

Fluency

Amanda Huensch

ORCID: <https://orcid.org/0000-0001-6229-3521>

Email Address: amanda.huensch@pitt.edu

Biographical Note

Amanda Huensch is Assistant Professor in the Department of Linguistics at the University of Pittsburgh. Her research examines the development of second language fluency, the acquisition of second language phonology, and the pronunciation pedagogy practices of foreign language instructors.

Abstract

This chapter begins with an overview of the ways in which fluency has been conceptualized and defined in second language (L2) research. It then identifies the key questions that have been investigated in research on L2 oral fluency. The research methods and tools section highlights common speaking tasks used to collect data in learner corpus research investigating L2 fluency as well as measurements of fluency and the software and tools used in their transcription, coding, and analysis. Three representative corpora are then introduced with a summary of the L2 fluency research conducted using them before the chapter concludes with suggestions of directions for future research.

Introduction

As has been pointed out on many occasions in work investigating oral fluency, the term fluency has multiple and varied definitions. Thus, it is important to begin by specifying how fluency is conceptualized in second language (L2) research. The first consideration is to differentiate between a “broad” view of fluency, which it is equated with general proficiency, and a “narrow” view of fluency in which fluency refers to the pace, flow, and tempo of a learner’s speech (Lennon, 1990, p. 389). It is the narrow definition of fluency that is the topic of inquiry in much second language acquisition (SLA) research. Lennon also focused on fluency as a perceptual phenomenon in claiming that “fluency is an impression on the listener’s part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently” (p. 391). Thus, fluency is not merely a concept that can be boiled down to measuring how many words a learner can utter between hesitations, but rather it represents a complex relationship between processes occurring in a learner’s planning and production of speech, characteristics of how that speech is uttered, and interpretations of how that speech is perceived by a listener. Segalowitz (2010) labelled these three senses of fluency as (a) cognitive fluency, (b) utterance fluency, and (c) perceived fluency. Cognitive fluency “refers to the fluid operation (speed, efficiency) of the cognitive processes responsible for performing L2 speech acts” (Segalowitz, 2016, p. 82) whereas utterance fluency refers to observable characteristics of the speech signal (e.g., features related to pausing, pace, hesitation), and perceived fluency refers to judgements made by listeners on the basis of utterance fluency features. Derwing (2017) summarized the relationship between these senses of fluency by stating that “cognitive fluency underlies utterance fluency, which affects listeners’ perception of fluency” (p. 250).

Identifying cognitive fluency as a learner's ability to efficiently plan and produce speech necessitates considerations of how speech production occurs. One model that has been highly influential in L2 research is Levelt (1999)'s model of speech production, which is made up of the stages of *Conceptualization* (i.e., message planning), *Formulation* (i.e., lexical, grammatical, phonological encoding), and *Articulation* (i.e., conversion into speech), along with self-monitoring. Segalowitz (2010) drew from the work of Levelt (1989, 1999), de Bot (1992), and Kormos (2006) to construct a "blueprint" of an L2 speaker including "fluency vulnerability points" which represent locations in the model where an L2 speaker might encounter different types of processing difficulties and therefore might exhibit disfluency (Segalowitz, 2010, p. 17). Measurements of utterance fluency have been used as a way to draw conclusions about successes/difficulties at different points in the speech production process (e.g., Skehan, Foster, Shum, 2016; Towell, Hawkins, Bazergui, 1996). The advent of learner corpus research (LCR) in the 1980s has resulted in an impressive and growing amount and variety of learner language data available to L2 acquisition researchers (Gilquin & Granger, 2015), including those that allow for investigations of L2 fluency. As recently argued in a special issue of the *International Journal of Learner Corpus Research* "the study of fluency and disfluency in L2 versus L1 speech with the help of corpora and tools for visualization allows a better and wider understanding of the phonetic mechanisms of cognitive processes involved in L2 discourse" (Trouvain et al., 2017, p. 111).

With these definitions in mind, we now turn to core issues and topics in L2 fluency research that have been investigated with learner corpora. However, as a final note, it is useful to clarify how 'learner corpora' and 'corpus-based techniques' are conceived of in this chapter as not all of the studies referenced self-identify as using corpora. A broad definition of these terms is adopted; thus,

the work focused on in this chapter is that which in some way has automated analyses of a collection of texts in its investigation of L2 fluency.

Core Issues and Topics

Native Speaker vs. Non-Native Speaker Fluency

One strand of research in which L2 fluency has been investigated with learner corpora explores the extent to which the utterance fluency of learners differs from that of native speakers (NS). While both learners and native speakers are expected to show signs of disfluency when speaking, lack of automaticity and limited linguistic knowledge may differentiate learner from native speaker speech (Kormos, 2006). As a general approach, work in this area compares the fluency characteristics of native and non-native speaker (NNS) speech from talkers who have completed identical or similar tasks (e.g., Belz, Sauer, Lüdeling, & Mooshammer, 2017; De Jong, 2016; Foster & Tavakoli, 2009; Götz, 2013; Gut, 2009; Huensch & Tracy-Ventura, 2017a; Kahng, 2014; Skehan et al., 2016; Tavakoli, 2011). Corpus-based comparisons of native speaker and learner speech have provided some evidence that, for example, learners demonstrate greater fluency variability across speaking tasks (e.g., read speech vs narrative retelling) than native speakers (Gut, 2009).

De Jong (2016) compared the speech of L1 and L2 speakers of Dutch who completed a variety of monologic speaking tasks (e.g., describe a crime you just witnessed to a police officer). Data, transcribed in the Computerized Language Analysis program (CLAN, MacWhinney, 2000), were explored with regard to the frequency, location, and duration of silent and filled (e.g., *um*, *uh*) pauses. Based on previous research (e.g., Davies, 2003), De Jong was particularly interested in

exploring the extent to which learner and native speaker pause behavior differed within versus between utterances (coded as Analysis of Speech Units [ASU], Foster, Tonkyn, Wigglesworth, 2000). Results indicated that at utterance boundaries, learners and native speakers did not significantly differ in their likelihood to pause nor in the length of their pauses. In contrast, within utterances, learners paused both more often and for longer durations than native speakers. Based on these findings, De Jong argued that the within-utterance pausing behavior of L2 speakers, as opposed to between-utterance, is reflective of trouble with *Formulation*, possibly due to limited L2 knowledge and skills.

L2 Fluency Development

A second focus of L2 fluency research is to better understand how L2 fluency develops over time as proficiency increases. Some corpora designed to answer these questions are structured cross-sectionally, such as the *Parallèle Oral en Langue Etrangère* ‘Parallel Oral Foreign Language’ (PAROLE) corpus (Hilton et al., 2008), designed to investigate learner language (L2 English, French, and Italian) at different proficiency levels and the *What is Speaking Proficiency* (WiSP) corpus (De Jong et al., 2012), which includes English and Turkish L1 learners of L2 Dutch. Both corpora gathered further information about learners that would be necessary in understanding their proficiency. For example, in PAROLE, learners completed a variety of other tasks to gain information about their language learning motivation, aptitude, experience, etc. In WiSP, learners completed a productive vocabulary task designed to be a separate measure of proficiency from the speaking tasks. With these data, the researchers were able to test whether learners at different proficiency levels exhibited different oral fluency characteristics.

In addition to cross-sectional corpora, longitudinal corpora have also been used to investigate L2 fluency development (e.g., the *Learning Prosody in a Foreign Language* [LeaP] corpus; Gut, 2009, 2017; the *Languages and Social Networks Abroad Project* [LANGSNAP] corpus, Huensch & Tracy-Ventura, 2017a). Some SLA researchers have argued that longitudinal data like these are critical for investigations of development because they allow explorations of how learning occurs over time (Myles, 2008; Ortega & Byrnes, 2008; Ortega & Iberri-Shea, 2005). Meunier & Littre (2013) reasoned that the use of longitudinal learner corpora “can enable researchers interested in L2 acquisition to test hypotheses on larger and better constructed databases, using the options offered by computer-based annotation and analysis corpus tools” (p. 72). Gut (2017) used a subcorpus of the LeaP corpus to investigate phonological development (including L2 fluency) over time for learners in three different contexts (i.e., study abroad, study abroad with phonology course, at-home phonology course). Of particular relevance to the current discussion is that while the corpus was originally designed to study phonological acquisition, it was not compiled specifically to explore the effects of learning context. Gut discussed the advantages and challenges of using corpora in this way (i.e., not for their originally intended purpose). On the one hand, using the corpus meant having missing data points and heterogeneous and unbalanced groups. Nevertheless, using a previously annotated corpus not only saved time, but perhaps more importantly allowed for the simultaneous investigation of a wide variety of phonological features (in addition to fluency) and both quantitative and qualitative data analyses.

L1-L2 Fluency Relationships

Beyond proficiency level and native speaker status, many other factors are likely to contribute to variation in an individual’s L2 utterance fluency and indeed have been the focus of research on L2

fluency. An important consideration in L2 fluency research is to differentiate L2-specific cognitive fluency from more general cognitive processing including that which regulates the L1 (Segalowitz, 2016). Factors such as speaking task, topic familiarity, planning time, first language, time spent in an immersion context, etc. are all likely to play a role in observed variation. One final area of L2 fluency research that has recently been of interest to scholars is the extent to which one's L1 speaking style relates to fluency characteristics in the same individual's L2 speech (e.g., De Jong et al., 2015; De Jong & Mora, 2017; Derwing et al., 2009; Garcia Lecumberri et al., 2017; Gósy, Gyarmathy, & Beke, 2017; Huensch & Tracy-Ventura, 2017b). In other words, this work attempts to tease apart the potential contribution of L1 speaking style in explaining L2 utterance fluency. In order to do so, studies have compared the speech of the same individuals in both their L1 and L2 on the same or similar tasks.

Gósy et al. (2017), for example, examined the frequency, form, location, and formant structure of filled pauses using the Hungarian English Database (HunEng-D) corpus, which is comprised of L1 Hungarian and L2 English speech from the same speakers who vary according to age and proficiency level. They found that while filled pauses were shorter in the L1 than L2, their form, location, and articulation were similar, demonstrating transfer of Hungarian filled pausing characteristics into the L2. In De Jong et al. (2015), L1 speakers of English and Turkish who were learning Dutch as a L2 (a subset from the WiSP corpus) completed similar monologic speaking tasks in both languages. Analyses indicated that all seven measures of fluency in their study (including those representing speed, breakdown, and repair fluency, see the Measurements of Oral Fluency section) were correlated between L1-L2 fluency ranging in strength from 0.37–0.76. Additionally, they conducted regression analyses to test whether L2 measures of fluency

‘corrected’ for L1 fluency (i.e., using the saved residuals from models predicting L2 fluency from L1 fluency) would better predict L2 proficiency as measured by a productive vocabulary task. The results indicated that for one of the seven measurements of fluency, mean syllable duration, the corrected measure was a stronger predictor of proficiency.

With these core issues and topics in mind, we next turn to the main research methods and tools that have been used in learner corpus research to investigate L2 fluency, including common types of speech data, utterance fluency measurements, and the software and tools used for data transcription, coding, and analysis.

Main Research Methods and Tools

Types of Speech Data

As mentioned previously, an important aspect of corpora designed to investigate SLA topics and issues is that they often include a variety of additional information about the participants’ proficiency, motivation, age, gender, language learning history, etc. Regarding the speech data itself, investigations of L2 fluency have been conducted with a variety of tasks ranging from tightly controlled passage reading to less controlled spontaneous speech tasks such as semi-structured interviews. Decisions made about which speaking tasks to include are often connected to the original purpose of compiling the corpus. For example, given its focus on the acquisition of phonology (and not only the development of L2 fluency), the LeaP corpus (Gut, 2009) included a word-list reading task to explore the acquisition of stress (in addition to three other tasks: an interview, a reading passage, and a story retelling). In order to investigate disfluency in dialogic speech Belz et al. (2017) used the Berlin Map Task Corpus (BeMaTaC, Sauer & Lüdeling, 2016)

in which participants instructed their partners (who could not see them) to recreate a route on a map that contained landmarks. Many oral corpora include multiple types of speech data, for instance in the PAROLE corpus (Hilton et al., 2008), learners and NSs completed the same three tasks: two narrative retellings based on videos and an autobiographical narrative describing an accident that had occurred in the past. Similarly, the HunEng-D corpus (Gósy et al., 2012) included responses to interview questions (e.g., give your opinion about a topic of current interest), retelling a story, completing a map-task/role-play with another learner, and a word-list reading. The use of a variety of speaking styles allows researchers to additionally explore the extent to which L2 fluency varies across tasks.

Measurements of Oral Fluency

When conducting a scan of the L2 fluency literature, it becomes quickly apparent that measures of oral fluency are varied and diverse. Some examples of tables listing measures used can be found in Kormos (2006, p. 163) and Derwing, (2017, p. 247). Skehan (2003) and Tavakoli and Skehan (2005) categorized commonly used measures as representing three types of fluency: speed, breakdown, and repair. Speed fluency represents dimensions of pace and includes measures such as speech rate (often words/syllables/characters per minute/second). Breakdown fluency relates to pausing phenomena and includes measures such as number of silent pauses per minute/X number of words. Breakdown fluency measures can be further categorized into those that provide information about the location, duration, and frequency of pauses. For example, the measure of mean silent pause duration within clauses provides information about both length and location whereas the measure of the number of filled pauses per 100 words provides information about frequency alone. As discussed in the Core Issues and Topics section, recent investigations of L1-

L2 and NS-NNS fluency provide compelling evidence that pause location is an important consideration when measuring aspects of breakdown fluency (see also Hilton, 2008; Skehan et al., 2016). Thus, when manually coding pauses in corpora,¹ it appears that including information about their location is particularly important. Finally, repair fluency is concerned with self-correction and reformulation and therefore includes measures such as the number of repetitions per minute/X number of words, the number of corrections per minute/X number of words, among others.

It is relevant to acknowledge the somewhat large number of possible measurements that can and have been used to investigate L2 fluency. While not specifically focused on LCR, a recent scoping review of fluency literature from the field of study abroad (SA) concluded “that oral fluency has been investigated with little methodological consistency in SA research (Tullock & Ortega, 2017, p. 13). Given this, it is relevant to identify the source of some of the inconsistency as well as how researchers justify their choices of utterance fluency measurements. Taking a closer look at the measures of breakdown fluency, which include silent pausing, a typical difference across studies relates to the threshold set for considering what should be coded as a silent pause. Durations in the fluency literature can range from 100ms (Riazzantseva, 2001) to 1000ms (Götz, 2013), with many set at 250ms or 400ms. De Jong & Bosker (2013) attempted to provide empirical evidence for an optimal cut-off point for silent pauses. They calculated measures of breakdown fluency with lower bound cut-offs ranging from 20ms to 1000ms and conducted Pearson correlations between those measurements and a measure of L2 proficiency based on vocabulary knowledge. The results indicated “that a lower cut-off point for silent pauses of 250–300ms leads to the highest correlation between the number of silent pauses and a measure of L2 proficiency (vocabulary knowledge)” (p. 20). Segalowitz (2016) argued that De Jong and Bosker’s approach of justifying their choice

of operationalization of silent pauses based on a “cognitive measure of L2 proficiency” is an important step in “the discussion of how utterance fluency reflects cognition” (p. 82).

One approach used to justify choices of utterance fluency measurements is to consider those measurements which best predict ratings of perceived fluency (e.g., Bosker et al., 2013; Kahng, 2018; Kormos & Dénes, 2004) and/or whether there are intercorrelations among measurements. Bosker et al. (2013) examined the extent to which measures of speed, breakdown, and repair fluency could predict fluency ratings (from 20 untrained raters) and demonstrated via linear regression analyses that measurements of speed and breakdown fluency best predicted the ratings, although measurements of repair fluency also contributed to the models but less so.

A final issue that arises in the measurement of oral fluency relates to potential cross-linguistic differences when comparing fluency across languages. Aspects of a language such as syllable inventories or morphological processes might contribute to differences in measurements of speed fluency such as mean syllable duration and mean length of run (Gut, 2009, p. 96). When comparing languages such as English and Spanish or German and French, the syllable inventories of one language (English and German in these cases) are such that it is possible that the number of phones within a syllable will be greater in those languages as opposed to the comparisons (Spanish and French). Studies have provided some indication that speech rates using these measures show slower rates for languages with greater syllable inventories (e.g., Pellegrino et al. 2011, Huensch & Tracy-Ventura, 2017b comparing English and Spanish; Trouvain & Mobius, 2014 comparing French and German). Garcia Lecumberri et al. (2017) addressed this issue by normalizing speech rate across speakers by taking into consideration average rates from native speakers. The effects

of phonotactics on speed fluency measurements are not the only cross-linguistic differences indicated by the literature. There is also evidence that cross-linguistic differences in pausing characteristics may also be present (see e.g., Riazantseva, 2001).

Software and Tools for Data Coding and Analysis

One advantage of using corpora or corpus-based techniques to investigate L2 fluency is that they allow for (at least partially) automated analysis of a large amount of data. However, it is often necessary to manually transcribe and code the data in order to conduct automated analyses that either output information such as the frequency, duration, and location of phenomena or simply calculate measures of oral fluency. Manually transcribing data and coding for fluency features is likely time-consuming and therefore also quite expensive (Ballier & Martin, 2013; Staples, 2015). This is partly why many scholars have argued for the public sharing of corpora that have been formatted with agreed-upon conventions (e.g., MacWhinney, 2017; Myles, 2008). Hilton (2009) provided a detailed description of the manual transcription and coding of the PAROLE corpus using the CLAN program and following transcription conventions in the format of the Codes for the Human Analysis of Transcripts (CHAT) as well as a description of some of the automated analyses they were able to conduct with such coding. For example, CLAN includes commands that will automatically count the frequency of repetitions coded with the symbol [/] (e.g., in [/] in the summer) as well as commands such as MLU which calculates the mean length of utterances. Transcripts following CHAT conventions in CLAN provide impressive interoperability with other programs (MacWhinney, 2017) commonly used for annotating aspects of fluency such as EUDICO Linguistic Annotator (ELAN, <https://tla.mpi.nl/tools/tla-tools/elan/>) and Praat (Boersma & Weenink, 2015). While software like ELAN and Praat have advantages for annotating, programs

such as CLAN or Annotation of Information Structure (ANNIS, Krause & Zeldes, 2016) are perhaps better suited for analysis as they are designed for more extensive queries. One of the goals of Hilton (2009) in providing such a detailed description of their manual transcription and coding was to inform and contribute to future automatization. For interested readers, Ballier and Martin (2013, 2015) provide useful summaries and comparisons of software that have been used in the annotation of spoken learner corpora.

Praat is another commonly used program in fluency analyses (see e.g., Gósy et al., 2017; Tracy-Ventura & Huensch, 2018), especially those that include the investigation of additional phonological phenomena such as intonation and vowel quality (e.g., Garcia Lecumberri et al., 2017; Gut, 2012). Praat includes built-in features for automating some of the coding often necessary for fluency analysis. For example, the *Annotate To TextGrid (silences)*... command automatically segments the sound file into silent and sounding segments (one can customize the length of the silences among other things). Of course, the program cannot differentiate between speech and noise of another form (e.g., laughter, filled pauses, a door slamming), so depending on the recording quality and the amount of background noise in the file, manual checking is necessary. Once the TextGrid is segmented, Praat scripts can be written to quickly and efficiently output data for simple analysis such as the number and duration of the pauses, etc. Additionally, De Jong and Wempe (2009) reported on a Praat script they developed to detect syllable nuclei which can be used to automatically count syllables. In conjunction with the automatic identification of silences, speech rate can be automatically calculated. De Jong and Wempe compared automated and manual coding and demonstrated high correlation ($r=0.8$) for a subset of the data. The differences found between manual and automatic coding were mainly the result of the script not identifying some of

the unstressed syllables that were coded manually. Hilton (2009) additionally pointed out that the script does not differentiate between speech and filled pauses, which for some of the participants in the PAROLE corpus would lead to an overestimation of speech rate. Thus, researchers using automated scripts for syllable counting are recommended to test the accuracy against manual coding with a subset of data.

Gut (2012) presented a detailed description of the transcription and coding of the LeaP corpus using Praat and additionally discussed the issue of interrater reliability for the different types of annotations completed. Perhaps not surprisingly, she reported that those annotations which were the most complex resulted in the lowest agreement. For example, one process of coding only required annotators to indicate whether something was a consonant, vowel, or pause. At this level, agreement was near perfect ($\kappa=0.99$). Another process of coding showed much lower agreement ($\kappa=0.23$) when annotators were required to first segment speech into syllables and then provide phonetic transcription (Gut, 2012, p. 11).

Representative Corpora and Research

In this section, three corpora are described along with some of the investigations of fluency that have been conducted using them. The corpora are the *Louvain International Database of Spoken English Interlanguage* (LINDSEI; Gilquin, De Cock, & Granger, 2010), the LANGSNAP corpus (Mitchell, Tracy-Ventura, & McManus, 2017), and the *Diapix Foreign Language Corpus* (DiapixFL, Garcia Lecumberri et al., 2017). These three corpora were chosen because their data are available to researchers (either freely or for a fee), they represent a variety of languages and

speaking tasks, and they were compiled to investigate different combinations of the core issues and topics presented earlier in this chapter.

LINDSEI

The LINDSEI corpus (Gilquin, De Cock, & Granger, 2010) is a collection of speech from advanced EFL learners from 11 different L1s (Bulgarian, Chinese, Dutch, French, German, Greek, Italian, Japanese, Polish, Spanish, Swedish), with additional L1s continuing to be added. LINDSEI was designed to be an oral counterpart to the *International Corpus of Learner English* (ICLE) corpus (Granger, 1998), which is a written corpus of argumentative essays. Each of the LINDSEI subcorpora were constructed following the same guidelines for comparison purposes. For each L1, the participants included approximately 50 university students typically in the third or fourth year of their studies. Speech data were collected from interviews comprised of three parts: a warm-up during which speakers completed a monologic task in which they spoke about a given topic, an informal interview (dialogic) in which they answered questions about their lives at university, hobbies, etc., and a picture description task. The *Louvain Corpus of Native English Conversation* (LOCNEC, De Cock, 2004) is a native speaker corpus of British university students that was compiled to allow for comparison with LINDSEI. The LINDSEI corpus (not including sound files) and handbook are available to the research community and require purchase. One of the main advantages of the LINDSEI corpus and its NS counterpart LOCNEC is that they were designed to be maximally similar to allow for both NS-NNS and cross-linguistic comparison.

LINDSEI has been used for an impressive number of investigations (see <https://uclouvain.be/en/research-institutes/ilc/cecl/lindsei-bibliography.html>), including those

focused on fluency (e.g., Brand & Götz, 2011; Götz, 2013; Quan & Weisser, 2015). Here, I focus on the work that has used the German subcorpus of LINDSEI (LINDSEI-GE) as well as the LOCNEC to investigate L2 fluency (e.g., Brand & Götz, 2011; Götz (2013). Brand & Götz conducted quantitative analyses of accuracy and fluency with the 50 German L1 learners and the NSs from the LOCNEC corpus. This analysis was supplemented by qualitative analysis with a subset of five speakers chosen based on their varying accuracy/fluency profiles (e.g., the least accurate learner, the most fluent learner). Finally, the speech samples from the qualitative analysis were used as stimuli in a perceived fluency experiment in which 50 NSs of English rated how proficient they thought the speakers were. One interesting finding from this study is that across both quantitative and qualitative analyses, much individual variation was found with the fluency variables whereas the same was not true for accuracy. With the same corpora, Götz (2013) provided a thorough comparative analysis of native and non-native fluency characteristics including the less frequently investigated discourse markers (e.g., *like*, *well*) and small words (e.g., *sort of*, *kind of*). Her analysis demonstrated that, in comparison to NSs, learners showed less variation in their use of both discourse markers and small words, and they typically repeated the same ones instead of varying them.

LANGSNAP

The LANGSNAP corpus (Mitchell, Tracy-Ventura, McManus, 2017) is the result of a 2-year longitudinal project investigating language development before, during, and after study/residence abroad. Participants were L2 learners of French ($n=29$) and Spanish ($n=27$) who were university language majors required to spend the third year of their four-year undergraduate program abroad. Data were collected at six times between May 2011 and February 2013: once before, three times

during, and two times after a nine-month stay abroad, and included a variety of tasks such as picture-based oral narratives, semi-structured interviews about the participants' experiences, daily lives and future plans, a measure of proficiency (elicited imitation test), etc. NSs of Spanish ($n=10$) and French ($n=10$) also completed the narrative and interview tasks for comparison purposes. The oral production data were transcribed in CLAN following CHAT conventions and are freely available with the audio files at <http://langsnap.soton.ac.uk/>.

Data in the LANGSNAP corpus have been used for several investigations of oral fluency including explorations of L1-L2 fluency relationships (Huensch & Tracy-Ventura, 2017b) and tracking development longitudinally (Huensch & Tracy-Ventura, 2017a). Huensch & Tracy-Ventura (2017a) used the Spanish subset of the LANGSNAP corpus to explore the development and retention of nine measures of speed, breakdown, and repair fluency. The results indicated differential trends in the development and maintenance of different measures of fluency which led the authors to argue that measures reflect different sub-dimensions of fluency. In 2016, three years after the final data collection wave in LANGSNAP, participants were invited to take part in a new round of data collection. Approximately 60% ($n=33$) agreed to participate (these data are available at <http://scholarcommons.usf.edu/langsnap/>). Huensch et al. (2019) investigated the possible outcomes of attrition/development/maintenance of L2 fluency three years after the end of formal instruction and explored the extent to which variables such as proficiency at the end of residence abroad and language exposure could predict changes in fluency. While previous research investigating first language attrition had not indicated reduced language exposure as a strong predictor of attrition (Mehotcheva, 2010), results from Huensch et al. with instructed learners indicated that the maintenance of some aspects of speed and breakdown fluency (e.g., speech rate

and silent pauses) were influenced by language exposure and not by proficiency at the end of study abroad. The LANGSNAP studies demonstrate just some of the possibilities for investigating L2 fluency with a longitudinal corpus.

DiapixFL

The *DiapixFL* corpus is a bi-directional corpus designed to allow for investigations that consider both individual differences in speaking style as well as potential cross-linguistic differences between the speaker's L1 and L2. Speakers in the corpus ($n=24$) include two groups: Spanish L1 learners of L2 English and English L1 learners of L2 Spanish. Both groups completed tasks in their L1 and L2. The task was a dialogic spot-the-difference task adapted from the DiapixUK materials (Baker & Hazan, 2011). In this task, participants were each presented with a picture which differed from their counterparts, and they worked together to identify the differences. Data were transcribed and annotated in *Mtrans* (Villegas et al. 2011) and are freely available at <https://datashare.is.ed.ac.uk/handle/10283/346>.

Using this corpus, Garcia Lecumberri et al. (2017) attempted to tease apart cross-linguistic factors from those of native/non-native speech to explore a variety of features of oral speech made possible by a dialogic corpus. Similar to Gut (2017), the corpus analysis presented in Garcia Lecumberri et al. allowed for not only an investigation of L2 fluency, but also phonological features such as pitch and vowel formant analyses. Of particular interest to issues of L2 fluency, the results of the analysis of speech rate (measured as words per minute) indicated effects of both nativeness and language being spoken (i.e., English vs. Spanish). Specifically, while it was the case that speech rate was generally slower for non-native speech, the greater number of monosyllabic words in English vs.

Spanish meant that this reduction was less pronounced for the Spanish native speakers when speaking their L2 (English). The findings from Garcia Lecumberri et al. demonstrate how this type of bi-directional corpus design can help further tease apart the effects of nativeness and cross-linguistic influence in the study of L2 fluency.

Future Directions

Investigations of L2 fluency have benefitted from the recent growth in learner corpus research. This section provides several suggestions for future work in this area. Given the fact that currently much of the transcription and coding of corpora for fluency analysis are done manually and furthermore that this coding is time-consuming and can result in low reliability for some of the most complex annotations, it appears that a pressing need still exists for principled ways of automatizing coding. This would include new developments in automated processes as well as comparisons with manually-coded corpora to allow for a better understanding of what can be reasonably expected when it comes to reliability.

In addition to automating processes, however, there are other ways to alleviate the time-consuming nature of coding and annotating data. The public sharing of annotated corpora using agreed-upon conventions for annotation allow for new investigations as well as potential for replication. While a great deal of data and tasks are available, the continued sharing, especially as pertains to coding decisions, could be another way to help ensure clarity (and perhaps encourage consistency) across studies. For example, Tracy-Ventura & Huensch (2018) in their critical reflection on the processes and decision-making involved in the creation of a publicly-shared, longitudinal corpus discussed the complexities of coding their data into utterances in CLAN (in this case, ASUs) particularly

given the fact that they were coding across multiple languages (English, French, and Spanish) but having to base decisions on literature published about the coding of English. The sharing of detailed coding procedures (e.g., Hilton, 2009) would allow for continued open discussion among the community as well as would make steps in ensuring the comparability across corpora, not to mention saving time.

Another area of future research that is just in its infancy in L2 fluency is further investigations into individual differences. Most studies exploring L2 fluency report large amounts of individual variation regardless of the aspect of fluency. Some work in LCR providing promising directions is that which combines quantitative analyses with the corpus as a whole with qualitative analyses of individual speakers (e.g., Brand & Götz, 2011; Gut, 2017). As described in the Representative Corpora and Research section, Brand and Götz (2011) qualitatively analyzed a subset of speakers who represented different learner profiles based on their quantitative analyses (e.g., those with the most/fewest grammatical errors, most fluent speech, average error and fluency scores). Using this approach, Brand and Götz demonstrated that the speaker with average scores for fluency and accuracy was rated as the most proficient indicating that raters relied on a variety of variables to rate proficiency. Learner corpora appear to be particularly well-suited for investigations of this kind that combine large-scale quantitative analyses with more detailed qualitative analyses with a subset of speakers.

Further Reading (annotated)

Götz, S. (2013). *Fluency in native and non-native English speech*. Amsterdam: John Benjamins. This book begins with a thorough introduction to issues in native and non-native fluency before

providing an in-depth analysis and comparison of fluency features in the L2 English LINDSEI-GE and the English L1 LOCNEC corpora.

Gut, U. (2009). *Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German*. Frankfurt: Peter Lang.

Arguing for the use of a corpus-based approach for investigations of second language acquisition, this book explores L2 phonological acquisition by investigating L2 English and L2 German speech from the LeaP corpus.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.

An in-depth introduction to L2 fluency from a cognitive science perspective that draws upon work from a variety of fields to bring together multiple perspectives.)

Related Topics

Ch 6 Annotating Learner Corpus Data, Ch 13 The TALKBANK System, Ch 24 Accuracy, Ch 25 Complexity, and Ch 30 Proficiency.

References

- Baker, R., & Hazan, V. (2011). DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, 43, 761–770.
- Ballier, N., & Martin, P. (2015). Speech annotation of learner corpora. In A. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 107–134). Cambridge: Cambridge University Press.
- Ballier, N., & Martin, P. (2013). Developing corpus interoperability for phonetic investigation of learner corpora. In N. Ballier, A. Diaz-Negrillo, & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 33–64). Amsterdam: John Benjamins.
- Belz, M., Sauer, S., Lüdeling, A., & Mooshammer, C. (2017). Fluently disfluent? Pauses and repairs of advanced learners and native speakers of German. *International Journal of Learner Corpus Research*, 3(2), 118–148.
- Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer (Version 5.4.18) [Computer program]. Retrieved from <http://www.praat.org/>
- Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed, and repairs. *Language Testing*, 30, 159–175.
- Brand, C., & Götz, S. (2011). Fluency versus accuracy in advanced spoken learner language: A multi-method approach. *International Journal of Corpus Linguistics*, 16, 255–275.
- Davies, A. (2003). *Native speaker: Myth and reality*. Clevedon: Multilingual Matters.
- De Bot, K. (1992). A bilingual production model: Levelt's "Speaking" model adapted. *Applied Linguistics*, 13, 1–24.

- De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures*, 2, 225–246.
- De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, 54, 113–132.
- De Jong, N. H., & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In R. Eklund (Ed.) *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech (DiSS)*, 17–20. Stockholm, Sweden: Royal Institute of Technology.
- De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, 36, 223–243.
- De Jong, N., & Mora, J. (2017). Does having good articulatory skills lead to more fluent speech in first and second languages? *Studies in Second Language Acquisition*, 1–13.
doi:10.1017/S0272263117000389
- De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, 34, 5–34.
- De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, 41, 385–390.
- Derwing, T. M. (2017). L2 Fluency Development. In S. Loewen, & M. Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 246–259). New York: Routledge.

- Derwing, T., Munro, M., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31, 533–557.
- Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning*, 59, 866–896.
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354–375.
- Garcia Lecumberri, M. L., Cooke, M., & Wester, M. (2017). A bi-directional task-based corpus of learners' conversational speech. *International Journal of Learner Corpus Research*, 3(2), 175–195.
- Gilquin, G., De Cock, S., & Granger, S. (2010). The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM. Louvain, Belgium: Presses universitaires de Louvain.
- Gilquin, G., & Granger, S. (2015). Learner language. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics* (pp. 418–436). Cambridge: Cambridge University Press.
- Granger, S. (1998) The computerized learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 3–18). London: Longman.
- Götz, S. (2013). Fluency in native and non-native English speech. Amsterdam: John Benjamins.
- Gósy, M., Gyarmathy, D., & Beke, A. (2017). Phonetic analysis of filled pauses based on a Hungarian-English learner corpus. *International Journal of Learner Corpus Research*, 3(2), 149–174.

- Gut, U. (2017). Phonological development in different learning contexts. *International Journal of Learner Corpus Research*, 3(2), 196–222.
- Gut, U. 2012. ‘The LeaP corpus. A multilingual corpus of spoken learner German and learner English. In Th. Schmidt, & K. Wörner, K. (Eds.), *Multilingual corpora and multilingual corpus analysis* (pp. 3–23). Amsterdam: John Benjamins.
- Gut, U. (2009). *Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German*. Frankfurt: Peter Lang.
- Hilton, H. (2014). Oral fluency and spoken proficiency: Considerations for research and testing. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 27–53). Buffalo, NY: Multilingual Matters.
- Hilton, H. (2009). Annotation and analyses of temporal aspects of spoken fluency. *CALICO Journal*, 26, 644–661.
- Hilton, H., Osborne, N. J., Derive, M.-J., Succo, N., O’Donnell, J., Billard, S. & Rutigliano-Daspet, S. (2008). Corpus PAROLE (Laboratoire LLS, Université de Savoie). TalkBank: SLA Data, <http://talkbank.org/data/SLA/>. Pittsburgh: Carnegie-Mellon University.
- Huensch, A., & Tracy-Ventura, N. (2017a). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *The Modern Language Journal*, 101, 275–293.
- Huensch, A., & Tracy-Ventura, N. (2017b). 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.

- Huensch, A., Tracy-Ventura, N., Bridges, J., & Cuesta-Media, J. (2019). Variables affecting the maintenance of L2 proficiency and fluency four years post-study abroad. *Study Abroad Research in Second Language Acquisition and International Education*, 4, 96–125.
- Kahng, J. (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, 39, 569–591.
- Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, 64, 809–854.
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32, 145–164.
- Krause, T., & Zeldes, A. (2016). ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities*, 31(1), 118–139.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387–417.
- Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83–122). Oxford: Oxford University Press.
- MacWhinney, B. (2017). A shared platform for studying second language acquisition. *Language Learning*, 67, 254–275.
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum.

- 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., & Littre, D. (2013). Tracking learners' progress: Adopting a dual "corpus cum experimental data" approach. *The Modern Language Journal*, 97, 61–76.
- Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). *The Anglophone student abroad: Identity, social relationships and language learning*. New York: Routledge.
- Myles, F. (2015). Second language acquisition theory and learner corpus research. In A. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 309–331). Cambridge: Cambridge University Press.
- 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.
- 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.
- Pellegrino, F., Coupé, C., & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language*, 87, 539–558.
- Quan, L., & Weisser, M. (2015). A study of 'self-repair' operations in conversation by Chinese English learners. *System*, 49, 39–49.
- Riazaantseva, A. (2001). Second language proficiency and pausing: A study of Russian speakers of English. *Studies in Second Language Acquisition*, 23, 497–526.

- Sauer, S., & Lüdeling, A. (2016). Flexible multi-layer spoken dialogue corpora. *International Journal of Corpus Linguistics*, 21, 419–438.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.
- Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics*, 54, 79–95.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36, 1–14.
- Skehan, P., Foster, P., & Shum, S. (2016). Ladders and snakes in second language fluency. *International Review of Applied Linguistics in Language Teaching*, 54, 97–111.
- Staples, S. (2015). Spoken discourse. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics* (pp. 271–291). Cambridge: Cambridge University Press.
- Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal*, 65, 71–79.
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–276). Philadelphia/Amsterdam: John Benjamins.
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17, 84–119.
- Tracy-Ventura, N., & Huensch, A. (2018). The potential of publicly shared longitudinal learner corpora in SLA research. In A. Gudmestad & A. Edmonds (Eds.), *Critical Reflections on data in second language acquisition*. Philadelphia/Amsterdam: John Benjamins. DOI: 10.1075/llt.51.07tra

- Trouvain, J., & Möbius, B. (2014). Sources of variation of articulation rate in native and non-native speech: comparisons of French and German. *Proceedings of Speech Prosody (SP7)*, 275–279.
- Trouvain, J., Zimmerer, F., Möbius, B., Gósy, M., & Bonneau, A. (2017). Segmental, prosodic and fluency features in phonetic learner corpora: Introduction to the special issue. *International Journal of Learner Corpus Research*, 3, 105–117.
- Tullock, B., & Ortega, L. (2017). Fluency and multilingualism in study abroad: Lessons from a scoping review. *System*, 71, 7–21.
- Villegas, J., Cooke, M., Aubanel, V., & Piccolino-Boniforti, M. A. (2011). MTRANS: A multi-channel, multi-tier speech annotation tool. *Proceedings of the 12th Annual Conference of the International Speech Communication Association: INTERSPEECH*. Florence, Italy, 3237–3240.

Notes

1 Further information about how this coding is accomplished can be found in the *Software and Tools for Data Coding and Analysis* section.