*y tomaba el sol en la pared.*
		'and he would sunbathe on the wall.'

In the French example, each ASU includes a subject pronoun, whereas in the Spanish example, only the first line includes an explicit subject. Ultimately, we decided upon coding conventions that seemed to be most appropriate for the languages in question, but without guidance from published work with agreed upon conventions, it is unclear whether researchers coding these phenomena across multiple languages are making similar decisions. This issue becomes critical when analyses are based on this information such as complexity and accuracy measures that include the number of clauses per ASU. Therefore, it would be beneficial for researchers to be more explicit about how data were coded using examples like the ones above of decisions made.

Because our picture-based narratives included prompts that participants used in their narratives, we decided to mark those examples of prompted language with codes that would allow us to ignore them in future analyses. The same was true for uses of English or a language other than the target language. These codes for prompted language can be seen in line 15 in Figure 1, and for English in line 14. With an additional step, transcripts of the oral data can be linked at the end of each ASU to the corresponding time location in the audio file. The black dot at the end of each line is evidence of the linking. This linking allows any line to be played by clicking a button and has proven useful for a number of analyses such as checking the number of syllables produced. We have found it easiest to complete this step after the transcript is as accurate as possible.

The writing data were collected on a computer using a program that automatically saves each file as a text document (.txt) that can be imported into CLAN. Some revising of the file is then necessary to update the header and to address spelling issues, which we noted in the transcripts with comments. One important measurement in our written data is fluency, which requires knowing the amount of time spent writing. Our computer program keeps track of this

information so it can easily be input into the header. Additionally, the written data is segmented by t-units, and although the CLAN program used to import the .txt files attempts to do this automatically, these may need to be checked manually and occasionally amended, particularly in our case because the participants produced quite complex sentences.

In the end, the transcripts that we shared publicly are basic versions that include anything uttered by the participants with some coding in the oral data of features such as repetitions, reformulated language, and filled pauses. We also used the part-of-speech tagger that is freely available as part of CLAN, called MOR, to tag the transcripts and shared those tagged files as well. Tagging the files with MOR can be lengthy process if new items need to be added to the lexicon which happened often in our case. We believe that sharing more basic copies of the transcripts encourages other researchers to familiarize themselves with the corpus, to make the data their own, and to annotate it in ways that contribute to answering their specific research questions. On our personal copies of the transcripts, we added several additional layers of annotation that were specific to our research questions. We acknowledge that other researchers might disagree with the transcription decisions that were made in the transcripts, such as what language qualifies as a correction. We also acknowledge that there are likely errors in the transcripts. We tried to limit these by having multiple people check the transcripts against the audio files but some error is expected. Furthermore, throughout a longitudinal project it is likely that transcription conventions would evolve to some extent, and that has been the case with our project as well. For example, some features annotated in the publicly available LANGSNAP 3.0 transcripts were not annotated in the original LANGSNAP transcripts (e.g., the prompted language). While we would like it to be completely consistent, currently we do not have the resources to go back and change the older transcripts. Again, for this reason it is important that

researchers wanting to use the corpus become familiar with the transcripts and amend them for their own purposes.

We are already aware of other researchers making use of the LANGSNAP corpus and hope that the LANGSNAP 3.0 data (available at http://scholarcommons.usf.edu/langsnap/) can also become useful to other researchers as well. For example, Edmonds and Gudmestad (to appear) investigated gender marking in the LANGSNAP French written data and found that targetlike rates of use improved during residence abroad. They also investigated gender marking in the year following residence abroad and found continued improvement with feminine nouns but not masculine nouns. It would be interesting to see whether similar results are found in the LANGSNAP 3.0 data, or whether evidence of attrition may also be found for gender marking with feminine nouns.

Looking ahead to the future and LANGSNAP 6.0

Our plan is to continue this longitudinal study for as long as we have participants willing to take part. We are currently planning our next data collection wave for May 2019 (LANGSNAP 6.0) and are actively searching for grant agencies that we can target for funding. We will invite all the original LANGSNAP participants to take part in the next data collection wave, even those who were not able to participate in 3.0. Our goal is investigating the long-term evolution of their French and Spanish proficiency which we acknowledge will be different for many of our participants, and so our study might eventually evolve into a more qualitative longitudinal case-study like those reported in Kinginger and Blattner (2008) and Spenader (2008). Currently we have some participants who are living abroad again but we realize that might change and they could move back to the UK. Conversely, some of our participants

currently living in the UK may decide to move abroad again. These diverging experiences with respect to the L2 need to be taken into consideration and accounted for in our research design.

Additionally, with this kind of research it is important to maintain flexibility in many different ways. For example, although we prefer meeting participants face-to-face, if it becomes difficult to financially support visiting participants where it is convenient for them, we may have to resort to doing more Skype interviews. Some of the data from LANGSNAP 3.0 were collected this way, but the quality of the sound recordings was not as good. We felt it was worth it, however, and much of the data is useable.

Another way in which we may need to be flexible is in the instruments and tasks that we administer. This last time we decided to use the same instruments that were used in LANGSNAP, but over time we may change the tasks or create parallel versions to reduce boredom and other issues that could affect validity of the data (see Ortega & Iberri-Shea, 2005). One issue that has already come up relates to the prompts in the argumentative writing tasks. When LANGSNAP started in 2011, gay marriage and the legalization of marijuana were hotly debated topics in Europe and many other parts of the world. Now, times are rapidly changing. Therefore, we will likely need to adapt the prompts or use others. Pilot testing will be important to ensure that the prompts elicit the same kind of language and are comparable. We may also end up administering other kinds of experimental tasks in the future if our data and research questions suggest that they are important and necessary.

Conclusion

Longitudinal data is necessary for many of the important questions in SLA, one of which is the question of L2/FL attrition and the factors that influence it. A longitudinal research agenda is the most reliable way to examine the long-term evolution of FL proficiency which for our

particular group of participants includes the possibilities of attrition, maintenance, and development. Due to limited research in this area, little is known about the personal (e.g., age, motivation, proficiency) and external factors (e.g., amount of language use, disuse) that are involved and their significance for attrition/retention years later. In general, the research suggests that proficiency level attained might predict the amount of attrition/retention (Mehotcheva, 2010). Of course every language learner is exposed to different amounts and types of input following the instructional period which could explain why learners show considerable variation in maintenance/attrition and why it is very important to find ways to measure this kind of data in a project like LANGSNAP 3.0. By building off the initial longitudinal LANGSNAP study, which includes the point of peak attainment when the participants were arguably the most proficient, we will hopefully be able to document what happens to our participants' L2 abilities over time for many years to come. We have just completed transcription and initial analyses but so far it appears that attained proficiency at the end of their stay abroad and language contact/use post-university both play a role in maintenance of their French and Spanish, particularly in the areas of fluency, lexical complexity, and oral proficiency.

We hope that more researchers will decide to devote time to a longitudinal research agenda and commit to data sharing in the form of learner corpora that are formatted with agreed-upon systems for transcription and analysis. The CHILDES tools are already available and useful for the kinds of questions important in SLA. Additionally, they can be adapted in many different ways to address specific types of analyses. We are committed to making our longitudinal learner corpus publicly available and formatted in CHAT. We encourage other researchers to investigate additional research questions in this corpus, and to share their longitudinal corpora to broaden the impact of the data collected.

References

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). *CELEX*. Philadelphia: Linguistic Data Consortium.

Baba, K., & Nitta, R. (2014). Phase transitions in development of writing fluency from a complex dynamic systems perspective. *Language Learning*, *64*, 1–35. https://doi.org/10.1111/lang.12033

Bardovi-Harlig, K., & Stringer, D. (2010). Variables in second language attrition. *Studies in Second Language Acquisition*, *32*, 1–45. https://doi.org/10.1017/S0272263109990246

Biber, D., Conrad, S., & Reppen. R. (1998). *Corpus linguistics: Investigating structure and use*. Cambridge: Cambridge University Press.

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer (Version 5.4.18) [Computer program]. Accessed 1 Jan. 2017 at http://www.praat.org/

Crossley, S., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning*, *60*, 573–605. https://doi.org/10.1111/j.1467-9922.2010.00568.x

Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*, 163–185. https://doi.org/10.1111/lang.12000

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*, 359–380. https://doi.org/10.1093/applin/amm041

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*, 533–557. https://doi.org/10.1017/S0272263109990015

Domínguez, L., Tracy-Ventura, N., Arche, M. J., Mitchell, R., & Myles, F. (2013). The role of dynamic contrasts in the L2 acquisition of Spanish past tense morphology. *Bilingualism: Language and Cognition*, *16*, 558–577. https://doi.org/10.1017/S1366728912000363

Edmonds, A., & Gudmestad, A. (to appear). Gender marking in written L2 French: Before, during, and after residence abroad. *Study Abroad Research in Second Language Acquisition and International Education*. https://doi.org/10.1075/sar.16018.edm

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21,* 354–375. https://doi.org/10.1093/applin/21.3.354

Harklau, L. (2008). Developing qualitative longitudinal case studies of advanced language learners. In L. Ortega, & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 23–35). New York: Routledge.

Howard, M. (2012). The advanced learner's sociolinguistic profile: On issues of individual differences, second language exposure conditions, and type of sociolinguistic variable. *The Modern Language Journal, 96,* 20–33. http://doi.org/10.1111/j.1540-4781.2012.01293.x

Huebner, T. (1983). *A longitudinal analysis of the acquisition of English*. Ann Arbor: Karoma.

Huensch, A., & Tracy-Ventura, N. (2017a). Understanding second language fluency behavior: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics*, *38,* 755–785. https://doi.org/10.1017/S0142716416000424

Huensch, A., & Tracy-Ventura, N. (2017b). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *The Modern Language Journal, 101,* 275–293. https://doi.org/10.1111/modl.12395

Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics*, *24*, 131–161. https://doi.org/10.1017/S0142716403000079

Keijzer, M. (2007). *Last in first out? An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada* (Unpublished doctoral dissertation). Vrije Universiteit Amsterdam, The Netherlands.

Kinginger, C., & Blattner, G. (2008). Histories of engagement and sociolinguistic awareness in study abroad: Colloquial French. In L. Ortega, & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 223–246). New York: Routledge.

Larsen-Freeman, D., & Cameron, L. (2008). Research methodology on language development from a complex systems perspective. *The Modern Language Journal*, *92*, 200–213. https://doi.org/10.1111/j.1540-4781.2008.00714.x

Leopold, W. F. (1939). *Speech development of a bilingual child: A linguist's record.* Evanston, IL: Northwestern University.

MacWhinney, B. (2000). *The Childes project: Tools for analyzing talk. Transcription format and programs*. Psychology Press.

MacWhinney, B. (2017). A shared platform for studying second language acquisition. *Language Learning*, *67*, 254–275. https://doi.org/10.1111/lang.12220

Maslen, R. J., Theakston, A. L., Lieven, E. V., & Tomasello, M. (2004). A dense corpus study of past tense and plural overregularization in English. *Journal of Speech, Language, and Hearing Research*, *47*, 1319–1333. https://doi.org/10.1044/1092-4388(2004/099)

McManus, K., Tracy-Ventura, N., & Mitchell, R. (September, 2016). *L2 linguistic development before, during, and after a nine-month sojourn: Evidence from L2 French and Spanish*. Paper presented at the Second Language Research Forum (SLRF) annual conference, New York, NY.

McManus, K., Mitchell, R., & Tracy-Ventura, N. (2014). Understanding insertion and integration in a study abroad context: The case of English-speaking sojourners in France. *Revue Française de Linguistique Appliquée*, *XIX*(2), 97–116.

Meara, P., & Milton, J. (2003). *X_lex: The Swansea levels test*. Newbury: Express.

Mehotcheva, T. H. (2010). *After the fiesta is over: Foreign language attrition of Spanish in Dutch and German Erasmus students* (Unpublished doctoral dissertation). University of Groningen, The Netherlands.

Meunier, F. (2015) Developmental patterns in learner corpora. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 379–400). Cambridge: Cambridge University Press.

Meunier, F., & Littre, D. (2013). Tracking learners' progress: Adopting a dual "corpus cum experimental data" approach. *The Modern Language Journal*, *97*, 61–76. https://doi.org/10.1111/j.1540-4781.2012.01424.x

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, *3*, 235–244. https://doi.org/10.1093/ijl/3.4.235

Mitchell, R., Domínguez, L., Arche, M. J., Myles, F., & Marsden, E. (2008). SPLLOC: A new database for Spanish second language acquisition research. *EuroSLA yearbook*, *8*, 287-304. https://doi.org/10.1075/eurosla.8.15smit

Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). *Anglophone students abroad: Identity, social relationships, and language learning*. New York: Routledge.

Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, *58*, 479–502. https://doi.org/10.1111/j.1467-9922.2008.00448.x

Myles, F. (2005). Interlanguage corpora and second language acquisition research. *Second Language Research*, *21*, 373–391. https://doi.org/10.1191/0267658305sr252oa

Myles, F. (2008). Investigating learner language development with electronic longitudinal corpora: Theoretical and methodological issues. In L. Ortega, & H. Byrnes, *The longitudinal study of advanced L2 capacities* (pp. 58–72). New York: Routledge.

Myles, F. (2015). Second language acquisition theory and learner corpus research. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 309–332). Cambridge: Cambridge University Press.

Myles, F., & Mitchell, R. (2004). Using information technology to support empirical SLA research. *Journal of Applied Linguistics*, *1*, 69–98. https://doi.org/10.1558/japl.v1i2.169

Ortega, L., & Byrnes, H. (2008). *The longitudinal study of advanced L2 capacities*. New York: Routledge.

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, *25*, 26–45. https://doi.org/10.1017/S0267190505000024

Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002, October). *An investigation of elicited imitation tasks in crosslinguistic SLA research*. Paper presented at the Second Language Research Forum (SLRF) annual conference, Toronto, Canada.

Perdue, C. (Ed.). (1993). *Adult language acquisition: Cross-linguistic perspectives*. Cambridge: Cambridge University Press.

Polat, B., & Kim, Y. (2014). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics, 35,* 184–207. https://doi.org/10.1093/applin/amt013

Rees, J., & Klapper, J. (2008). Issues in the quantitative longitudinal measurement of second language progress in the study abroad context. In L. Ortega, & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 89–106). New York: Routledge.

Schmid, M. S., & Mehotcheva, T. (2012). Foreign language attrition. *Dutch Journal of Applied Linguistics*, *1*, 102–124. https://doi.org/10.1075/dujal.1.1.08sch

Schumann, J. (1978). Second language acquisition: The pidginization hypothesis. In E. Hatch (Ed.), *Second language acquisition: A book of readings* (pp. 256–271). Rowley, MA: Newbury House.

Skiba, R., Dittmar, N., & Bressem, J. (2008). Planning, collecting, exploring, and archiving longitudinal L2 data: Experiences from the P-MoLL project. In L. Ortega, & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 73–88). New York: Routledge.

Spenader, A. (2008). Acquiring oral language skills over the course of a high school year abroad: What's in it for absolute beginners. In L. Ortega, & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 247–263). New York: Routledge.

Tracy-Ventura, N., McManus, K., Norris, J. M., & Ortega, L. (2014). "Repeat as much as you can": Elicited imitation as a measure of oral proficiency in L2 French. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 143–166). Bristol: Multilingual Matters.

Tracy-Ventura, N., Mitchell, R., & McManus, K. (2016). The LANGSNAP longitudinal learner corpus: Design and use. In M. Alonso Ramos (Ed.), *Spanish learner corpus research: State of the art*. Amsterdam: John Benjamins Publishing.

van der Zee, K. I., & van Oudenhoven, J. P. (2000). The multicultural personality questionnaire: A multidimensional instrument of multicultural effectiveness. *European Journal of Personality*, *14*, 291–309. https://doi.org/10.1002/1099-0984(200007/08)14:4<291::AID-PER377>3.0.CO;2-6

Vyatkina, N. (2013). Specific syntactic complexity: Developmental profiling of individuals based on an annotated learner corpus. *The Modern Language Journal*, *97*, 11–30. https://doi.org/10.1111/j.1540-4781.2012.01421.x

Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

Young, C. H., Savola, K. L., & Phelps, E. (1991). *Inventory of longitudinal studies in the social sciences*. Newbury Park, CA: Sage.

Young-Scholten, M., & Langer, M. (2015). The role of orthographic input in second language German: Evidence from naturalistic adult learners' production. *Applied Psycholinguistics*, *36*, 93–114. https://doi.org/10.1017/S0142716414000447

Yuldashev, A., Fernandez, J., & Thorne, S. L. (2013). Second language learners' contiguous and discontiguous multi-word unit use over time. *The Modern Language Journal*, *97*, 31–45. https://doi.org/10.1111/j.1540-4781.2012.01420.x

Zheng, Y. (2012). Exploring long-term productive vocabulary development in an EFL context:

The role of motivation. *System*, *40*, 104–119.

https://doi.org/10.1016/j.system.2012.01.007