



UNIVERSITY OF LEEDS

This is a repository copy of *New directions and developments in defining, analyzing and measuring L2 speech fluency*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/105129/>

Version: Accepted Version

---

**Article:**

Wright, CEM [orcid.org/0000-0003-3962-7903](http://orcid.org/0000-0003-3962-7903) and Tavakoli, P (2016) New directions and developments in defining, analyzing and measuring L2 speech fluency. *International Review of Applied Linguistics in Language Teaching*, 54 (2). pp. 73-77. ISSN 0019-042X

<https://doi.org/10.1515/iral-2016-9990>

---

© 2016, De Gruyter. This is an author produced version of a paper published in *International Review of Applied Linguistics in Language Teaching*. Uploaded in accordance with the publisher's self-archiving policy.

**Reuse**

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

Clare Wright and Parvaneh Tavakoli

## **New directions and developments in defining, analyzing and measuring L2 speech fluency**

### Introduction to the special issue

The investigation of L2 speech fluency over the past decade has brought significant progress in understanding its multi-faceted nature and its role in SLA, particularly in temporal terms of fluid automatic speech production (Kormos 2006; Segalowitz 2010). Researchers have highlighted the importance of different language typologies, clarifying the relationship between L2 and L1 fluency (de Jong 2012), working towards careful and consistent measurements of fluency (Skehan 2009) and bringing rigour to models of L2 speech production (Kormos 2006). Recent research findings highlight L2 fluency as a reliable predictor of L2 proficiency (de Jong et al. 2012; Revesz et al. 2014), but also a characteristic that retains some traits of L1 speech production (de Jong et al. 2012), with implications for SLA research on L2 development and ultimate attainment. Notwithstanding the progress research has made in this area, L2 fluency still remains a complex research construct in SLA, an aspect of performance difficult to define and measure consistently across different tasks and conditions, and a characteristic of language use that many L2 learners may find difficult to develop in and out of the classroom.

The special issue provides a timely opportunity to revisit some of the several unknowns about L2 fluency, particularly to refine the current range of theoretical and empirical approaches to defining the construct of L2 fluency, and implications for measuring fluency (Segalowitz). The empirical studies from the other invited authors focus specifically on four main issues: exploring what may affect variability in performance and development in different tasks (Tavakoli), to consider how measures of fluency can differentiate underlying cognitive demands at planning and utterance level of speech (Skehan et al.), to examine factors that affect both L1 and L2 fluency (de Jong), and to include the significance of listener perception and comprehensibility (Prefontaine and Kormos). Some of the papers combine theoretical, contextual and empirical insights, while others are more specific in addressing current debates over standardising the way fluency measures are operationalised. Both quantitative and qualitative approaches are used, providing therefore a useful multi-faceted comparative collection of constructs, methods and evidence to take our understanding of fluency development forward.

In the first scene-setting paper, Segalowitz moves the framework of the fluency agenda forward by extending the current *descriptive* approach to an *exploratory* framework. Drawing on a dynamic systems perspective and considering language in its broader sociolinguistic context, Segalowitz introduces a fresh perspective that can

potentially allow for both identification of mechanisms and processes underlying fluency and emergence of common patterns of fluency and disfluency, driven by language use in authentic communicative contexts. He proposes that combining rigorous cognitive science with communicative learning research provides a broader framework that would enable researchers to study fluency more insightfully within the larger context of second language acquisition. In his paper, after a detailed discussion of the differences between cognitive, utterance and perceived fluency, Segalowitz argues that research in this area so far has sought to establish a catalogue of L2-specific utterance fluency features and an indication of how these features are linked with underlying cognitive operations. What is more urgently needed, he argues, is to provide a detailed account of the challenges which these cognitive factors impose on L2 learners' speech development in social interaction, and how these challenges can be overcome. Working towards this broad perspective, the paper suggests that using existing theories of learning such as a *usage-based* approach to language acquisition, and a *transfer-appropriate processing* approach to memory and knowledge retrieval, can be helpful in understanding the wider issues affecting communicative fluency in ways that could open new horizons to understanding and operationalising fluency.

Paper two, from Skehan, Foster and Shum, reports on a study comparing first and second language fluency, in which they examine the influences on fluency caused by the demands of the conceptualisation and formulation stages of speech production. By making a distinction between clause-level and discourse-level fluency, the authors explain the relationship between dysfluencies caused by different demands of processing, and attempt to investigate measures that can represent these two levels of fluency in the light of the need for running parallel processes of macroplanning and microplanning in speech production. The authors argue that distinguishing between discourse-based and clause-based fluency not only provides a more reliable basis for comparing native speaker and non-native speaker fluency, but it allows researchers to identify and measure the influence of task design on fluency that has so far remained under researched.

The third paper from De Jong introduces a new lexical perspective to understanding the role of pauses before or during runs in L1 and L2 speech, focusing in depth on the relationship between location of pauses, level of proficiency and use of low frequency vocabulary in spontaneous speech production. De Jong adopts a detailed and systematic statistical approach to analysing data from Turkish and English L2 Dutch speakers on the one hand and L1 Dutch speakers on the other. The analysis demonstrates important differences between pause patterns external and internal to utterances (defined here in the well-established sense of AS units – Foster et al., 2000), and finds significant connections between frequency of words and pause location. However, in both dimensions, such pausing patterns were more similar than different across both L1 and L2 speech. The findings of the study are crucial to our understanding of existing speech production models, as they provide robust evidence to support the claim that pausing can be the opportunity for conceptual planning not only in L1 but also L2 production processes (see also Skehan et al. in this volume). The other key

contribution De Jong's study makes to the field is the introduction of lexical frequency as another contextual factor to be taken into account in pausing patterns, regardless of the degree of automaticity with which the language is produced.

In paper four, Tavakoli challenges current approaches to defining and measuring L2 fluency, and argues that research in this area has paid minimal attention to conceptualising and operationalising fluency in interaction – i.e. in dialogic mode. By comparing the performance of L2 speakers on both monologic and dialogic tasks, Tavakoli's paper demonstrates the differences between the same speaker's fluency profiles in the two modes, and indicates which measures can more reliably capture fluency in each mode. She includes some of the principles of discourse analysis and conversation analysis for analysing aspects of fluency in a dialogic performance, particularly in relation to the very thorny and under-researched issue of what role is played by pauses in between turns. Like Segalowitz, she thus adds a discourse dimension for operationalising and measuring interactive features of temporal fluency, in a new direction for researching communicative speech.

In the final fifth paper, Kormos and Prefontaine add a novel perspective to discussions of L2 fluency by considering how L2 speech is perceived by the listener. This brings a more holistic approach to the construct of L2 fluency, by advocating the notion that fluency as "speech competence" also involves being successfully comprehended, not just produced. Listener ratings of fluency have been studied before, often using generic ratings to see how listeners perceive rate, effortlessness, richness of vocabulary and comprehensibility, but Kormos and Prefontaine promote the importance of more qualitative perceptions in terms of prosody and stress patterns at suprasegmental or discourse level. They present data from a cross-sectional study of adult English learners of French on immersion programmes in Canada, across a range of proficiency levels, performing narrative tasks at using differing levels of task complexity. Naïve raters, who were deliberately not given a prior definition of fluency, were asked to write their impressions of what most influenced their perceptions of L2 fluency in French, which were then subjected to careful thematic analysis. Raters valued, as in other studies, temporal measures such as speech rate, number of pauses and amount of self-correction, but they also highlighted their prioritisation of rhythm and stress over the temporal measures. The study thus foregrounds the importance in gaining speech fluency of developing L2-based prosody, which can remain challenging even at high levels of proficiency. This becomes particularly important for overcoming transfer effects from non-stress-timed languages such as English, when acquiring French or other stress-timed languages.

By combining the range of perspectives here across different aspects of L2 fluency, investigating both theoretical and empirical issues, this special issue brings much needed light on the complexities involved in defining and measuring L2 fluency, and drives forward the research agenda on L2 fluency and its place in SLA research. We promote in this collection a new way of operationalising L2 speech research by bringing together approaches based on specific utterance-level analysis with work investigating how speech fluency is affected by social and contextual demands. Inevitably, in

broadening the field of enquiry, and deliberately setting out to bring different research paradigms together, we raise questions of how to assure rigour, systematicity and clarity in working on fluency as such a multi-faceted construct. It is important to engage with these questions to avoid L2 fluency becoming too narrow. Back in 1979 Fillmore identified four dimensions of L1 fluency as time filled with talk, incorporating semantic density, communicative appropriacy and creative, imaginative use of language. This is recognisably the basis of the goal of communicative competence that has underpinned much modern L2 teaching but that can elude so many L2 learners. We hope in this issue we have re-emphasised the value for SLA research on fluency of moving away from narrow if rigorous analysis of the temporal dimension in Fillmore's original model. We suggest it may be better to talk not of fluency, but fluencies, as a way of capturing both the breadth and depth of L2 speech research going forward within this new utterance/discourse perspective – in this way we can find new research insights to refresh the value to teachers and learners of what communicative competence really is.

- de Jong, Nivja, Margarita Steinel, Arjen Florijn, Rob Schoonen & Jan Hulstijn. 2012. Facets of speaking proficiency. *Studies in Second Language Acquisition* 34 (1). 5–34.
- Fillmore, Charles. 1979. On fluency. In Charles Fillmore, Daniel Kempler & William Wang (eds.), *Individual differences in language ability and language behavior*, 85–102. New York: Academic Press.
- Kormos, Judit, 2006. *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Pauline Foster, Alan Tonkyn & Gillian Wigglesworth. 2000. Measuring spoken language: a unit for all reasons. *Applied Linguistics* 21 (3): 354–375.  
doi:10.1093/applin/21.3.354
- Revesz, Andrea, Rebecca Sachs & Mika Hama. (2014). The effects of task complexity and input frequency on the acquisition of the past counterfactual construction through recasts. *Language Learning* 64 (3). 615–650.
- Segalowitz, Norman. 2010. *Cognitive bases of second language fluency*. New York: Routledge.
- Skehan, Peter. 2009. Lexical performance by native and non-native speakers on language learning tasks. In Brian Richards, Helmut Daller, David Malvern & Paul Meara (eds.), *Vocabulary studies in first and second language acquisition: The interface between theory and application*, 107–124. London: Palgrave Macmillan.



# Second language fluency and its underlying cognitive and social determinants

NORMAN SEGALOWITZ

## Abstract

*In studying second language (L2) fluency attainment, researchers typically address questions about temporal and hesitation phenomena in a descriptive manner, cataloguing which features appear under which learning circumstances. The goal of this paper is to present a perspective on L2 fluency that goes beyond description by exploring a potential explanatory framework for understanding L2 fluency. This framework focuses on the cognitive processing that underlies the manifestation of fluency and disfluency, and on the ways social context might contribute to shaping fluency attainment. The framework provides a dynamical systems perspective of fluency and its development, with specific consequences for a research program on L2 fluency.*

*This framework gives rise to new questions because of its focus on the intimate link between cognitive fluency and utterance fluency, that is, between measures of the speed, efficiency and fluidity of the cognitive processes thought to underlie implementation of the speech act and measures of the oral fluency of that speech act. Moreover, it is argued that cognitive and utterance fluency need to be situated in the social context of communication in order to take into account the role played by the pragmatic and the sociolinguistic nature of communication in shaping L2 fluency development.*

Keywords: L2 fluency; second language learning; cognitive processing

## 1. Introduction

Discussions about how to assess second language (L2) fluency often begin by acknowledging that the meaning of the term *fluency* is quite difficult to pin down. For example, in English the word *fluency* can mean different things in different contexts. Sometimes it refers to a person's global competence or proficiency (*She is fluent in Japanese*), sometimes to the fluidity of speech (*He is a fluent public speaker*), sometimes separately to speaking, listening, reading, or writing abilities. Also, translating the term *fluency* into other languages can be difficult; French *aisance à l'oral* 'ease of speaking'

focuses on the speaker's experience whereas the Spanish *fluidez* 'fluidity' focuses on the quality of the action. This variability is problematic because a meaningful discussion about fluency requires agreement on what is being talked about. The solution to this problem has typically been to narrow the focus to just one meaning. In this regard, a distinction is often made between knowledge of the L2 (e.g., of phonology, vocabulary, syntax, semantics, sociolinguistic and pragmatic considerations) and the *fluency* or fluidity with which a speaker is able to implement that knowledge (rate of speech, pausing, hesitation and other temporal phenomena). There remains, however, a deeper problem.

This deeper problem has to do with the goal one has in mind when investigating L2 fluency. Often the goal involves *describing* L2 speakers' fluidity in order to highlight differences between their speech and that of native speakers, and then (sometimes) following up with qualitative analyses to make inferences about possible reasons for the observed disfluencies (e.g., Tavakoli 2011). Such descriptions provide valuable insights into the nature of fluency and can be useful for comparing the impact of different learning experiences on fluency attainment or for studying the relationship between a particular variable (age, aptitude, ethno-linguistic identity, intelligence, learning style, personality) and fluency development. As a strategy, however, this approach has limitations; it does not address the problem of how to decide which speech features to look at or how best to operationalize them (for reviews see Kormos 2006; Segalowitz 2010). Without a principled way to narrow down choices, the field risks becoming populated by a collection of studies whose results are difficult to relate to one another and from which to draw clear generalizations. This poses a challenge for defining what constitutes progress in the field, especially with respect to developing a theory of L2 fluency acquisition. However, an alternative exists.

This alternative involves trying, from the outset, to *explain* L2 fluency. Here the aim is to identify, in a theory-driven way, the mechanisms and processes responsible for L2 (dis)fluency. An explanatory approach would make it possible to chart progress in the field; as mechanisms and processes underlying fluency phenomena become identified, the initially very large number of possible fluency phenomena to study becomes reduced. Patterns begin to emerge and fluency can be situated in the larger context of L2 acquisition as a whole. This paper reviews a framework for such an approach (originally presented in Segalowitz 2010, but discussed here in light of recent developments).

Three ideas are central to the framework. The first comes from Goldman-Eisler (1951, 1961, 1968) whose pioneering work set the stage for subsequent research on L2 fluency. She wrote that "the complete speech act is a dynamic process, demanding the mobilization in proper sequence of a series of complex procedures and is the temporal integration of serial phenomena" (1968: 6). Goldman-Eisler points to the central role played by cognitive mechanisms in shaping the temporal phenomena of oral fluency and she draws attention to how these mechanisms are organized into a dynamic system. The second insight comes from Rehbein (1987) for whom being fluent "means that the activities of planning and uttering can be executed nearly simultaneously by the speaker of the language" (p. 104). Rehbein points to the rapid speed, automaticity and efficiency

of the underlying mechanisms responsible for fluent L2 speech. Finally, Meisel (1987) emphasizes the importance of the communicative acceptability of the speech act, that is, its communicative fit according to the expectations of the interlocutor. Taken together, these three insights suggest that the study of L2 fluency needs to focus "on features of L2 performance that are reliable indicators of how efficiently a speaker is able to mobilize and temporally integrate, in a nearly simultaneous way, the underlying processes of planning and assembling an utterance in order to perform a communicatively acceptable speech act" (Segalowitz 2010: 165). This goes well beyond that of describing and documenting fluency phenomena, and it has implications for how to approach questions about L2 fluency in a systematic way.

## **2. L2 utterance, cognitive and perceived fluency**

For the perspective presented here, we first need to distinguish among three aspects of L2 fluency—L2 utterance fluency, L2 cognitive fluency, and L2 perceived fluency.

### *2.1. L2 utterance fluency*

L2 utterance fluency refers to the fluidity of the observable speech as characterized by measurable temporal features, such as syllable rate, duration and rate of hesitations, filled and silent pauses, and including what Skehan (2003) has identified as breakdown fluency and repair fluency. Often such features can be operationally defined in more than one way (Hilton 2014; Kormos 2006; Segalowitz 2010) and indeed, for some features debate exists about how best to do this. For example, De Jong and Bosker (2013) recently addressed the problem of how to choose the lower threshold (minimum duration) for defining disfluent silent pauses, long believed to be an important marker of oral fluency. As they pointed out, until recently researchers have used a wide variety of lower thresholds, from 100 to 1000 milliseconds. De Jong and Bosker (2013) analyzed the speech of L2 speakers of Dutch performing a variety of speaking tasks. They looked at how speakers' pause rates correlated with a vocabulary size measure of overall proficiency as a function of 21 different lower cut-off thresholds (20, 50, 100, and then up to 1000 milliseconds in 50 milliseconds steps). They found that a minimum duration threshold of around 250-300 milliseconds yielded the highest correlation between silent pause rate and L2 proficiency, peaking around  $r = -.53$  (lower pause rate associated with greater vocabulary size). When silent pauses were defined by shorter or longer duration thresholds, the correlations were much weaker. Interestingly, in contrast to silent pause *rate*, mean pause *duration* did not yield significant correlations with vocabulary size. This study represents an important advance in fluency research because, for the first time a cognitive measure of L2 proficiency (here, vocabulary size) was used to justify selection of a particular operational definition of an utterance fluency feature (here, the minimum duration for defining silent pauses) and to demonstrate its superiority over competing operational definitions (the other proposed cut-off threshold levels). This finding moves the discussion of how utterance fluency reflects cognition from

speculation to empirical test. Moreover, the strategy of looking at the association between a cognitive measure and an oral fluency measure led to the conclusion that silent pause *rate*, and not mean silent pause *duration*, was the relevant pause feature to focus on.

## 2.2. L2 cognitive fluency

L2 cognitive fluency refers to the fluid operation (speed, efficiency) of the cognitive processes responsible for performing L2 speech acts. This includes not just the articulatory act itself but the mobilization and temporal integration of mental processes that give rise to the utterance (Goldman-Eisler 1968). These cognitive processes thus include the speed and efficiency of semantic retrieval, the handling of the attention-focusing demands inherent in utterance construction, operations in working memory, among others.

Before continuing to explore cognitive fluency as it relates to utterance fluency, it is important to contrast the cognitive fluency under discussion here from two other kinds encountered in the literature. One of these concerns the fluency of *general-purpose* cognitive control processes involved in the regulation of all mental activities and behaviours, including in the L1. These include monitoring and updating operations in working memory, shifting focus of attention between mental sets, and inhibition or the overriding of responses, among others (Miyake and Friedman 2012). The fluidity (speed, smoothness, efficiency) with which these processes operate can vary across individuals and are treated as relatively stable person characteristics. These individual differences may explain some of the variability across individuals in the L2 (e.g., poor general working memory has been shown to affect L2 learning; Linck et al. 2014; Williams 2011), but individual differences in *general-purpose* cognitive processing alone cannot explain disfluency that is *specific* to the L2 (e.g., slower lexical access in L2 than L1; Segalowitz and Freed 2004). For this reason, the focus is on *L2-specific* modes of cognition that might underlie L2 utterance fluency.

Of course, measures of some cognitive skill related to L2 processing, for example speed of lexical access in the L2, are likely to also pick up on aspects of related, general-purpose skills (e.g., aspects of lexical access that are not specific to any given language). This is certainly true for utterance fluency, where people's general speaking style (e.g., speech rate tendencies) can result in L1 and L2 utterance fluency measures correlating with each other (De Jong et al. 2013). Thus, for example, some aspect of L2 speech rate may partly reflect a habitual way of speaking, in addition to an aspect that is L2-specific. The same is most certainly likely to be true for measures of cognitive fluency (Segalowitz 2010). Because the L1 is highly overlearned and practiced, L1 performance can be thought of as providing baseline levels of cognitive and utterance fluency characteristic for each individual. L1 measures thus provide a good way to control for such potential confounds and can be used to obtain *L2-specific* measures (e.g., by residualizing the L2 data against L1 data). Using L1 measures as baseline has other benefits. It helps to control for individual differences in handling particular task demands that are otherwise unrelated to language. For example, individuals may differ in basic motor speed when

pressing a reaction time panel or differ in intelligence, motivation, or personality in ways that affect task performance, thereby adding noise to the data. For all these reasons, research needs to focus on L2-*specific* measures of cognitive fluency and utterance fluency, something that is still not common practice in fluency research.

The other kind of fluency not to be confused with the cognitive fluency under consideration here is *subjective cognitive fluency*. This refers to the idea that when people perform cognitive acts they often have a sense of the ease with which they carried them out, such as the ease of recalling a word or recognizing a picture (Tversky and Kahneman 1973; Unkelbach and Greifeneder 2013). These experiences can sometimes be misleading to the person having them and result in cognitive illusions, such as believing that a stimulus that feels very familiar must have been recently presented when in fact it had not. Such cognitive illusions—and the subjective experiences of cognitive fluency that give rise to them—may even play a role in L2 fluency development by, say, affecting motivation to learn or to use the target language in a given situation (see Segalowitz 2010). However, the mechanisms underlying these subjective experiences must be distinguished from the cognitive mechanisms that give rise to the observable features of utterance fluency.

Returning now to L2-specific cognitive fluency, there are several promising candidate measures to consider, among them speed and efficiency of making word-meaning links, operationalized respectively as reaction time (RT) speed and RT stability (coefficient of variability (CV) of RT; Segalowitz and Segalowitz 1993). Note that De Jong and Bosker (2013), in the study cited earlier, used a proficiency-related cognitive measure, namely L2 vocabulary size, to evaluate a potential utterance fluency measure. Vocabulary size, however, is not a cognitive *fluency* measure. It is a knowledge measure, although it may be strongly associated with cognitive fluency (and even with spoken fluency; Hilton, 2008). It would be interesting, therefore, to see whether a cognitive *fluency* measure would support, perhaps even more strongly, the specific results obtained in De Jong and Bosker's (2013) study. Regarding the RT and CV measures, initial support for these as L2 cognitive fluency measures comes from a study by Segalowitz and Freed (2004). They obtained RTs and CVs from a visual word semantic categorization (living-nonliving) task and found them to be associated with L2 fluency (degree to which speech was free of filled pauses). In that study, the cognitive fluency measures were L2-specific (residualized against the L1) but the utterance fluency measures were not (L1 oral measures were not available). (For more recent discussion of RT and CV as predictors of L2 proficiency, see Ankerstein 2014, and Lim and Godfroid 2014).

Another cognitive fluency measure proposed to underlie L2 utterance fluency is flexibility in the control of *linguistic attention* (Segalowitz 2010). This relates to the way grammatical elements (e.g., spatial prepositions; conjunctions) direct attention to relationships between elements within utterances. Linguistic attention flexibility can be operationalized as a switch cost measure (in milliseconds) obtained from an alternating runs experimental design (Rogers and Monsell 1995). Using this design, Taube-Schiff and Segalowitz (2005), Segalowitz and Frenkiel-Fishman (2005) and more recently Duncan, Segalowitz and Phillips (2014) have shown, in different ways, that linguistic attention is

related to L2 proficiency. In these experiments, participants performed two different but closely related tasks in a sequence that involved repeats and shifts of attention focus. For example, in Taube-Schiff and Segalowitz (2005), in Task A participants had to judge the verticality meaning (ABOVE/BELOW) of sentence fragments containing phrases such as *over the spot*. In Task B the same participants judged the proximity meaning (CLOSE/DISTANT) of sentence fragments containing phrases such as *near the place*. The tasks were sequenced to repeat and shift in the pattern ...AABBAABB... so that on half the trials attention focus was on a repeat of the previous task type and on half the trials attention focus had to shift. The RT difference between shift and repeat trials provided an index of attention focus flexibility. The results revealed an L2-specific cognitive shift cost—that is, a linguistic attention effect. This study, however, looked only at the relationship between L2 linguistic attention and L2 proficiency, and not L2 utterance fluency as such, something future research could address. It is reasonable to suppose that linguistic attention skill underlies some aspects of utterance fluency. This is because speaking fluidly requires shifting attention focus continuously while packaging information to make the utterance unfold properly. Poor cognitive control of linguistic attention may thus underlie aspects of L2 utterance disfluency (see Segalowitz 2010, for fuller discussion).

To date, RT and CV measures of cognitive fluency have always been obtained from visual, receptive tasks (i.e., judgments of visually presented words or sentence fragments). In contrast, utterance fluency measures are based on oral production tasks. This potential mismatch merits some comment. The cognitive tasks used are generally very simple and would not seem to pose modality-specific challenges (e.g., visual perceptual difficulties) that are otherwise unrelated to the language performance of interest. For example, in the task aimed at measuring lexical access (e.g., Segalowitz and Freed 2004), participants are simply asked to indicate whether *boat* refers to a living or nonliving object. In the task aimed at measuring sentence construction skill (Lim and Godfroid 2014), people have to indicate, for example, which word—*does* or *he*— best continues the sentence fragment *I wonder what...* Moreover, while these cognitive tasks do have a receptive aspect—namely, the stimulus must be read—they nevertheless also possess a production aspect in that one must mentally generate a word's meaning or mentally construct a sentence or sentence fragment. Also, these tasks are relatively free of articulation demands (they do not require oral production) and therefore they overlap little in demand characteristics with the tasks yielding the utterance fluency data. For these reasons, RT and CV measures of performance are suitable for studying the cognitive fluency underlying L2 speech production.

In sum, L2 cognitive fluency is the rapid and fluid mobilization of the complex cognitive procedures referred to by Goldman-Eisler (1968), and this includes the automatization of these processes which Rehbein (1987) called the nearly simultaneous execution of planning and uttering activities (corresponding to Levelt's (1989, 1999) formulator and articulator levels of the speaking process). Both L2 cognitive fluency and L2 utterance fluency can be operationalized and measured in both L2 and L1 contexts, thus making it possible to obtain L2-specific measures of each. In this way, L2 fluency can be unpacked into two separate but related components—cognitive and utterance

fluency. Before continuing discussion of these two and the proposed framework, however, it will be useful to consider one more dimension of L2 fluency—fluency as experienced by the listener/observer.

### 2.3. *L2 perceived fluency*

While the framework discussed in this paper focuses on L2 utterance and L2 cognitive fluency, it is important to consider briefly L2 *perceived* fluency and to distinguish it from L2 utterance fluency. L2 perceived fluency refers to subjective judgments of L2 speakers' oral fluency. Researchers often use such judgments to assign fluency levels to the L2 speakers under study (Bosker et al. 2012; Derwing et al. 2004; Préfontaine et al. 2015). Perceived fluency can reflect something about the objective characteristics of oral fluency. For example, Préfontaine (2013) collected fluency measures from L2 learners' of French, using three different speech elicitation tasks. She found that native speakers' ratings of L2 fluency correlated significantly with the L2 speakers' self-ratings of their fluency, and that these self-ratings did correlate significantly with objective utterance fluency measures. The strength of these significant correlations varied as a function of the speaking task, ranging from around .31 to around .65, indicating that perceived fluency is reliably related to objective measures of utterance fluency but that nevertheless there remains a great deal of variance in the objective measures not accounted for by the perceived fluency measures. It should also be recognized that an interlocutor's perception of and judgment about a speaker's fluency could potentially have an impact on the course of an interaction. For example, if an interlocutor's perceptions are somehow communicated to the speaker, this might lead the speaker to reallocate the amount of attention devoted to speech, thereby influencing the fluency and other characteristics of the L2 output. For theoretical perspectives relevant to this point see Michel (2011), Robinson (2011), and the volume edited by Housen et al. (2012). In sum, research on the topic of perceived fluency is important for at least two reasons. One is that it is useful to understand what speech features listeners focus in when drawing conclusions about a speaker's L2 fluency and proficiency. The other is that listeners' judgments of an L2 speaker's fluency may in some circumstances affect how speaker and listener interact, with consequences for the speaker's fluency.

That said, from the perspective of the framework under discussion here, it is nevertheless important to keep in mind that perceived fluency can only provide a *subjective* measure of utterance fluency and is only moderately associated with the objective measures of oral fluency. Moreover, measures of perceived fluency are seldom, if ever, adjusted to take into account fluency in the speaker's L1 (presumably because the speaker is assumed to be maximally fluent as a native speaker), despite the importance of this adjustment for obtaining L2-specific measures. For these reasons, perceived fluency is not an appropriate way to assess utterance fluency if the goal is discover links between cognitive fluency and utterance fluency.

### 3. A framework for understanding L2 fluency

So far, discussion has focused only on how to identify L2-specific utterance fluency features, especially those related to L2-specific aspects of cognitive fluency. Such identification will yield a catalogue of utterance features that goes beyond simply describing speech because it will also specify the connections between these features and their cognitive underpinnings. However, while useful, such a catalogue is nevertheless somewhat limited. These L2-specific cognitive-utterance fluency associations should also be situated within in a broader, theoretical perspective that can provide a basis for understanding the challenges that fluency poses to learners and possible routes for overcoming these challenges. As a step toward creating this broader perspective, it is useful to consider two ideas in particular. One comes from a *usage-based* approach to language acquisition and communication, and the other comes from a *transfer appropriate processing* approach to memory.

#### 3.1. A usage-based approach to language acquisition and communication

Up to this point, the discussion of L2 fluency has been largely decontextualized from the social and communicative situations in which language is acquired. What is missing is recognition of Meisel's (1987) point that speech acts must also have good communicative fit with interlocutors' expectations. Tomasello's (2003) usage-based approach to language acquisition provides a way to repair this (see also Barlow and Kemmer 2000, on usage-based theory in applied linguistics). Tomasello (2003) and Lieven and Tomasello (2008) point out that, normally, when people speak to each other, they engage in two important activities. The first is to establish joint attention, that is, getting each other to attend to objects, ideas and their inter-relationships in a specific way. People communicate not (only) about specific things and ideas, but about perspectives and ways of construing the world (e.g., *The man stood in front of the tree* conveys a different perspective from *The tree was located behind the man* even though both describe the same basic scene). The linguistic tools for establishing joint attention include, among other things, grammatical devices for conveying a particular perspective of the situation being talked about (here, insights from cognitive linguistics and construction grammar theorists are especially relevant; see Fauconnier 1994; Goldberg 1995; Langacker 1987, 1991; Talmy 2008). The second activity that interlocutors engage in is that they try to read and convey messages about social intentions (e.g., is the message meant to be informative, an admonishment, supportive, sarcastic, solicitous, etc.). The social message is a subtext conveyed in parallel with the main cognitive message. People always try to deal with the social message, even if it does not seem to be the main focus of the conversation. In sum, according to this attention/intention perspective on the nature of language communication, normal communication involves interlocutors attempting to establish joint attention and reading each other's social intentions.

Tomasello (2003) and Lieven and Tomasello (2008) developed this attention/intention perspective in terms of its implications for L1 acquisition. With respect to L2 acquisition, the question of interest here is how these attention/intention demands of communication might have an impact on a person's ability to speak fluently. There are two important points to consider here. One concerns the linguistic knowledge needed for carrying out the attention/intention functions of communication, and the other the role played by the attention/intention aspect of communication in memory retrieval. Regarding linguistic knowledge, to achieve a high level of fluency one clearly needs to master the target language's devices used for establishing joint attention and for conveying and reading social intentions. Establishing joint attention will require vocabulary knowledge for naming objects, events and their properties plus knowledge of the structural devices for appropriately conveying a perspective on the relationships among what is named (knowledge of how function words convey relationships, of word order conventions, agreement patterns, etc.). Conveying social intentions will require knowledge of the sociolinguistic and pragmatic dimensions of language use—choice of register, register shifting, idioms and fixed expressions, prosody, etc. Poor knowledge of these aspects of the target language could compromise the ability to communicate fluently by leading to inefficient word searches and awkward attempts to produce appropriately structured utterances. Beyond this, however, there is a second, less obvious way that the attention/intention aspect of communication may have an impact on L2 fluency. This brings us to the topic of transfer appropriate processing in memory retrieval.

### *3.2. Transfer appropriate processing*

Transfer appropriate processing refers to the idea that "memories are represented in terms of the cognitive operations engaged by an event as it is initially processed, and that successful memory retrieval occurs when those earlier operations are recapitulated" (Rugg et al. 2008:340; see also Danker and Anderson 2010; Roediger et al. 2002; Roediger and Gynn 1996; Tulving and Thompson 1973; Wing et al. 2015). This means a person's memory for recently learned information is linked to representations of the perceptual and cognitive activities that were engaged in when acquiring the information earlier. This is why, for example, during recall we often remember "irrelevant" pieces of accompanying information, such as what we were doing at the time we learned something. Transfer appropriate processing has the following implication for a framework for thinking about L2 fluency. Fluent speech requires rapid, smooth retrieval of information for formulating and articulating the intended message (Levelt 1989, 1999). This retrieval takes place under communication conditions that normally involve having to handle attention/intention demands that were described earlier. According to the principle of transfer appropriate processing, retrieval at the time of need will be facilitated (become more rapid, smooth, efficient) if, at the time of original learning, the learner also had to deal with attention/intention demands similar to those required at the time of need. An implication of this idea is that developing fluency requires L2 learning that takes place in genuinely communicative contexts, that

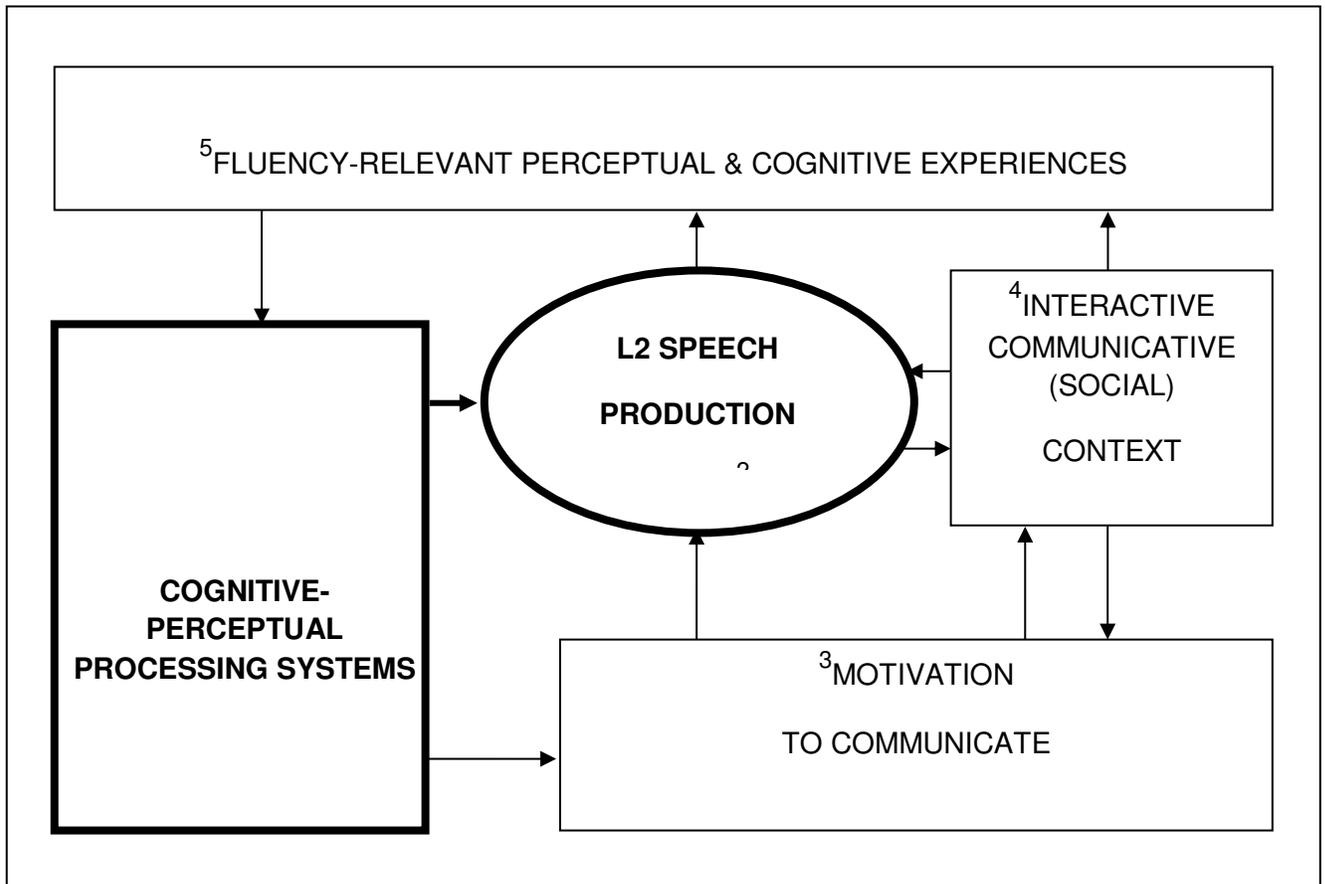
is, contexts that include dealing with the attention/intention demands of normal communication.

Incidentally, the attention/intention aspect of communication may have implications for L2 fluency researchers regarding how best to obtain speech samples in an ecologically valid manner. Most researchers attempt to elicit speech by using tasks that aim to be authentic or genuine in some way with respect to real world communication. Does this mean that speech elicitation tasks in L2 fluency research should always include an attention/intention aspect in their design? This question is important because it is known that the nature of a speech task can affect speech production (Tavakoli and Foster 2008; Tavakoli and Skehan 2005). Thus, it would be useful to know whether utterance fluency changes as a function of the presence or absence of the attention/intention demands (compared to story recall, text reading or other minimally interactive tasks). A challenge facing researchers, of course, will be to include attention/intention demands while keeping the elicitation task as controlled as possible.

Putting it all together, the framework that emerges can be summarized as in Figure 1. The core phenomena addressed by the framework are L2-specific speech features that characterize L2 fluency and the L2-specific cognitive operations associated with those speech features. Disfluent execution of these cognitive operations is what underlies L2 utterance disfluency. It is experience in using the language that sharpens the learner's cognitive-perceptual systems so that these cognitive operations become rapid, efficient and fluid, resulting in speech output that is fluent. For this cognitive fluency to develop, however, there must be repeated experiences in producing speech. Because the need to be fluent normally arises in interactive social contexts characterized by the attention/intention demands of communication, by the principle of transfer appropriate processing learning should also take place in contexts involving attention/intention demands if learning experiences are to facilitate fluency development. The figure also shows that motivation plays a role fluency development. Motivation not only energizes learners to use the L2, it can also shape the nature of the communicative situations in which learners use the L2, which may or may not be optimal for promoting fluency. Moreover, motivation itself can be enhanced or diminished by the learner's subjective experience of trying to use the L2, both in terms of the cognitive effort involved and in terms of certain psychological experiences regarding self and identity (see Dörnyei [2009] and Henry [2015] on the development of the *L2-self* in L2 motivation, and Segalowitz et al. [2009] on the link between fluency and ethno-linguistic identity). If learners cognitive and social experiences result in increased motivation to communicate, then they will engage in more L2 use, creating a positive feedback loop to the cognitive-perceptual processing system that enhances cognitive fluency, leading to improved utterance fluency and more successful L2 encounters. Thus, as shown in the figure, L2 fluency is the outcome of the operation of a dynamical system where cognitive, motivational, social, sociolinguistic, pragmatic and psycholinguistic considerations interact in complex ways (for more on dynamical system theory applied to L2 issues, see: de Bot and Freeman 2011; Larsen-Freeman and Cameron 2008; and Larsen-Freeman 2015).

---

Figure 1. Framework for L2 fluency (from Segalowitz 2010: 164, Figure 7.1).



Notes: <sup>1</sup>Cognitive fluency features include processing speed, stability and flexibility in the planning, assembly and execution of utterances in terms of lexical access and the use of linguistic resources (linguistic affordances) to express construals, handle sociolinguistic functions, and pursue psychosocial goals. <sup>2</sup>Utterance fluency features include speech rate, hesitation and pausing phenomena, etc. <sup>3</sup>Motivation includes willingness to communicate, beliefs about communication, language and identity, and the concept of the L2-Self. Motivation influences speech production and the selection of social contexts in which to speak. <sup>4</sup>The social context influences speech production by setting the cognitive task demands associated with communication and is the source for learning about linguistic affordances. <sup>5</sup>Experiences include frequency of exposure, opportunities for repetition practice, etc.

#### 4. Conclusion

This framework can be summarized in terms of three main points.

- (1) *Identifying the features of L2 fluency to study.* In order to identify the features of L2 fluency that are truly reflective of how a speaker handles the L2 as opposed to other co-occurring demands, it is important to distinguish among three different aspects of fluency—utterance, cognitive, and perceived fluency. Of central concern in the framework presented here are measures of utterance fluency that correlate highly with measures of cognitive fluency, thereby pointing to cognitive operations that underlie L2 speech production. Moreover, measures of utterance and cognitive fluency should, ideally, be made as L2-specific as possible by controlling for corresponding measures in the L1, to avoid confounds with general cognitive and language abilities and with abilities related to handling task-specific demand characteristics. Focusing research in this way will yield a set of L2-specific utterance fluency features that are related to the fluency of underlying L2-specific cognitive processes.
- (2) *Situating L2 fluency in a larger theoretical context.* In order to go beyond simply describing L2 fluency, it is important to locate discoveries about the cognitive-utterance fluency associations mentioned above in a theoretical context that can address how experience might shape fluency acquisition. Two considerations for this were presented. The first, derived from a usage-base theory of language acquisition, is that when learners develop L2 fluency through communicative experiences, what is learned is embedded in a neurocognitive environment of operations for establishing joint attention and for reading social intentions. The second, derived from psychological research on memory, is that memory retrieval is facilitated when the neurocognitive environment (the set of cognitive operations in play) that exists at the time of need matches in significant ways the neurocognitive environment that existed at the time of learning. Thus, the cognitive demands encountered at the time of learning should match as much as possible the anticipated future demands when there is a need to retrieve what was learned. Because these future demands will arise in the context of normal communication, the relevant cognitive operations are those associated with handling the attention/intention aspects of communication. An implication of this view is that the cognitive underpinnings of L2 fluency are affected by the cognitive consequences of engaging in social interaction during learning.
- (3) *Viewing L2 fluency as reflecting the operation of a dynamical system.* The view outlined above suggests that the cognitive operations underlying utterance fluency are themselves affected by fluency-relevant experiences shaped by social interactions, motivational states, and subjective experiences associated with using the L2. From this it follows that that L2 fluency attainment is the outcome of the operation of a complex system of mechanisms and processes that are dynamically interacting at all times. The implication here is that to investigate

the nature of L2 fluency one needs to take into account the many factors contributing to its development.

In sum, understanding the determinants of L2 fluency requires an appreciation of the cognitive underpinnings of L2 fluency phenomena, and this in turn requires an understanding of how these cognitive factors themselves are intimately bound up in the social-motivational matrix in which language learners find themselves.

## References

- Ankerstein, Carrie A. 2014. A psycholinguistic measurement of second language proficiency: The coefficient of variation. In Pascale Leclercq, Amanda Edmonds & Heather Hilton (eds.), *Measuring L2 proficiency: Perspectives from SLA*, 109-121. Bristol, UK: Multilingual Matters.
- Barlow, Michael. & Suzanne Kemmer (eds.). 2000. *Usage based models of language*. Stanford: CSLI Publications.
- Bosker, Hans R., Anne-France Pinget, Hugo Quené, Ted Sanders & Nivja. H. De Jong. 2012. What makes speech sound fluent? The contributions of pauses speed and repairs. *Language Testing*. doi: 10.1177/0265532212455394
- Danker, Jared F. & John. R. Anderson. 2010. The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin* 136(1). 87-102.
- de Bot, Kees., & Diane Larsen-Freeman. 2011. Researching second language development from a dynamic systems theory perspective. In Marjolijn H. Verspoor, Kees de Bot, & Wander Lowie. 2011. *A dynamic approach to second language development*, 5-23. Amsterdam: John Benjamins.
- De Jong, Nivja. H. & Hans R. Bosker. 2013. Choosing a threshold for silent pauses to measure second language fluency. In Robert Eklund (Ed.), *Proceedings of disfluency in spontaneous speech.*, 17–20. Stockholm: Royal Institute of Technology (KTH).
- De Jong, Nivja. H., Rachel Groenhout, Rob Schoonen & Jan H. Hulstijn. 2013. Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*. doi:10.1017/S0142716413000210
- Derwing, Tracey M., Marian J. Rossiter, Murray J. Munro & Ron I. Thomson. 2004. Second language fluency: Judgments on different tasks. *Language Learning* 54. 655-679.
- Dörnyei, Zoltán., 2009. The L2 motivational self system. In Zoltán Dörnyei & Ema Ushioda (eds.) *Motivation, language identity and the L2 Self* , 9-42. Bristol, UK: Multilingual Matters.
- Duncan, Hilary D., Norman Segalowitz & Natalie A. Phillips. 2014. Differences in L1 linguistic attention control between monolinguals and bilinguals. *Bilingualism: Language and Cognition*. doi:10.1017/S136672891400025X.

- Fauconnier, Gilles. 1994. *Mental spaces*. Cambridge, UK: Cambridge University Press.
- Goldberg, Adele. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: The University of Chicago Press.
- Goldman-Eisler, Frieda. 1951. The measurement of time sequences in conversational behaviour. *British Journal of Psychology* 42. 355-362.
- Goldman-Eisler, Frieda. 1961. Hesitation and information in speech. In Colin Cherry (ed.), *Information theory*, 162-174. London: Butterworths.
- Goldman-Eisler, Frieda. 1968. *Psycholinguistics experiments in spontaneous speech*. London: Academic Press.
- Henry, Alastair. 2015. The dynamics of possible selves. In Zoltán Dörnyei, Peter D. MacIntyre & Alastair Henry (eds.), *Motivational dynamics in language learning*, 83-94. Bristol, UK: Multilingual Matters.
- Hilton, Heather. 2008. The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal* 36(2). 153-166.
- Hilton, Heather. 2014. Oral fluency and spoken proficiency: Considerations for research and testing. In Pascale Leclercq, Amanda Edmonds & Heather Hilton (eds.), *Measuring L2 proficiency: Perspectives from SLA*, 27-53. Bristol, UK: Multilingual Matters.
- Housen, Alex., Folkert Kuiken, & Ineke Vedder. (Eds.) 2012. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins.
- Kormos, Judit. 2006. *Speech production and second language acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar, Volume 1: Theoretical prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald W. 1991. *Foundations of cognitive grammar, Volume 2: Descriptive application*. Stanford: Stanford University Press.
- Larsen-Freeman, Diane., & Lynne Cameron. 2008. *Complex systems and applied linguistics*. Oxford, UK: Oxford University Press.
- Larsen-Freeman, Diane. 2015. Ten 'lessons' from complex dynamic systems theory: What is on offer. In Zoltán Dörnyei, Peter D. MacIntyre & Alastair Henry (eds.), *Motivational dynamics in language learning*, 11-19. Bristol, UK: Multilingual Matters.
- Levelt, Willem. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, Willem. 1999. Producing spoken language: a blueprint of the speaker. In Colin M. Brown & Peter Hagoort (eds.), *The neurocognition of language*, 83-122. Oxford: Oxford University Press.
- Lieven, Elena. & Michael Tomasello. 2008. Children's first language acquisition from a usage-based perspective. In Peter Robinson & Nick C. Ellis (eds.), *Handbook of cognitive linguistics and second language acquisition*, 168-196. New York: Routledge.
- Lim, Hyojung. & Aline Godfroid. 2014. Automatization in second language sentence processing: A partial, conceptual replication of Hulstijn, Van Gelderen, and

- Schoonen's 2009 study. *Applied Psycholinguistics*, doi:10.1017/S0142716414000137
- Linck, Jared A., Peter Osthus, Joel T. Koeth & Michael F. Bunting. 2014. Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-013-0565-2
- Meisel, Jürgen. 1987. A note on second language speech production. In Hans W. Dechert & Manfred Raupach (eds.), *Psycholinguistic models of production*, 83-90. Norwood, NJ: Ablex Publishing Corporation.
- Michel, Marije. C. 2011. Effects of task complexity and interaction on L2 performance. In P. Robinson (Ed.), *Second language task complexity*, 141-173. Amsterdam: John Benjamins.
- Miyake, Akira., & Naomi P. Friedman. 2012. The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21(1). 8-14. doi: 10.1177/0963721411429458.
- Préfontaine, Yvonne. 2013. Perceptions of French fluency in second language speech production. *Canadian Modern Language Review* 69. 324-348.
- Préfontaine, Yvonne., Judit Kormos, & Daniel E. Johnson,. 2015. How do utterance measures predict raters' perceptions of fluency in French as a second language? *Language Testing*. doi: 10.1177/0265532215579530.
- Rehbein, Jochen. 1987. On fluency in second language speech. In Hans W. Dechert & Manfred. Raupach (eds.), *Psycholinguistic models of production*, 97-105. Norwood, NJ: Ablex Publishing Corporation.
- Robinson, Peter. 2011. Second language task complexity, the Cognition Hypothesis, language learning, and performance. In Peter Robinson (Ed.), *Second language task complexity*, 3-37. Amsterdam: John Benjamins.
- Roediger, Henry L., David A. Gallo, & Lisa Geraci. 2002. Processing approaches to cognition: the impetus from the levels of processing framework. *Memory* 10. 319-332.
- Roediger, Henry L. & Melissa J. Guynn. 1996. Retrieval processes. In Elizabeth L. Bjork & Robert A. Bjork (eds.), *Memory*, 197-236. New York: Academic Press.
- Rogers, Robert D. & Stephen Monsell. 1995. Costs of a predictable switch between simple and cognitive tasks. *Journal of Experimental Psychology: General* 124. 207-231.
- Rugg, Michael D., Jeffrey. D. Johnson, H. Park & Melina R. Uncapher. 2008. Encoding-retrieval overlap in human episodic memory: a functional neuroimaging perspective. In Wayne S. Sossin, Jean-Claude Lacaille, Vincent F. Castellucci & Sylvie Belleville (eds.), *Progress in Brain Research* (Vol. 169), 339-352. Amsterdam: Elsevier.
- Segalowitz, Norman. 2010. *Cognitive bases of second language fluency*. New York: Routledge.
- Segalowitz, Norman. & Barbara F. Freed. 2004. Context, contact and cognition in oral fluency acquisition: Learning Spanish in At Home and Study Abroad contexts. *Studies in Second Language Acquisition* 26. 173-199.
- Segalowitz, Norman. & Sarah Frenkiel-Fishman. 2005. Attention control and ability level

- in a complex cognitive skill: attention-shifting and second language proficiency. *Memory & Cognition* 33. 644-653.
- Segalowitz, Norman, Elizabeth Gatbonton & Pavel Trofimovich. 2009. Links between ethnolinguistic affiliation, self-related motivation and second language fluency: Are they mediated by psycholinguistic variables? In Zoltán Dörnyei & Ema Ushioda (eds.), *Motivation, language identity and the L2 Self*, 172-192. Bristol, UK: Multilingual Matters.
- Segalowitz, Norman & Sidney J. Segalowitz, S. 1993. Skilled performance, practice and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics* 14. 369-385.
- Skehan, Peter. 2003. Task based instruction. *Language Teaching* 36. 1-14.
- Talmy, Leonard. 2008. Aspects of attention in language. In Peter. Robinson & Nick C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 27-38). New York: Routledge.
- Taube-Schiff, Marlene. & Norman Segalowitz. 2005. Within-language attention control in second language processing. *Bilingualism: Language and Cognition* 8. 195-206.
- Tavakoli, Parvaneh. 2011. Pausing patterns: differences between L2 learners and native speakers. *ELT Journal* 65(1). 71-79. doi:10.1093/elt/ccq020.
- Tavakoli, Parvaneh. & Pauline Foster. 2008. Task design and second language performance: the effect of narrative type on learner output. *Language Learning* 58. 439-473.
- Tavakoli, Parvaneh. & Peter Skehan. 2005. Strategic planning, task structure, and performance testing. In Rod Ellis (ed.), *Planning and task performance in a second language*, 239-273. Amsterdam: John Benjamins.
- Tomasello, Michael. 2003. *Constructing a language*. Cambridge, MA: Harvard University Press.
- Tulving, Endel & Donald Thomson. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychological Review* 80. 352-373.
- Tversky, Amos. & Daniel Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5. 207-232.
- Unkelbach, Christian. & Rainer Greifeneder (eds.). 2013. *The experience of thinking: How the fluency of mental processes influences cognition and behaviour*. London: Psychology Press.
- Williams, John. 2011. Working memory and SLA. In Susan M. Gass & Alison Mackey (eds.), *The handbook of second language acquisition*, 427-441. New York: Routledge.
- Wing, Erik, Maureen Ritchey & Roberto Cabeza. 2015. Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. *Journal of Cognitive Neuroscience* 27(4). 679-691. doi:10.1162/jocn\_a\_00740

#### ACKNOWLEDGEMENT

The author would like to thank the Social Sciences and Humanities Research Council of Canada and the Australian Research Council for grants supporting this work in part. I

also would like to thank Elizabeth Gatbonton, Yeji Han, Momyka Rodrigues, and an anonymous reviewer for their constructive feedback on an earlier draft. The author may be contacted at [norman.segalowitz@concordia.ca](mailto:norman.segalowitz@concordia.ca).

# Ladders and Snakes in Second Language Fluency

PETER SKEHAN, PAULINE FOSTER, AND SABRINA SHUM

## Abstract

*This article reports a study comparing first and second language fluency during narrative retelling tasks of varying degrees of tightness in structural organisation, exploring in particular a distinction between discourse-based and clause-based fluency. We argue that positive and negative influences on fluency are linked to the Conceptualiser and Formulator stages of Levelt's model of speaking. Task structure and degree of subordination, which were related to greater fluency for both native and non-native speakers, are Conceptualiser and discourse oriented. Formulaic language, which was also related to fluency, is more Formulator and clause oriented. Contrastingly, higher lexical sophistication and longer clauses are associated with clause-linked fluency problems, but only for native speakers.*

Keywords: L2 fluency; second language fluency; discourse-based fluency; clause-based fluency

## 1. Introduction

Based on Levelt's three-stage model of speaking (Levelt 1999; Kormos 2006) we will assume that speaking fluently involves:

- *knowing what you want to say,*
- *having the means to say it,*
- *not changing your mind,*
- *anticipating problems effectively.*

*Knowing what you want to say* requires concepts to be selected and organised as a pre-verbal message (Conceptualisation), which is input to the next stage (Formulation) which involves *having the means to say it*. Lemmas are retrieved from the mental lexicon to build syntactic and phonological frames, and these are converted into speech (Articulation). If these stages operate smoothly and ideas are translated into spoken language without mishap, fluency results.

*Not changing one's mind* obviously supports the fluent delivery of a message. However, as messages are delivered they remain under review. The speaker may realise that a

formulation can be improved, or a non-verbal reaction from the interlocutor (a raised eyebrow, the absence of a nod) could signal that the message needs to be reconsidered. In such cases, fluency may be disrupted. Similarly, *anticipating problems effectively* (e.g. difficulties in lemma retrieval) maintains fluency, just as *not* doing so can make speakers prone to dysfluency.

For clarity, the above analysis covers just one cycle of operation from Conceptualisation to Articulation. Obviously, normal communication proceeds through successive cycles of parallel processing. If this were not the case, speech would necessarily be filled with pauses between stages. But in ordinary discourse, conceptualising a prelinguistic message proceeds while a previous one is in the process of being formulated and then articulated. Fluency therefore rests on a harmonious and balanced relationship between the three major components of the model, operating simultaneously and in parallel. In first language performance at least, lemma access presupposes the speaker has a mental lexicon rich in size, in elaboration, in organisation, in speed of access, and in availability of a repertoire of formulaic language, all of which enable a smooth process of syntax building and articulation.

A key difference in second language (SL) performance is the mental lexicon it draws upon, which is smaller, less organised, likely slower in access, less elaborated with syntactic and collocational information, and contains a narrower repertoire of formulaic language (Bolibaugh and Foster 2013). Consequently, the demands of the pre-verbal message cannot be met so easily, or even at all. Moreover, the Formulator is likely to be demanding of attentional resources, to the extent that there is little remaining for the Conceptualiser or the Articulator to work with in parallel mode. This means the SL speaker must switch to serial processing, where problems at one stage must be solved before parallel processing can resume (Kormos 2006; Skehan 2014b). This leads to a further consequence – the need to have resources to recapture a parallel processing, and re-launch fluent speech production.

If we relate the earlier discussion of fluency to stages of speech production, two aspects emerge which are the focus of this article. The first concerns dimensions of fluency – how it can be broken down into sub-components. The second concerns influences on fluency (or more often, dysfluency). As we will see, a comparison between native and non-native performance is revealing in each of these cases, regarding the positive and negative influences, the ladders and snakes of our title.

Previous approaches to the dimensions of fluency in L2 speaking (e.g. Tavakoli and Skehan 2005) suggested a three-way separation between breakdown, repair, and speed fluency. Here we propose instead a distinction between discourse-level issues and AS-unit<sup>1</sup>/clause level issues, with speed continuing to be a largely separate factor (Wang

---

<sup>1</sup> We use the term ‘AS-unit’ defined as “an independent clause, or sub-clausal unit, together with any dependent clauses associated with either” (Foster, Tonkyn and Wigglesworth 2000). For our purposes,

and Skehan in prep.) Clause<sup>2</sup> processing takes place *within* a clause and the focus is on lexical choices, connecting to Formulator and Articulator stages. Discourse-level issues concern to processing *above* the level of the clause, and problems that occur in conjoining units and developing an argument with them. These are more likely to be associated with Conceptualiser operations. We will argue that this distinction (discourse/clause) makes more sense psycholinguistically than the breakdown/repair distinction.

This view of dimensions of fluency suggests that dysfluencies occurring at clause boundaries are associated with Conceptualiser-linked discourse decisions and are evidence of macro-planning. By contrast, within-clause dysfluencies have a Formulator origin and are evidence of micro-planning. The distinction between discourse dysfluency and clause dysfluency is shown in Table One.

*Table 1. Processes and measurement options in discourse and clause dysfluency*

| Discourse dysfluency   | Clause dysfluency  |
|--|--|
| <ul style="list-style-type: none"> <li>- filled and unfilled end-clause pauses;</li> <li>- linked to the Conceptualiser;</li> <li>- evidence of discourse decisions;</li> <li>- indicates macro-planning.</li> </ul> | <ul style="list-style-type: none"> <li>- filled and unfilled mid-clause pauses, reformulations and false starts;</li> <li>- linked to the Formulator and Articulator;</li> <li>- evidence of language decisions;</li> <li>- indicates micro-planning.</li> </ul> |

The fluency link to macro- or micro-planning provides a more satisfactory base for comparing native speaker (NS) and non-native speaker (NNS) performance. Everyone punctuates the speech stream with pauses that do not arise from processing blips. So while we expect both NSs and NNSs to pause, the difference suggested by the discourse and clause distinction is that NSs will pause more for macro-planning and less for micro-planning. Indeed, previous research suggests as much. A synthesis of seven research studies (Skehan and Foster 2008) showed that NSs in these studies paused at least as often as NNSs at AS unit boundaries, and far less often mid-clause.

Viewing fluency through the discourse vs. clause contrast unlocks a different perspective on the *influences* upon fluency, such as task structure. Several studies have explored how different types of narrative structure impact on performance. Typically, structure is regarded as involving degree of tightness in the organisation of the narrative. Tavakoli

---

AS-unit is a more precise term than 'sentence' as it can capture the fragmentary nature of spoken language.

<sup>2</sup> AS-units can contain one clause only, or more than one clause. For convenience, from now on, we use the term 'clause' in this article, but this can refer to the clauses in a one clause AS-unit, or a multi-clause AS-unit.

and Skehan (2005) contrasted very loose narratives (i.e. with a rather general beginning-middle-end organisation) with different and tighter degrees of problem-solution discourse structure (Hoey 1983). Broadly, increasing degrees of structure in this way were associated with greater accuracy and fluency. Skehan and Foster (1999) report similar effects with video-retelling tasks. A tight narrative structure has also been linked to greater fluency (Skehan and Foster 1999; Skehan and Shum 2014) because it lends itself more easily to macro-planning. One can also explore how fluency is influenced by the Formulator and micro-planning. For example, access to a wide repertoire of formulaic language, may have a considerable 'easing' effect on processing (Foster 2001).

The discussion so far points to three general research questions:

RQ 1: How justifiable is it to discuss fluency in terms of the Discourse vs. Clause distinction? Specifically: (a) does language complexity in performance impact differently upon discourse-linked fluency and clause-linked fluency; and (b) does lexical complexity in performance impact differently upon discourse and clause-linked fluency?

RQ 2: How does task structure impact upon fluency? Specifically, does task structure influence discourse-linked and clause-linked fluency in different ways?

RQ 3: How does a NS/NNS comparison illuminate the nature of SL fluency? Specifically, for the NSs and NNSs, are there relationships between on the one hand language complexity, lexical complexity, and structure, and on the other between discourse and clause-linked fluency?

## **2. The research study**

### *2.1 The narrative tasks*

The study used four Mr. Bean video excerpts, of 5-7 minutes each. They were chosen because Mr Bean has international appeal, and almost no spoken language. The four selections represented increasing degrees of narrative structure, from loose to tight, established by a combination of analysis by the authors and ratings of ELT professionals.

1. *Crazy Golf*: Mr. Bean plays a round of Crazy Golf. A series of unconnected misadventures ensues. This narrative has a beginning and an end but no tight structure between these points.
2. *Christmas*: Mr. Bean meets his girlfriend on Christmas Eve. She sees a ring she would like. On Christmas Day, his girlfriend arrives, and he gives her the picture and a hook he saw in the jewellery store. This narrative has a clear beginning, middle and end, but no other tight narrative links.

3. *Funfair*: Going to a funfair, Mr. Bean's car gets hooked to a baby's pram. He 'parks' the baby in a rocking car. He goes on various rides and comes back to find the baby still crying. He buys helium balloons to amuse the baby, which carry the pram up into the sky. He bursts the balloons with an arrow, and the pram descends to earth. This narrative has a clear beginning and end, and some causal links between the various parts..
4. *Thief*: In a park, Mr. Bean fails to take a photo of himself, and recruits a passer-by who steals his camera. He searches for the thief, finds him, but the thief escapes. Later, at the police station, Mr Bean identifies the thief in an amusing manner. This narrative was deemed to have the tightest structure and the strongest causal links between each part of the story.

## 2.2. Participants

Data were gathered from 28 NNSs and 28 NSs of English. The NNSs were 15 female and 13 male second-year students at a university in Southern China, age range 19 to 22, (mean = 21). Their proficiency, based on their College English Test scores, was low intermediate. They were able to do comparable tasks as part of their English classes. All had L1 Cantonese or Mandarin. The NSs were 15 female and 13 male international students at a university in Hong Kong. They ranged from 21 to 32 (mean = 26). They were familiar with narrative retellings.

## 2.3. Procedure

One-on-one meetings were arranged between a researcher and each participant. It was explained what was to be done. The videos were shown on a computer screen, and the participant watched each story and narrated it afterwards. Instructions were provided on the computer screen. All participants completed all four narratives, in a counter-balanced order to control for sequence effects.

## 2.4. Data processing

Each narrative was transcribed in modified CHAT format (MacWhinney 2000). Each AS unit was represented in three lines. The first was the CHAT line. The second contained start-finish timing information for the AS unit. The third line was coded according to TaskProfile<sup>3</sup> conventions.

---

<sup>3</sup> TaskProfile (Skehan 2014b) is software written to produce automatically-generated scores for task-based second language performance.

## 2.5. Measures

There are two independent variables: *structure of narrative* and *speaker type*. The former has four values: a) no structure (*Golf*), b) a clear beginning-middle-end structure (*Christmas*), c) loose problem-solution structure (*Funfair*), d) tight problem-solution structure (*Thief*). It is a within-subject variable. The independent between-subjects variable of speaker type contrasted NSs and NNSs.

The dependent variables comprised measures of fluency, and structural and lexical complexity<sup>4</sup> (see Table Two). Pauses were defined as an interruption to the speech flow of more than 400 milliseconds, following previous research using comparable data, and human (as opposed to machine) identification of pauses (e.g. Tavakoli and Skehan 2005). Filled pauses were any sound element lacking referential meaning ('uh', 'um'.) Reformulations were syntactic or morphological changes to an utterance. Repetitions were the number of times a word or phrase was repeated.

Structural complexity was calculated by dividing the data into syntactic clauses and AS-units (Foster et al. 2000) and expressed as the ratio of clauses to AS-units. The second complexity measure, (Norris and Ortega 2009), was of the number of words per clause, reflecting the internal complexity within clauses.

Lexis was measured in two ways. Lexical diversity was measured through the VocD subprogram (Malvern and Richards 2002). This captures the extent to which the speaker uses a variety of words within the sample. Lexical sophistication (Read 2000) concerns the selection of less common words, defined through frequency lists. In the present case, following Meara and Bell (2001), a version of PLEX was used. This computes a value, Lambda, which reflects the 'penetration' in a short text of less frequent lexical items (Skehan 2009).

*Table 2. Summary of the variables used in the study*

| <b>Fluency</b>                       |  |
|--------------------------------------|--|
| <i>Discourse</i>                     |  |
| Clause-boundary pauses per 100 words | Number of pauses greater than 0.4 sec. at clause boundary position, standardised |
| <i>Clause</i>                        |  |
| Mid-clause pauses per 100 words      | Number of pauses greater than 0.4 sec. in mid-clause position                    |
|                                      | Number of times a word or expression was   |

<sup>4</sup> Measures of structural and lexical complexity are dependent variables regarding structure and speaker status since one could compare, for example, lexical complexity as a function of task structure. But they also function as possible influences on fluency, and so they also have an independent variable role with respect to the various fluency measures.

|                              |  |
|------------------------------|--|
| Reformulations per 100 words | reformulated   |
| Repetition per 100 words     | Number of times a word or expression was repeated<br>Number of times pauses such as 'um', 'ah' were used |
| Filled pauses per 100 words  |  |
| <b>Structural Complexity</b> |  |
| Subordination index          | Number of clauses divided by number of AS-units  |
| Words per clause             | The average number of words used per clause  |
| <b>Lexis</b>                 |  |
| Lexical diversity            | The value D produced by the program VocD, a length-corrected type-token ratio.                           |
| Lexical sophistication       | The value Lambda (Meara and Bell, 2001), an index of the use of lower frequency words                    |

## 2.6. Analyses

There are two independent variables in this study; *structure*, a within-subject variable with four values; and *speaker type*, a between-subjects variable with the two values. There are nine dependent variables: two measures of complexity, two of lexis, and five of fluency. Simple correlations assessed the strength of relationship between the measures of structural and lexis, and the measures of fluency. One-way repeated measures ANOVAs were conducted to explore the impact of task structure, with post-hoc testing where appropriate. The data for all variables were within acceptable parameters for normality of distribution.

## 3. Results

First we will explore the data relevant to Research Question 1, which focusses on the influences of lexis and structural complexity on discourse and clause-based fluency measures. The basic descriptive information on the fluency measures is provided, for reference, in Table 5 below, which we will turn to in detail later. Our focus here is the association between the complexity and fluency measures. The data for structural complexity are shown in Table Three.

*Table 3. Correlations between Structural Complexity and Fluency Measures*

| Fluency Measure | Structural Complexity Measure: Subordination Index |            |        |            |        |
|-----------------|--|------------|--------|------------|--------|
|                 | Reform. per 100                                    | Repet. per | Filled | Mid-clause | Clause |
|                 |  |            |        |            |        |

|                  | words        |              | 100 words |      | pauses per<br>100 words |           | pauses per<br>100 words |           | boundary<br>pauses per<br>100 words |      |
|------------------|--------------|--------------|-----------|------|-------------------------|-----------|-------------------------|-----------|-------------------------------------|------|
| Task             | NS           | NNS          | NS        | NNS  | NS                      | NNS       | NS                      | NNS       | NS                                  | NNS  |
| <i>Golf</i>      | -0.41<br>**  | -0.63<br>*** | 0.24      | 0.23 | 0.16                    | 0.33      | 0.08                    | 0.34      | 0.08                                | 0.2  |
| <i>Christmas</i> | -0.32        | -0.50<br>**  | 0.23      | 0.14 | 0.24                    | 0.28      | 0.11                    | 0.37      | 0.44<br>*                           | 0.11 |
| <i>Funfair</i>   | -0.58<br>*** | -0.40<br>*   | 0.25      | 0.16 | 0.1                     | 0.3       | 0.17                    | 0.43<br>* | -0.11                               | 0.01 |
| <i>Thief</i>     | -0.54<br>**  | -0.53<br>**  | 0.24      | 0.31 | 0.14                    | 0.39<br>* | 0.44<br>*               | 0.45<br>* | -0.51<br>**                         | 0.17 |

| Structural Complexity Measure: Words per clause |                          |             |                         |       |                                   |      |                                       |            |   |       |
|---|--------------------------|-------------|-------------------------|-------|-----------------------------------|------|---------------------------------------|------------|---|-------|
| Fluency Measure                                 | Reform. per<br>100 words |             | Repet. per<br>100 words |       | Filled<br>pauses per<br>100 words |      | Mid-clause<br>pauses per<br>100 words |            | Clause<br>boundary<br>pauses per<br>100 words |       |
| Task  | NS                       | NNS         | NS                      | NNS   | NS                                | NNS  | NS                                    | NNS        | NS  | NNS   |
| <i>Golf</i>                                     | -0.57<br>**              | -0.09       | 0.42<br>*               | 0.2   | 0.17                              | 0.02 | 0.40<br>*                             | 0.16       | 0.07  | 0.27  |
| <i>Christmas</i>                                | -0.14                    | -0.32       | 0.51<br>**              | 0.12  | 0.51<br>**                        | 0.18 | 0.68<br>***                           | 0.14       | 0.44<br>*                                     | 0.01  |
| <i>Funfair</i>                                  | 0.12                     | -0.44<br>*  | 0.34                    | -0.07 | 0.46<br>**                        | 0.13 | 0.48<br>**                            | 0.09       | 0.61<br>**                                    | -0.22 |
| <i>Thief</i>                                    | -0.45<br>*               | -0.53<br>** | 0.48<br>**              | 0.09  | 0.39<br>*                         | 0.14 | 0.41<br>*                             | 0.46<br>** | 0.45<br>*                                     | -0.21 |

\* =  $p < .05$ : \*\* =  $p < .01$ : \*\*\* =  $p < .001$ : N for NS and NNS = 28.

As Table 3 shows, subordination is related to a reduction in the number of pauses at clause-boundary points, for both NNS and NS speakers on all tasks; correlations range from a low of -0.41 to a high of -0.63, with a median correlation of -0.54. If speakers produce multi-clausal utterances, they tend to pause less often.

The other measure of complexity, words per clause, is associated, for the NSs only, with an increase in reformulation, repetition, and filled pauses, and mid-clause pausing. Of the sixteen possible correlations, thirteen reach significance, with a highest value of 0.68. In other words, the longer the clause, the more likely that some form of repair will

be required.

The results for the lexical measures are shown in Table Four. Lexical diversity (VocD) has no significant correlations for the NSs. The NNS data shows much stronger relationships here, with less repair, generally and slightly increased clause-boundary pausing in two of the tasks. In other words, greater lexical diversity is associated with greater fluency. The reduction in repair is striking, with many correlations reaching significance and some being especially high, e.g. for greater lexical diversity and fewer filled pauses (showing correlations of 0.61 [*Golf*], 0.79 [*Christmas*], 0.76 [*Funfair*] and 0.80 [*Thief*]). This indicates a relationship in the NNS data between avoiding recycling words and avoiding repair; speakers can comfortably handle the demands of task processing, using varied words with little need for repair, or conversely they can recycle smaller word sets with a greater need for repair.

Table 4. Correlations between lexical and fluency measures

| Lexical Measure  | Lexical Diversity     |            |                      |              |                             |              |                                 |              |                                      |       |
|------------------|-----------------------|------------|----------------------|--------------|-----------------------------|--------------|---------------------------------|--------------|--------------------------------------|-------|
|                  | Reform. per 100 words |            | Repet. per 100 words |              | Filled pauses per 100 words |              | Mid-clause pauses per 100 words |              | Clause boundary pauses per 100 words |       |
| Fluency Measure  | NS                    | NNS        | NS                   | NNS          | NS                          | NNS          | NS                              | NNS          | NS                                   | NNS   |
| Task             | NS                    | NNS        | NS                   | NNS          | NS                          | NNS          | NS                              | NNS          | NS                                   | NNS   |
| <i>Golf</i>      | -0.18                 | 0.21       | 0.01                 | -0.61<br>*** | -0.3                        | -0.54<br>**  | -0.07                           | -0.61<br>*** | -0.24                                | -0.24 |
| <i>Christmas</i> | 0.43<br>*             | 0.44<br>*  | -0.15                | -0.52<br>**  | -0.38                       | -0.60<br>*** | -0.27                           | -0.79<br>*** | -0.21                                | -0.29 |
| <i>Funfair</i>   | -0.1                  | 0.05       | -0.14                | -0.58<br>*** | -0.45<br>**                 | -0.59<br>*** | -0.36                           | -0.76<br>*** | -0.49<br>**                          | -0.31 |
| <i>Thief</i>     | 0.16                  | 0.52<br>** | -0.27                | -0.55<br>**  | -0.2                        | -0.67<br>*** | -0.24                           | -0.80<br>*** | -0.15                                | -0.15 |

| Lexical Measure  | Lexical Sophistication |       |                      |       |                             |      |                                 |       |                                      |       |
|------------------|------------------------|-------|----------------------|-------|-----------------------------|------|---------------------------------|-------|--------------------------------------|-------|
|                  | Reform. per 100 words  |       | Repet. per 100 words |       | Filled pauses per 100 words |      | Mid-clause pauses per 100 words |       | Clause boundary pauses per 100 words |       |
| Fluency Measure  | NS                     | NNS   | NS                   | NNS   | NS                          | NNS  | NS                              | NNS   | NS                                   | NNS   |
| Task             | NS                     | NNS   | NS                   | NNS   | NS                          | NNS  | NS                              | NNS   | NS                                   | NNS   |
| <i>Golf</i>      | 0.47<br>**             | -0.08 | -0.07                | 0.01  | -0.27                       | 0.11 | -0.37                           | 0.02  | -0.39<br>*                           | -0.27 |
| <i>Christmas</i> | 0.46                   | 0.14  | -0.35                | -0.22 | -0.28                       | 0.01 | -0.22                           | -0.15 | -0.07                                | -0.22 |

|                |      |      |       |       |       |       |       |       |       |       |
|----------------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|
|                |      |      | **    |       |       |       |       |       |       |       |
| <i>Funfair</i> | 0.38 | 0.23 | -0.41 | 0.13  | -0.35 | 0.22  | -0.40 | 0.2   | -0.24 | 0.18  |
|                | *    |      | *     |       |       |       | *     |       |       |       |
| <i>Thief</i>   | 0.12 | 0.01 | 0.12  | -0.32 | 0.24  | -0.06 | 0.04  | -0.14 | 0.19  | -0.38 |
|                |      |      |       |       |       |       |       |       |       | *     |

\* =  $p < .05$ : \*\* =  $p < .01$ : \*\*\* =  $p < .001$ : N for NS, NNS = 28.

The results for lexical sophistication contrast with this. There are hardly any significant relationships for the NNS group, suggesting that if NNSs use less frequent words, they are able to access lemma-based information without sacrificing fluency. In contrast, there are effects with the NS group. Repair and mid-clause pauses show no relationships to Lambda. But there are relationships between Lambda and the number of clause-boundary pauses. In other words, for the NS group, less frequent lexical items are associated with more end-clause pausing.

For task structure, the relevant descriptive statistics are shown in Table Five. Following previous research we predicted for the NNSs that greater narrative structure would be associated with greater fluency. The data shows that only two fluency measures were related to task structure: clause-boundary pauses and reformulation. A repeated measures one-way ANOVA, with structure as the only factor, showed that the clause-boundary pauses measure was significant for the NSs ( $p < .01$ ) and NNSs ( $p < .001$ ). For the NSs, post-hoc testing showed that the *Funfair* score was lower than *Golf* and *Christmas*, and *Thief* was significantly lower than *Christmas*, indicating a reasonable (if not perfect) relationship with degree of structure. For the NNSs, the most structured task, *Thief*, was performed significantly more fluently than the other three, which did not differ from one another. For the one other significant result, NNSs reformulation, post-hoc testing located this as a difference between *Golf*, the least structured narrative, and *Funfair*, the third most structured.

Table 5. Descriptive Statistics for Fluency Measures by Structure

| Task        | Fluency Measures        |                |                       |                |                            |                |                             |                  |                                 |                 |
|-------------|-------------------------|----------------|-----------------------|----------------|----------------------------|----------------|-----------------------------|------------------|---------------------------------|-----------------|
|             | AS Pauses per 100 words |                | Reform. per 100 words |                | Repetitions. per 100 words |                | Filled pauses per 100 words |                  | Mid-clause pauses per 100 words |                 |
|             | NS                      | NNS            | NS                    | NNS            | NS                         | NNS            | NS                          | NNS              | NS                              | NNS             |
| <i>Golf</i> | 5.20<br>(1.63)          | 7.59<br>(2.37) | .61<br>(.61)          | 3.17<br>(2.19) | 1.63<br>(1.69)             | 9.21<br>(7.78) | 2.55<br>(2.82)              | 10.13<br>(10.78) | 2.66<br>(1.58)                  | 12.43<br>(7.68) |

|                  |                |                |               |                |                |                |                |                  |                |                 |
|------------------|----------------|----------------|---------------|----------------|----------------|----------------|----------------|------------------|----------------|-----------------|
| <i>Christmas</i> | 5.85<br>(2.43) | 7.94<br>(2.70) | .68<br>(.72)  | 3.56<br>(2.01) | 1.51<br>(1.67) | 8.51<br>(8.17) | 2.7<br>(2.51)  | 11.15<br>(12.17) | 2.55<br>(1.48) | 11.61<br>(6.51) |
| <i>Funfair</i>   | 4.41<br>(2.04) | 7.38<br>(3.02) | .67<br>(1.01) | 4.65<br>(2.82) | 1.62<br>(1.74) | 9.92<br>(7.65) | 2.20<br>(2.65) | 10.83<br>(11.33) | 2.11<br>(1.36) | 11.62<br>(7.51) |
| <i>Thief</i>     | 4.77<br>(1.83) | 5.82<br>(2.66) | .51<br>(.48)  | 3.84<br>(2.55) | 1.49<br>(1.37) | 9.53<br>(7.23) | 2.15<br>(1.98) | 10.62<br>(10.50) | 2.36<br>(1.65) | 10.85<br>(7.20) |

Mean scores are shown, with standard deviations in parentheses: N =28, for both NS and NNS groups.

Although not central to an analysis of fluency, it is intriguing that narrative structure was related to greater structural complexity, as shown in Table Six. The mean scores for the subordination measure for the NNSs were 1.30 (*Golf*), 1.30 (*Christmas*), 1.37 (*Funfair*) and 1.53 (*Thief*). The first two values did not differ significantly from one another, but they did from the third, *Funfair* ( $p < .05$ ), which was also significantly and appreciably less than the fourth, *Thief* ( $p < .001$ ). The NSs data reflects a similar trend. Interestingly, the reverse trend occurs with the other measure of complexity, words per clause. Here the NNSs values are 6.53 words per clause (*Golf*), 5.54 (*Christmas*), 5.64 (*Funfair*) and 5.38 (*Thief*). The value for *Golf* is significantly greater ( $p < .001$ ) than those for *Christmas* and *Funfair* (which do not differ), but all values are significantly greater than the 5.38 for *Thief* ( $p < .001$ : *Golf*;  $p < .05$ : *Christmas*, *Funfair*). These results too are (broadly) reflected mirrored in the NS data.

Table 6. Complexity scores as a function of task structure

| Task             | Subordination measure |                | Words per clause measure |                |
|------------------|-----------------------|----------------|--------------------------|----------------|
|                  | NS                    | NNS            | NS                       | NNS            |
| <i>Golf</i>      | 1.39<br>(0.26)        | 1.30<br>(0.13) | 6.02<br>(0.48)           | 6.53<br>(1.48) |
| <i>Christmas</i> | 1.38<br>(0.16)        | 1.30<br>(0.13) | 5.17<br>(0.42)           | 5.54<br>(0.39) |
| <i>Funfair</i>   | 1.52<br>(0.23)        | 1.37<br>(0.12) | 5.43<br>(0.49)           | 5.64<br>(0.59) |
| <i>Thief</i>     | 1.53<br>(0.19)        | 1.53<br>(0.14) | 5.30<br>(0.40)           | 5.38<br>(0.48) |

#### 4. Discussion

Three themes underlie the discussion in this section: discourse-level and clause-level fluency contrast; the portrayal of positive and negative factors, referenced by the 'ladders and snakes' of our title; and the relationship of our findings to the Levelt Model for native and non-native speakers.

Two issues emerge regarding influences on discourse-level fluency. Positively, there are the influences of subordination and narrative structure on clause-boundary pausing; more structured tasks and greater syntactic subordination are associated with less pausing at boundaries. The broad arc of a tightly structured narrative means the speaker does not need to pause to muster resources for the next speech plan. Additionally, subordination within AS-units might support fluency because the speaker is oriented to a larger discourse unit. But there is also a negative influence, a 'snake' in terms of our title. Greater lexical sophistication is related to more frequent and longer clause-boundary pauses for the NSs group. Accessing less frequent lexis seems to manifest its processing load at the end of the syntactic plan as the speaker regroups. Rather curiously, the NNSs show no relationship between lexical sophistication and any pausing location. This might well be related to the fact that NNSs have a smaller mental lexicon to draw on, and also that they are aware of their limitations and work within them.

We turn next to clause-level influences, and will consider the ability to draw on a repertoire of formulaic language, a 'ladder', and the consequence of using longer clauses, a 'snake'. Regarding formulaic language, Pawley and Syder (1983) point out that while the potential for infinite originality exists in speech, it would come at the cost of very frequent pausing to allow time for original combinations to be assembled. But idiomatic 'chunks' of language are acquired and stored as single entries in a speaker's lexicon: stringing these together to make utterances takes less time and effort than stringing together individual words, thus easing Formulator operations. Pawley and Syder describe NS speech as constructed one clause at a time, and memorised sequences are a requirement for being able to deliver a fluent stream of speech.

“We may speak, then, of a 'one clause at a time facility' as an essential constituent of communicative competence [.....]; the speaker must be able regularly to encode whole clauses, in their full lexical detail, in a single encoding operation and so avoid the need for mid-clause hesitations.” (Pawley and Syder 1983:204)

Aligning this with our title, we note there is a board game called Snakes and Ladders, but no board game called Ladders and Snakes. The phrase 'ladders and snakes' comprises three lexemes put together for the nonce, and requires three unrelated selections from the lexicon. 'Snakes and Ladders' on the other hand, is only one lexeme, requiring a

single selection that does not allow individual preference over which noun goes first. It follows that the greater the store of formulaic chunks a NNS has acquired in an L2, the more likely he or she is to be able to produce fluent clauses. Yet, there is also a negative influence, or 'snake', in the use of longer clauses by NSs in our data. There is a clear relationship between longer clauses and more repair and clausal breakdown indices which is not shown in the NNS data, possibly because these speakers are aware of their limits and work within them.

Relating these findings to the Levelt model complements the title of this article as it references positive and negative factors in SL fluency. We will consider one positive and one negative factor each for Conceptualisation and Formulation (and the factors apply to both NS and NNS groups, unless specified otherwise). For Conceptualisation, the positive factor will be the association of subordination and clause-boundary pausing, and the negative factor is the association, for the NS group, between clause-boundary pausing and lexical sophistication. The discussion attempts to bring out the fruitfulness of using such a speech production model to understand fluency effects. (In passing, it might be noted that lexical measures and syntactic complexity do not correlate in this data.)

A key issue concerns the units of speech production. Earlier we discussed discourse vs. clause-level processing. The data reported here fit in with that interpretation well. Positively, it seems as if the Conceptualiser is able to offer clause and AS-unit plans which span the entire unit, or even more than one unit. These plans then sustain performance at a general, discourse level, and so pre-empt the need for clause-boundary pauses. Ideas, if they have any internal complexity and structure, seem a good way to underpin and even drive forward fluency, and to scaffold narrative performance as the connected ideas they embody link series of AS-units to one another. It is important to note, though, that task structure may have a slight confounding effect here and part of the subordination-complexity effect may be attributable to the underlying effect of structure on both aspects of performance.

In contrast, there is the negative influence of higher lexical sophistication scores for the NS group. It is interesting here that using such lower frequency items is not related to dysfluency within clauses, but appears to defer this to the clause boundary. The use of such lower frequency lexical items has no impact on the performance of the NNS group. It is not clear why Lambda is associated with clause-boundary pausing in the present NSs dataset. This interesting effect seems to implicate the Conceptualiser, but for unclear reasons, suggesting future research is needed here.

Next we turn to the Formulator influences. Regarding the positive influence, there is an association between VocD (lexical diversity) and less repair for the NNSs. The less the NNSs recycle the same words, the fewer reformulations and repairs that they produce. This seems counter-intuitive at first sight but this association may be a reflection of effective proficiency. NNSs who are more proficient, and more able to mobilise that

proficiency, are able to draw on the lexicon that they have in a more efficient manner. Those who are less proficient, in contrast, are not able to access such a wide range of words, and are more often in trouble with those that they can access.

We have noted a relationship in the NS data between longer clauses, on the one hand, and greater repair and mid-clause pausing on the other. Yet this relationship does not occur for the NNSs. Perhaps the key, though, is a feature of fluency that we have not emphasised here – speed. The NSs and NNSs differ dramatically on this index, producing respective means of 135 and 69 words-per-minute (a very significant difference, with a huge effect size). A possible explanation is that the NS group are speaking at speed and the greater repair reflects them having pushed their speaking resources to the limit, whereas the NNS group were more conservative.

## 5. Conclusions and Limitations

These generalisations about positive and negative influences are intriguing, and offer a more nuanced view of how NNSs achieve fluency in their performance. They also demonstrate that significant insights can derive from research designs where NSs and NNSs do the same tasks, so that a NS baseline teases out which performance features are the result of the task and which arise from limited language resources (e.g. Foster 2001; Foster and Tavakoli 2009).

Three limitations mean that the interpretations and conclusions drawn here need to be treated with caution. First, the data comparing NSs and NNSs used a between-subjects design, and therefore individual variability is involved. Further research comparing NS and NNS performance in the same individuals would avoid this. Second, this study sampled mid-intermediate level NNSs; information about performance on comparable tasks of high intermediate or advanced students is needed to sustain more robust claims. Finally, there is the issue of task. Video-based material has implications for time-pressure which picture narratives do not. Beyond this particular comparison there is the need for tasks other than narratives. The research literature has used a variety of task types performed under a range of pre-task, during-task and post-task conditions. As a result, generalising too much from the present data would be premature.

## References

- Bolibaugh, Cylcia & Pauline Foster. 2013. Memory-based aptitude for nativelike selection: The role of phonological short-term memory. In Gisela Granena & Michael Long (eds.) *Sensitive periods, language aptitude, and ultimate L2 attainment*. 205-230..Amsterdam: John Benjamins.
- Foster, Pauline. 2001. Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In Martin Bygate, Peter Skehan and Merrill Swain (eds.) *Researching Pedagogic Tasks: Second language*

- learning, teaching, and testing*. 75-93. London: Longman.
- Foster, Pauline, Alan Tonkyn & Gillian Wigglesworth. 2000. Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354–375.
- Foster, Pauline & Parvaneh Tavakoli. 2009. Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning*, 59, 866-896.
- Hoey, Michael 1983. *On the Surface of Discourse*. London: George Allen and Unwin.
- Kormos, Judit. 2006. *Speech production and second language acquisition*. Mahwah, N.J.: Lawrence Erlbaum.
- Levelt, William. 1999. Producing spoken language: a blueprint of the speaker. In C. Brown and P. Hagoort (eds.). *Neurocognition of language*, 83-122. Oxford: Oxford University Press.
- MacWhinney, Brian. 2000. *The CHILDES Project: Tools for analyzing talk*. 3<sup>rd</sup> edn. Mahwah, NJ: Lawrence Erlbaum Associates.
- Malvern, David & Brian Richards. 2002. Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19, 85-104.
- Meara, Paul & Huw Bell. 2001. P\_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, 16, 5-19.
- Norris, John & Lourdes Ortega. 2009. Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30, 555-578.
- Pawley, Andrew & Frances Syder. 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In Jack Richards & Richard Schmidt, (eds.). *Language and communication*, 191-227. London: Longman.
- Read, John. 2000. *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Skehan, Peter. 2014a. (ed.). *Processing Perspectives on Task Performance*, Amsterdam: John Benjamins
- Skehan, Peter. 2009. Lexical Performance by Native and Non-native Speakers on Language-Learning Tasks. In Brian Richards, Helmut Daller, David Malvern & Paul Meara (eds.). *Vocabulary Studies in First and Second Language Acquisition: The interface between theory and application*. 107-124. London: Palgrave Macmillan.
- Skehan, Peter. 2014b. Limited attentional capacity, second language performance, and task-based pedagogy. In Peter Skehan. (ed.). 211-260. *Processing Perspectives on Task Performance*, Amsterdam: John Benjamins.
- Skehan, Peter & Pauline Foster. 1999. The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49, 93-120.
- Skehan, Peter & Pauline Foster, P. 2008. Complexity, accuracy, fluency and lexis in task-based performance: a meta-analysis of the Ealing research. In Van Daele, S., Housen, A., Kuiken, F., Pierrard, M., and Vedder, I. (eds.). *Complexity, Accuracy, and Fluency in Second Language Use, Learning, and Teaching*. Brussels: University of Brussels Press.
- Skehan, Peter & Sabrina Shum. 2014. Structure and processing condition in video-based narrative retelling. In Peter Skehan, (ed.). 187-210. *Processing Perspectives on Task Performance*, Amsterdam: John Benjamins.

- Tavakoli, Parvaneh. & Peter Skehan, P. 2005. Planning, task structure, and performance testing. In Rod Ellis R. (ed.) *Planning and Task Performance in a Second Language*. 239-273. Amsterdam: John Benjamins.
- Wang, Z. & Peter Skehan. (in preparation). The effects of time pressure and L2 proficiency level on speaking performance.

# Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency

NIVJA H DE JONG

## Abstract

*This paper compares the distribution of silent and filled pauses in first (L1) and second language (L2) speech. The occurrence of pauses of 52 L2 and 18 L1 Dutch speakers was evaluated with respect to utterance boundaries and word frequency. We found that L2 speakers paused more often than L1 speakers within utterances; but not between utterances. Similarly, only within utterances, L2 pauses were longer than L1 pauses. Regarding word frequency, both L1 and L2 speakers are more likely to pause before lower frequency words as compared to higher frequency words. These findings imply that L1 and L2 speakers' production processes may be similar in that (1) pauses at utterance boundaries are used for conceptual planning mostly and (2) lexical retrieval difficulties are comparable for L1 and L2 speakers. These findings furthermore imply that when using fluency for L2 testing, pause locations must be taken into account.*

Keywords: pause distribution; fluency; speech production; second language speaking

## 1. Introduction

Being fluent in an L2 is defined as being able to smoothly and effortlessly translate intended messages to speech (e.g., Schmidt 1992) and, hence, fluent speech is defined as speech without (unnatural) hesitations. Fluency, together with complexity (with respect to syntax and lexis) and accuracy in L2 speech (with respect to morpho-syntax, lexis, and pronunciation), is one of the three perceptual dimensions of speech that develop as L2 learners' proficiency progresses (Housen and Kuiken 2009). It is therefore no wonder that fluency in speech is used as a diagnostic in second language (L2) assessment. Indeed, human judgements on overall L2 proficiency are related to aspects of fluency (e.g., Iwashita et al. 2008). Additionally, for most speakers, their L2 speech is less fluent than their L1 speech and L2 speakers progress in certain aspects of fluency over time (Derwing et al. 2009; Towell et al. 1996). Aspects of fluency are indicative of proficiency and can be discriminatory between L1 and L2 speech. So could one simply count the number of disfluencies such as silent and filled pauses to distinguish between L1 and L2 speech or to measure a fluency aspect of L2 proficiency? The answer is no, because between L1 speakers there are also differences in levels of fluency (Bortfeld et al. 2001) and between L2 speakers matched on overall L2 proficiency, large differences with respect to measures of fluency likewise exist.

Davies (2003) suggested that it might not be the amount of disfluencies, but rather their distribution, that will be indicative of L2 proficiency. Davies (2003) refers to Pawley and Syder (1983), who suggest that native speakers produce speech “one clause at a time”. Because native speakers can produce multi-word chunks, they direct attention to planning the upcoming message, without having to pay attention to the linguistic formulation of the current message. This will result in pauses that occur at clause boundaries mostly. L2 speakers, however, may not (yet) have a large repository of pre-fabricated chunks and therefore pauses may also occur clause-medially.

To investigate how L2 speakers’ pause distribution is different from L1 speakers’ pause distribution, one first needs to consider the distribution of pauses for L1 speakers. The current study will investigate two aspects that have been shown to influence pause occurrence in L1 speech, and extend it to L2 speech: (1) pause occurrence between and within utterances, and (2) pause occurrence before low and high frequency words. In L1 speech, pauses have been found to be more likely to occur before (major) constituents, at syntactic boundaries (Swerts 1998), presumably reflecting pauses in which speakers plan what to say. Within clauses, pauses are more likely to occur before open-class words (Maclay and Osgood 1959), that are low in predictability (Goldman-Eisler 1958) and (therefore) pauses are more likely to occur before low-frequency words (Hartsuiker and Notebaert 2010; Kircher et al. 2004).

Previous research (e.g., Riazantseva 2001; Skehan and Foster 2007) has already investigated the potential difference between L1 and L2 pause distribution between versus within ASU’s (Analysis of Speech Units, Foster et al. 2000) and (major) constituents, and investigated how L2 proficiency moderates pause distributions. As such, this study will partly be a replication study. However, as argued below, the present study adopts better-suited analyses to investigate these questions. The factor “word frequency” has (to my knowledge) not been investigated in L2 speech before.

Additionally, the current study will investigate potential differences in pause durations between L1 and L2 speech, taking pause placement into account. Riazantseva (2001) also investigated differences in pause duration between L1 and L2 speech and between speakers with different levels of proficiency. She found that higher proficient speakers paused, on average, longer than lower proficient speakers. Other studies, however, have suggested that pause duration is not related to proficiency (De Jong et al. 2013; De Jong et al. 2015; Towell et al. 1996). However, with the exception of De Jong et al. (2015), these studies did not take pause placement into account. The current study will therefore test whether pause duration is dependent on pause placement (between versus within ASU’s) for L1 and L2 speakers.

### *1.1. Previous studies investigating the distribution of pauses*

From previous research one can conclude that indeed, L2 speakers pause more often within ASU’s (Skehan and Foster 2007), clauses (Tavakoli 2011) or constituents (Riazantseva 2001) than L1 speakers do. However, the measures and analyses that were used in these studies make firm conclusions difficult. Riazantseva (2001) used percentage of pauses within constituents and compared this measure for the same

speakers in their L1 (Russian) and L2 (English). It is likely, however, that the speakers produced more complex and longer constituents in their L1 compared to their L2. If this is indeed the case, there were more opportunities to pause in the longer constituents (L1) than in the shorter constituents (L2). The found difference in percentage of pauses within constituents between L1 and L2 might therefore have been underestimated. Rianzantseva (2001) did not find a difference between the two L2 proficiency levels in her sample (higher and intermediate). Because it may very well have been the case that the higher proficient L2 speakers produced longer constituents, comparing the percentage of pauses within constituents between these groups may not have been valid.

The studies by Tavakoli (2011) and Skehan and Foster (2007) have the same issue with the measures used to compare the distribution of pauses between L1 and L2 speakers. Skehan and Foster (2007) report a *ratio* of number of pauses within ASU's to number of pauses between ASU's. Again, when longer ASU's are produced, there are more opportunities to pause within them than when learners produce only short ASU's. In addition, they report the mean number of pauses either within or between ASU's for the speaking performances. From these measures, it is found that L1 speakers pause *more* at the boundaries than L2 speakers do. However, if L1 speakers in total produce more speech and more ASU's, comparing the number of pauses for L1 and L2 speakers at this position is not a valid comparison.

Finally, Tavakoli (2011) counted the mean number of pauses either within or between clauses produced by L1 speakers and compared these to the mean number of pauses for a group of L2 speakers. She also reported that L2 speakers pause more often within clauses than L1 speakers do. However, as explained before, when clauses are longer, which is likely the case for the L1 speakers, there is more opportunity to pause within them. In the current study, we will therefore employ logistic regression analyses, where each word boundary is considered as potential pause position. With such analyses, length of ASU's or number of ASU's in the sample are no longer confounding factors.

## 1.2. Research questions

The previous discussion leads to the following four research questions:

RQ1: Do L1 and L2 speakers differ in pause distribution with respect to within/between ASU's?

RQ2: Do L1 and L2 speakers differ in pause distribution with respect to word frequency?

RQ3: What is the relation between pause placement (between/within ASU's) and pause duration for L1 and L2 speakers?

RQ4: Does L2 proficiency moderate the pause occurrence and pause duration patterns of L2 speakers?

## 2. Method

The current study used the same L2 data as reported on in De Jong et al. (2015), but now adding transcripts and measures from L1 speakers performing the same speaking tasks.

### 2.1. *Participants*

Twenty-nine native speakers of English, 25 native speakers of Turkish, and 18 native speakers of Dutch were paid to take part in our experiment, with signed informed consent. This research was part of a larger project with, for some participants, more tasks than are reported on here (see De Jong et al. 2012, 2013, and 2015). Depending on how many tasks the participants completed, they were paid between 30 and 50 euros. The Turkish (8 male, 17 female; mean age = 32, range = 23–48) and English participants in the current study (11 male, 18 female; mean age = 31, range = 23–43) had come to the Netherlands between the ages of 18 and 40 (English range = 22–40, Turkish range = 18–35). Most of these participants had lived in the Netherlands for fewer than 10 years (English mean = 4.5 years, range = 1 month to 21 years; Turkish mean = 7 years, range = 9 months to 20 years). All L2 participants were at an intermediate to advanced level of Dutch as an L2 and the majority was taking intermediate or advanced level Dutch courses to prepare for enrollment at the University of Amsterdam. Most of the 18 native speakers (7 male, 11 female); mean age = 26, range = 19 – 45) were students at the same university.

### 2.2. *Materials: speaking tasks*

Speech was elicited by using eight speaking tasks as described in De Jong et al. (2012). The speaking tasks differed in difficulty, formality, and discourse mode. Instructions for each task contained specific information about the speaking task itself, which was provided by one or several visual-verbal cues on a computer screen. No additional knowledge about the topic beyond the information provided in the tasks was needed to successfully complete each speaking task. Information about the purpose and audience of the task was also provided.

### 2.3. *Materials: Vocabulary test*

As a proxy for overall L2 proficiency, to be assessed separately from the speaking performances, we chose to use a productive vocabulary task. Vocabulary knowledge has been shown to be a good predictor of overall L2 proficiency (Beglar and Hunt 1999; Zareva et al. 2005). Moreover, the vocabulary test currently used has been shown to be a strong predictor of overall speaking proficiency. In De Jong et al. (2012), using structural equation modeling, the vocabulary knowledge score showed a strong relation to ratings of overall speaking proficiency ( $r = 0.79$ ).

The paper-and-pencil task (with instructions in the L2, Dutch) elicited knowledge of 90 words, and of 26 multi-word units. The total score for each participant was calculated as the total number of correct responses. The scoring procedure was lenient towards spelling mistakes and errors in inflectional variants. This vocabulary test was piloted and improved before first used in De Jong et al. (2012). In their original sample of 181 L2 learners of Dutch, Cronbach's alpha for this test was high ( $\alpha = 0.98$ ), indicating good reliability.

#### 2.4. Procedure

Participants completed the eight tasks in an office with a native Dutch-speaking experimenter present. The tasks were presented on a computer screen and the participants' speech was recorded. Participants navigated the experiment and instructions themselves. Each task consisted of screens containing information about the task in Dutch, including pictures. For each task, participants had 30 seconds of preparation time and 120 seconds of speaking time, which was shown by a status bar at the bottom of the screen. Participants could press a "finished" button if they finished the task before the 120 seconds had gone by.

The average time used for completing all eight tasks was around 25 minutes. The vocabulary task was done in a separate session and took between twenty minutes and an hour.

#### 2.5. Data analyses

All speech recordings (72 participants, eight tasks; totaling roughly 15 hours of speaking data) were transcribed and annotated by three native speakers of Dutch. They followed precise guidelines to annotate as similarly as possible. For two (English) participants, more than half of the performances were not recorded well (interference with the computer caused a strong hum). Therefore, these two participants were discarded from further analyses. In addition, eight recordings (from 6 participants) were not recorded with sufficient quality to make precise transcriptions. These recordings were also discarded. From 70 participants we thus obtained speech performances from at least six, but for most participants eight tasks. The transcriptions were made in CLAN (MacWhinney 2000). Besides orthographic transcriptions, information relevant for measuring pauses was inserted. Silent pauses were detected by careful listening and by using the waveform (as shown in CLAN). The transcribers added silent pause boundaries in the CLAN-waveform manually. The silent pauses were subsequently measured in milliseconds automatically (using the boundaries). The lower silent pause threshold was set at 250ms, to exclude short so-called micropauses (Riggenbach 1991), which are irrelevant for measures of L2 fluency (De Jong and Bosker 2013).

The transcriptions were also split up into so-called analysis of speech units (ASU). Foster et al. (2000) have shown that using the ASU is the optimal way of dividing transcribed data into analyzable units. As defined by Foster et al. (2000: 365), an ASU is "a single speaker's utterance consisting of an independent clause, or a subclausal unit,

together with any subordinate clause(s) associated with either.” In the current analyses, we will use such ASU’s (the actual speech units) to mark major boundaries in speech. Silent pauses were categorized as being either between or within ASU. Furthermore, the transcripts were annotated with nonlexical filled pauses (such as “uh,” “uhm,” “er,” “mm”). Finally, for nouns, CELEX word frequency was added (Baayen et al. 1995).

### 3. Results

In total, from 70 participants, 84599 words were transcribed, 1209 on average per participant (SD 369) with a range from 380 to 2013. Below, we report on a number of analyses. We ran generalized linear mixed effect models as implemented in the lme4 library (Bates et al. 2010) in R to predict either silent or filled pause occurrence separately. With such analyses, we were able to consider each word transition as a data point while accounting for variance between speakers and variance between words. Taking this variance into account is important, as for both speakers and words we have repeated measures. Each word transition can be considered a data point because between two words either a pause occurred, or did not occur. For both silent and filled pauses, we first ran models to test for differences with respect to ASU-position (between or within ASU). Taking only the transitions within ASU’s into account, for all transitions to nouns, a potential effect of Word Frequency was tested. To test interactions with Proficiency (measured with Vocabulary knowledge) and MotherTongue (English/Turkish), analyses were run on the subset of L2 speakers.

Finally, we ran linear mixed models to predict silent pause duration. Again, we ran models investigating the relation between pause placement and pause duration with all participants to test interactions with L1/L2 and separate models on the subset of L2 speakers, to test interactions with Proficiency and MotherTongue. No analyses to test for the effect of word frequency on pause duration for the subset of nouns was viable, as this restriction led to a too small dataset.

#### 3.1. *Predicting pause occurrence: between or within ASU’s*

Table 1 shows the counts of silent and filled pauses either between or within ASU’s for both L1 and L2 speech. The percentages are based on the total number of ASU’s for L1 and L2 speakers (1797 and 5115, respectively) and on the total number of words (minus all the initial words of ASU’s) produced for L1 and L2 speakers (20660 and 50115, respectively). From these numbers, we can already see that L1 speakers, on average, had longer ASU’s than L2 speakers (13.5 and 11.8 words per ASU, respectively;  $\chi^2(1) = 26.12, p < 0.001$ ).

*Table 1. Number of silent and filled pauses (percentages\*) for L1 and L2 speakers.*

|             | Nr of silent pauses<br>(% of total) |                 | Nr of filled pauses<br>(% of total) |                |
|-------------|-------------------------------------|-----------------|-------------------------------------|----------------|
|             | L1                                  | L2              | L1                                  | L2             |
| Between ASU | 1320<br>(73.5%)                     | 3805<br>(74.4%) | 175<br>(9.7%)                       | 533<br>(10.4%) |
| Within ASU  | 1599<br>(7.7%)                      | 7572<br>(15.1%) | 684<br>(3.3%)                       | 4257<br>(8.5%) |

\* Percentages based on total number of ASU's for L1 and L2 (1797 and 5115) and total within ASU-words for L1 and L2 (20660 and 50115).

To test for differences between L1 and L2 speech with respect to pause occurrence, we first ran generalized linear mixed models predicting silent pause occurrence for each word with participant ( $N = 70$ ) and word ( $N = 4110$ ) as crossed random effects, using the Laplace approximation. Both random effects were highly significant, as shown by likelihood ratio tests (Pinheiro and Bates 2000). We then proceeded with adding fixed effects; again using likelihood ratio tests to test whether adding these fixed effects significantly improved the model. We added Task (tasks 1 – 8), Position (within or between ASU), and NativeSpeaker (L1 or L2) as fixed effects. Task as effect was not to answer a research question, but was added to get more precise estimations of the remaining effects (the effects of the tasks are therefore mentioned in the tables presenting the models, but will not be commented on). Finally, to test whether a possible effect for Position was similar across L1 and L2 speakers, we added the interaction between Position and NativeSpeaker. The first three columns of Table 2 show the resulting model. This model took L2 speaker in task 1 with position between ASU's as its intercept, and all effects should be interpreted relative to this intercept. Between ASU, there was no effect for NativeSpeaker. There was an effect of Position: for each word transition, there were lower probabilities of pauses to occur when the word would be within ASU's as compared to between ASU's. Crucially, there was also a significant interaction between NativeSpeaker and Position showing that the effect of Position (fewer pauses within than between ASU's) is stronger for L1 speech than for L2 speech.

For filled pauses, we ran the same model, but now predicting filled pause occurrence. The results of this model are shown in the last three columns of Table 2. The results were very similar to the model for silent pauses: no significant effect for NativeSpeaker between ASU's, a significant effect of Position (lower chance of pauses within ASU's), and a significant interaction: the effect of Position was again stronger for L1 speech as compared to L2 speech.

*Table 2. Results of generalized linear mixed models predicting silent pause occurrence and filled pause occurrence for L1 and L2 speakers' word transitions.*

|                              | Predicting silent pause occurrence |          |          | Predicting filled pause occurrence |          |          |
|------------------------------|------------------------------------|----------|----------|------------------------------------|----------|----------|
|                              | Estimates (SE)                     | z-values | p-values | Estimates (SE)                     | z-values | p-values |
| <i>Fixed effects</i>         |                                    |          |          |                                    |          |          |
| (Intercept)                  | 0.742<br>(0.086)                   | 8.63     | <0.001   | -1.960<br>(0.131)                  | -14.95   | <0.001   |
| Nativespeaker                | -0.098<br>(0.144)                  | -0.68    | 0.495    | -0.011<br>(0.234)                  | -0.05    | 0.961    |
| ASUwithin                    | -2.659<br>(0.042)                  | -63.27   | <0.001   | -0.674<br>(0.059)                  | -11.52   | <0.001   |
| Nativespeaker with ASUwithin | -0.719<br>(0.074)                  | -9.75    | <0.001   | -0.856<br>(0.106)                  | -8.09    | <0.001   |
| Task2                        | 0.153<br>(0.053)                   | 2.89     | 0.004    | 0.028<br>(0.069)                   | 0.41     | 0.683    |
| Task3                        | -0.094<br>(0.047)                  | -1.99    | 0.047    | -0.312<br>(0.065)                  | -4.83    | <0.001   |
| Task4                        | 0.027<br>(0.049)                   | 0.55     | 0.581    | -0.042<br>(0.064)                  | -0.65    | 0.516    |
| Task5                        | 0.187<br>(0.048)                   | 3.88     | <0.001   | -0.044<br>(0.064)                  | -0.68    | 0.498    |
| Task6                        | 0.165<br>(0.046)                   | 3.56     | <0.001   | 0.248<br>(0.059)                   | 4.22     | <0.001   |
| Task7                        | -0.136<br>(0.047)                  | -2.91    | <0.001   | -0.153<br>(0.062)                  | -2.49    | 0.013    |
| Task8                        | -0.015<br>(0.045)                  | -0.33    | 0.745    | -0.039<br>(0.060)                  | -0.65    | 0.516    |
| <i>Random effects</i>        |                                    |          |          |                                    |          |          |
| Words                        | 0.511                              |          |          | 0.410                              |          |          |
| Speakers                     | 0.217                              |          |          | 0.603                              |          |          |

We then tested, for the subset L2 speakers, whether there was an effect of Proficiency (estimated by the vocabulary size measure) and MotherTongue (Turkish or English). This model predicted silent pause occurrence for each word with L2 participant (N = 52) and word (N = 3350) as crossed random effects, using the Laplace approximation. The first three columns of Table 3 show the result of this model, which took English speakers' pausing behavior between ASU's in task 1 as intercept. Relative to this intercept, there was no overall effect of MotherTongue. There was an effect of Proficiency (higher

Proficiency resulted in a lower probability of pausing between ASU's) and an effect of Position (within ASU's there was a lower probability of a silent pause than between ASU's). The interaction between Proficiency and Position was also significant: for participants with a higher Proficiency, there was a stronger effect of Position (higher probability of pausing within ASU's as compared to between ASU's). There was no interaction between MotherTongue and Proficiency, indicating that Turkish and English native speakers do not show differential pause probabilities in their L2 (Dutch).

For filled pauses, we tested the same model, but now had filled pause occurrence as the dependent variable. This model is shown in the last three columns of Table 3. Similar to the model for silent pauses, there was no significant effect for MotherTongue. The significant interaction between MotherTongue and Position showed, however, that within ASU's, the Turkish speakers used more filled pauses than the English speakers. There was again an effect of Proficiency: higher proficiency was associated with fewer filled pauses. Because there was no interaction between Position and Proficiency, we can conclude that this effect of Proficiency was similar between and within ASU's.

*Table 3. Results of generalized linear mixed models predicting silent pause occurrence and filled pause occurrence for L2 speakers' word transitions.*

|                            | Predicting silent pause occurrence |          |          | Predicting filled pause occurrence |          |          |
|----------------------------|------------------------------------|----------|----------|------------------------------------|----------|----------|
|                            | Estimates (SE)                     | z-values | p-values | Estimates (SE)                     | z-values | p-values |
| <i>Fixed effects</i>       |                                    |          |          |                                    |          |          |
| (Intercept)                | 0.645<br>(0.102)                   | 6.34     | <0.001   | -2.174<br>(0.178)                  | -12.23   | <0.001   |
| Proficiency                | -0.139<br>(0.063)                  | -2.21    | 0.027    | -0.239<br>(0.115)                  | -2.08    | 0.037    |
| L1Turkish                  | 0.182<br>(0.128)                   | 1.42     | 0.157    | 0.280<br>(0.238)                   | 1.18     | 0.238    |
| ASUwithin                  | -2.632<br>(0.055)                  | -47.91   | <0.001   | -0.813<br>(0.081)                  | -10.04   | <0.001   |
| Proficiency with ASUwithin | -0.164<br>(0.037)                  | -4.44    | <0.001   | 0.026<br>(0.048)                   | 0.54     | 0.590    |
| L1Turkish with ASUwithin   | -0.126<br>(0.074)                  | -1.71    | 0.087    | 0.303<br>(0.103)                   | 2.95     | 0.003    |
| Task2                      | 0.157<br>(0.059)                   | 2.66     | 0.008    | 0.007<br>(0.076)                   | 0.10     | 0.923    |
| Task3                      | -0.043<br>(0.053)                  | -0.80    | 0.422    | -0.292<br>(0.070)                  | -4.15    | <0.001   |
| Task4                      | 0.070<br>(0.054)                   | 1.28     | 0.199    | -0.036<br>(0.070)                  | -0.52    | 0.606    |
| Task5                      | 0.243                              | 4.53     | <0.001   | -0.017                             | -0.25    | 0.804    |

|                       |         |       |        |  |         |       |        |
|-----------------------|---------|-------|--------|--|---------|-------|--------|
|                       | (0.054) |       |        |  | (0.070) |       |        |
| Task6                 | 0.201   | 3.82  | <0.001 |  | 0.227   | 3.47  | <0.001 |
|                       | (0.053) |       |        |  | (0.066) |       |        |
| Task7                 | -0.131  | -2.48 | 0.013  |  | -0.079  | -1.18 | 0.240  |
|                       | (0.053) |       |        |  | (0.067) |       |        |
| Task8                 | -0.050  | -0.98 | 0.329  |  | 0.003   | 0.04  | 0.968  |
|                       | (0.052) |       |        |  | (0.065) |       |        |
| <i>Random effects</i> |         |       |        |  |         |       |        |
| Words                 | 0.518   |       |        |  | 0.433   |       |        |
| Speakers              | 0.153   |       |        |  | 0.602   |       |        |

### 3.2. Predicting pauses: word frequency

For this analysis, we decided to focus on content words and restrict the analysis to nouns only. We excluded adjectives and verbs as it is unclear which frequency measure would actually be relevant due to the different inflectional variants of the adjectival and verb forms. In our corpus of speech, 14431 words were classified as nouns (within ASU's), 10644 tokens by L2 speakers, and 3787 tokens by L1 speakers. The range, median, and mean CELEX noun frequency was slightly lower for L1 compared to L2 speakers when calculated over all tokens (L1: range 0 – 55630, median: 2518, mean: 8098; L2: range 0 – 55630, median: 3112, mean: 8536). Comparing the mean frequencies for L1 and L2 in a linear mixed model with NativeSpeaker as a predictor variable (and participant as random effect) indeed showed a significant difference between the frequencies of the nouns used by the two groups ( $p = 0.038$ ).

Similarly as for the analysis on all word transitions, for the 14431 transitions to nouns both participant ( $N = 70$ ) and word (i.e., nominal word,  $N = 1221$ ) were highly significant random factors, both in the model predicting silent pauses (first three columns of Table 4) as for the model predicting filled pauses (last three columns of Table 4). Likewise for both models, the effect of NativeSpeaker was significant: before nouns L1 speakers are less likely to pause than L2 speakers. Also the factor Word Frequency proved to significantly improve both model; transitions to high-frequent nouns are less likely to contain a pause than transitions to low-frequent nouns. The interactions between NativeSpeaker and Word Frequency, however, were not significant.

*Table 4. Results of generalized linear mixed models predicting silent pause occurrence and filled pause occurrence for L1 and L2 speakers' transitions to nouns.*

|  | Predicting silent pause occurrence |          |          | Predicting filled pause occurrence |          |          |
|--|------------------------------------|----------|----------|------------------------------------|----------|----------|
|  | Estimates (SE)                     | z-values | p-values | Estimates (SE)                     | z-values | p-values |
|  |                                    |          |          |                                    |          |          |

| <i>Fixed effects</i>                 |                   |        |        |                   |        |        |
|--------------------------------------|-------------------|--------|--------|-------------------|--------|--------|
| (Intercept)                          | -2.271<br>(0.146) | -15.51 | <0.001 | -2.845<br>(0.175) | -16.22 | <0.001 |
| Nativespeaker                        | -1.064<br>(0.187) | -5.69  | <0.001 | -0.856<br>(0.248) | -3.45  | 0.001  |
| Noun frequency                       | -0.289<br>(0.053) | -5.41  | <0.001 | -0.361<br>(0.054) | -6.65  | <0.001 |
| Nativespeaker with<br>Noun frequency | -0.157<br>(0.083) | -1.88  | 0.060  | 0.157<br>(0.095)  | 1.64   | 0.101  |
| Task2                                | -0.400<br>(0.194) | -2.07  | 0.039  | -0.673<br>(0.208) | -3.24  | 0.001  |
| Task3                                | -0.566<br>(0.183) | -3.10  | 0.002  | -0.442<br>(0.198) | -2.24  | 0.025  |
| Task4                                | -0.243<br>(0.174) | -1.40  | 0.162  | -0.105<br>(0.187) | -0.56  | 0.572  |
| Task5                                | -0.088<br>(0.195) | -0.45  | 0.652  | 0.044<br>(0.209)  | 0.21   | 0.832  |
| Task6                                | 0.280<br>(0.153)  | 1.83   | 0.067  | 0.302<br>(0.164)  | 1.84   | 0.067  |
| Task7                                | -0.219<br>(0.159) | -1.37  | 0.170  | -0.415<br>(0.176) | -2.36  | 0.018  |
| Task8                                | -0.148<br>(0.160) | -0.93  | 0.354  | -0.053<br>(0.177) | -0.30  | 0.763  |
| <i>Random effects</i>                |                   |        |        |                   |        |        |
| Words                                | 0.605             |        |        | 0.525             |        |        |
| Speakers                             | 0.326             |        |        | 0.631             |        |        |

To see whether Mothertongue and Proficiency were significant predictors for pauses occurring before nouns, we also ran generalized linear mixed models on the subset of L2 speaker data of all nouns (N = 10644). Again, Participant (N = 52) and Word (nominal word; N = 988) were the crossed random effects. For silent pauses (see columns 1 – 3 of Table 5), the effects of Mothertongue, Proficiency, and Noun Frequency were significant: Turkish speakers produced more silent pauses before nouns than English speakers did; higher proficient speakers produced fewer silent pauses than lower proficient speakers; and there was a higher probability of silent pauses before lower-frequency words. There were no significant interactions. In the model predicting filled pauses, the results showed that the effect of Proficiency was not significant. Turkish speakers were more likely to use a filled pause than English speakers. Like in the model for silent pauses, there were no significant interactions.

*Table 5. Results of generalized linear mixed models predicting silent pause occurrence and filled pause occurrence for L2 speakers' transitions to nouns.*

|                                 | Predicting silent pause occurrence |          |          | Predicting filled pause occurrence |          |          |
|---------------------------------|------------------------------------|----------|----------|------------------------------------|----------|----------|
|                                 | Estimates (SE)                     | z-values | p-values | Estimates (SE)                     | z-values | p-values |
| <i>Fixed effects</i>            |                                    |          |          |                                    |          |          |
| (Intercept)                     | -2.473<br>(0.163)                  | -15.18   | <0.001   | -3.225<br>(0.218)                  | -14.76   | <0.001   |
| Proficiency                     | -0.246<br>(0.071)                  | -3.45    | 0.001    | -0.193<br>(0.113)                  | -1.70    | 0.088    |
| L1Turkish                       | 0.386<br>(0.145)                   | 2.64     | 0.008    | 0.591<br>(0.236)                   | 2.50     | 0.012    |
| Noun frequency                  | -0.262<br>(0.064)                  | -4.09    | <0.001   | -0.348<br>(0.067)                  | -5.17    | <0.001   |
| Proficiency with Noun frequency | 0.033<br>(0.033)                   | 0.99     | 0.320    | -0.032<br>(0.035)                  | -0.89    | 0.373    |
| L1 Turkish with Noun frequency  | -0.058<br>(0.068)                  | -1.00    | 0.315    | -0.005<br>(0.077)                  | -0.06    | 0.952    |
| Task2                           | -0.339<br>(0.206)                  | -1.65    | 0.099    | -0.630<br>(0.222)                  | -2.84    | 0.005    |
| Task3                           | -0.562<br>(0.194)                  | -2.90    | 0.004    | -0.406<br>(0.211)                  | -1.92    | 0.055    |
| Task4                           | -0.263<br>(0.187)                  | -1.41    | 0.160    | -0.149<br>(0.203)                  | -0.73    | 0.463    |
| Task5                           | -0.054<br>(0.208)                  | -0.26    | 0.796    | 0.019<br>(0.223)                   | 0.09     | 0.930    |
| Task6                           | 0.265<br>(0.167)                   | 1.59     | 0.112    | 0.352<br>(0.179)                   | 1.96     | 0.049    |
| Task7                           | -0.211<br>(0.172)                  | -1.23    | 0.220    | -0.307<br>(0.188)                  | -1.64    | 0.101    |
| Task8                           | -0.219<br>(0.174)                  | -1.26    | 0.207    | -0.004<br>(0.191)                  | -0.02    | 0.982    |
| <i>Random effects</i>           |                                    |          |          |                                    |          |          |
| Words                           | 0.607                              |          |          | 0.490                              |          |          |
| Speakers                        | 0.203                              |          |          | 0.590                              |          |          |

### 3.3. Predicting silent pause duration: between or within ASU's

To further investigate differences in fluency between L1 and L2 speakers, we also analysed silent pause duration, taking into account pause location. Table 6 shows the

means and standard deviations of all silent pause durations (of all silent pauses >250ms and <3000ms) between and within ASU's for L1 and L2 speech separately.

*Table 6. Means and standard deviations of silent pauses for L1 and L2 speech in milliseconds.*

|                        | Mean duration of silent pauses (SD) |           |
|------------------------|-------------------------------------|-----------|
|                        | L1                                  | L2        |
| Between ASU (N = 4995) | 729 (407)                           | 794 (464) |
| Within ASU (N = 9075)  | 565 (329)                           | 687 (439) |

In the analyses predicting silent pause duration, we used linear mixed models. We had Participant and Word as crossed random effects, as in the analyses predicting pause occurrence. We first investigated whether (log) pause duration in our dataset was dependent on Position (within or between ASU) and NativeSpeaker. There were 14070 silent pauses in the dataset. Both random effects Participant (N = 70) and Word (N = 1484) were significant. In addition, the fixed effect Task proved to be a significant factor. Table 7 shows the results of this model. The number of degrees of freedom required for statistical significance testing of the  $t$  values was calculated as  $J - m - 1$  (Hox 2010), where  $J$  is the most conservative number of a random effect (70 speakers) and  $m$  is the total number of explanatory variables in the model ( $m = 13$ ) resulting in 56 degrees of freedom. The model took L2 speaker with position between ASU's as its intercept, and the effects should be interpreted relative to this intercept. Between ASU's, there was no effect for NativeSpeaker. There was an effect of Position: within ASU's pauses were shorter as compared to between ASU's. Crucially, there was also a significant interaction between NativeSpeaker and Position: the effect of Position (shorter pauses within than between ASU's) proved to be stronger for L1 speech than for L2 speech.

*Table 7. Results of linear mixed model predicting (log) silent pause duration for L1 and L2 speakers.*

|                                     | Predicting (log) silent pause duration |          |          |
|-------------------------------------|--|----------|----------|
|                                     | Estimates (SE)                         | t-values | p-values |
| <i>Fixed effects</i><br>(Intercept) | 6.627                                  | 205.91   | <0.001   |

|                                 |         |        |        |
|---------------------------------|---------|--------|--------|
|                                 | (0.032) |        |        |
| Nativespeaker                   | -0.084  | -1.50  | 0.139  |
|                                 | (0.056) |        |        |
| ASUwithin                       | -0.181  | -14.51 | <0.001 |
|                                 | (0.012) |        |        |
| Nativespeaker<br>with ASUwithin | -0.071  | -3.04  | 0.004  |
|                                 | (0.023) |        |        |
| Task2                           | -0.021  | -1.06  | 0.294  |
|                                 | (0.020) |        |        |
| Task3                           | -0.036  | -1.92  | 0.060  |
|                                 | (0.019) |        |        |
| Task4                           | -0.084  | -4.28  | <0.001 |
|                                 | (0.020) |        |        |
| Task5                           | -0.041  | -2.17  | 0.034  |
|                                 | (0.019) |        |        |
| Task6                           | -0.055  | -3.03  | 0.004  |
|                                 | (0.018) |        |        |
| Task7                           | -0.093  | -4.91  | <0.001 |
|                                 | (0.019) |        |        |
| Task8                           | -0.090  | -5.04  | <0.001 |
|                                 | (0.018) |        |        |
| <i>Random effects</i>           |         |        |        |
| Words                           | 0.0058  |        |        |
| Speakers                        | 0.0275  |        |        |
| Residual                        | 0.2315  |        |        |

We then tested, for the subset of L2 speakers, whether there was an effect of Proficiency and Mothertongue. This model predicted silent pause duration for each silent pause in this subset with participant (N = 52) and Word (N = 1332) as crossed random effects and with df = 36 (Hox 2010). Table 8 shows the result of this model, which had English speakers between ASU's as intercept. Relative to this intercept, there was no effect of Mothertongue or Proficiency, which means that between ASU's, Turkish and English speakers had comparable pause durations, no matter what their Proficiency was. The main effect for Position shows that the pauses within ASU's were significantly shorter than pauses between ASU's. Additionally, there were two significant interactions. First of all, the effect of Position was modulated by Proficiency: the higher the Proficiency of the L2 speakers, the shorter the pause durations within ASU's tended to be. Secondly, it turned out that within ASU's, the Turkish speakers tended to have shorter pause durations than the English speakers.

*Table 8. Results of linear mixed model predicting (log) silent pause duration for L2 speakers.*

| Predicting (log) silent pause duration |                   |          |          |
|--|-------------------|----------|----------|
|  | Estimates<br>(SE) | t-values | p-values |
| <i>Fixed effects</i>                   |                   |          |          |
| (Intercept)                            | 6.523<br>(0.038)  | 172.53   | <0.001   |
| Proficiency                            | 0.040<br>(0.025)  | 1.65     | 0.108    |
| L1Turkish                              | 0.060<br>(0.050)  | 1.22     | 0.230    |
| ASUwithin                              | -0.121<br>(0.015) | -7.85    | <0.001   |
| Proficiency with<br>ASUwithin          | -0.052<br>(0.010) | -5.05    | <0.001   |
| L1Turkish with<br>ASUwithin            | -0.043<br>(0.020) | -2.09    | 0.044    |
| Task2                                  | -0.014<br>(0.021) | -0.65    | 0.520    |
| Task3                                  | -0.037<br>(0.020) | -1.89    | 0.067    |
| Task4                                  | -0.050<br>(0.020) | -2.45    | 0.019    |
| Task5                                  | -0.026<br>(0.020) | -1.32    | 0.195    |
| Task6                                  | -0.042<br>(0.019) | -2.20    | 0.034    |
| Task7                                  | -0.067<br>(0.020) | -3.42    | 0.002    |
| Task8                                  | -0.074<br>(0.019) | -3.95    | <0.001   |
| <i>Random effects</i>                  |                   |          |          |
| Words                                  | 0.006             |          |          |
| Speakers                               | 0.028             |          |          |
| Residual                               | 0.243             |          |          |

#### 4. Discussion and conclusion

This paper examined how L1 and L2 speech differs in pause placement regarding utterance boundaries and word frequency. Additionally, the paper investigated whether

L2 proficiency is a mediating factor in predicting pause distribution. Finally, silent pause duration between and within utterances was investigated.

Previous studies had already shown that L2 speakers tend to pause more within constituents, clauses, and ASU's as compared to L1 speakers (Riazantseva 2001; Skehan and Foster 2007; Tavakoli 2011). The measures and analyses used in these studies, however, potentially underestimated such differences. The current study therefore used logistic analyses, taking all word transitions as potential pause locations to predict, for each potential pause location, whether a pause actually occurred or not.

With respect to ASU's, this study found that L2 speakers and L1 speakers do not differ at ASU-boundaries: the likelihood of both silent and filled pauses is not significantly different. However, within ASU's, L2 speakers are more likely to pause – either with a filled or with a silent pause – than L1 speakers are. Whereas Riazantseva (2001) did not find any significant differences with respect to pause distribution for low and high proficient speakers, the results of the current study did show a mediating effect of (gradient) L2 proficiency: the interaction (for silent pauses only) between proficiency and location with respect to ASU showed that as learners were more proficient, they produced fewer silent pauses within ASU's.

The study also investigated whether pauses are more likely to occur before low-frequency words in (semi-)spontaneous speech. Previous research has shown that at least in constrained description tasks (in which participants had to describe a path from picture to picture on a computer screen), L1 speakers are more likely to pause before lower frequency nouns in comparison to higher frequency nouns (Hartsuiker and Notebaert 2010; Kircher et al. 2004). The current study showed that, in more spontaneous speech, both L1 and L2 speakers are more likely to pause before lower frequency nouns than before higher frequency nouns and that L2 proficiency did not modulate this effect. It is likely, however, that the range of which words are considered to be “lower-frequent” and therefore more likely to induce a pause, differs for the L1 and L2 speakers. This option is in fact borne out by the data: it was found that the nouns used by L2 speakers were, overall, higher in frequency than those by L1 speakers. The relation between frequency of nouns and likelihood to pause may be the same (or at least, in this corpus we were unable to prove that the relation was different), but the relation seems to hold for a slightly different range of frequencies: somewhat higher frequency nouns for L2 speakers and the lower frequency nouns for L1 speakers

The current findings have implications for our understanding of speech production processes in L1 and L2. Because both L1 and L2 speakers pause before ASU's with filled or silent pauses to the same extent, we may infer that at these positions, speakers make their conceptual plan and this causes the speakers to pause, no matter what language they speak. This conclusion is validated by the finding that between ASU's, L1 and L2 speakers tend to pause for similar durations. Before starting an ASU, a speaker plans his conceptual message and only the beginning of the linguistic message, both in L1 and in L2. It is within ASU's that pausing behaviour differs for L1 and L2 speech. Within ASU's, L2 speakers more often run into trouble while formulating the linguistic message than L1 speakers do and are therefore more likely to pause, for longer durations than L1

speakers do. The pauses within ASU's that L2 speakers use (in excess of those that L1 speakers would use), we may hypothesize, are due to less L2 knowledge and lower L2 skills. Again, this is corroborated by the moderating effect we found for the proficiency of the L2 speakers: the higher the L2 proficiency, the less likely speakers were to pause within ASU's and the shorter the pauses within ASU's tended to be.

This study also found an effect of word frequency on the likelihood to pause: L1 speakers are more likely to pause before lower frequency nouns as compared to higher frequency nouns. This may be explained by postulating that speakers may have trouble while retrieving low frequency referents. The likelihood that speakers run into trouble when retrieving low frequency words is of the same order of magnitude for L1 and L2 speakers, of different levels of proficiency (as no interactions for the effect were found).

It should be noted that the current study is limited in that only intermediate to advanced learners of Dutch participated. It is unclear whether beginner to intermediate learners may show differential effects. Additionally, with the current sample of L1 Turkish and L1 English speaking in their L2 Dutch it is difficult to draw conclusions on differential effects for different language backgrounds. In some cases, we found an effect of language background. Firstly, compared to English speakers, Turkish speakers tended to use more filled pauses within utterances (and before nouns, also more silent pauses). Secondly, Turkish speakers' silent pauses within utterances tended to be shorter. It is beyond the scope of this paper, however, to investigate such potential differences further.

In spite of these limitations, we may conclude that L2 speakers pause more often and for longer durations within utterances than L1 speakers do; at utterance boundaries, the current study did not find any differences. This finding is compatible with the claim that L1 speakers are able to speak 'one clause at a time', but that L2 speakers may need to pause to formulate their message mid-utterance. For the language testing practice, these findings have implications in that (at least for intermediate to advanced learners) pauses between utterances are not informative because they reflect conceptual planning. It is only the pauses that occur within utterances that are indicative of L2 proficiency and are therefore informative for distinguishing on the basis of aspects of fluency. This suggests that both human judgements and automatic scores in language testing should take pause locations into account. As a matter of fact, human judges will already partly do this even without having been told to do so. For instance, Butcher (1981) showed that pauses at junctures need to be longer in order to be perceived by at least 75% of the listeners (> 220 ms) than pauses within utterances (> 80 ms).

## References

- Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers. 1995. The Celex lexical database (CD-ROM).
- Bates, Douglas, Martin Maechler & Ben Bolker. 2012. lme4: Linear mixed-effects models using Eigen and syntax. R package version 0.999375-39.

- Beglar, David & Alan Hunt. 1999. Revising and validating the 2000 word level and university word level vocabulary tests. *Language testing* 16(2). 131-62.
- Bortfeld, Heather, Silvia D. Leon, Jonathan E. Bloom, Michael F. Schober & Susan E. Brennan. 2001. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and speech* 44(2). 123-47.
- Davies, Alan. 2003. *Native Speaker : Myth and Reality*. Clevedon: Multilingual Matters.
- De Jong, Nivja H. & Hans Rutger Bosker. 2013. Choosing a threshold for silent pauses to measure second language fluency. In Robert Eklund (ed.), *Proceedings of Disfluency in Spontaneous Speech Diss 2013*, 17-20. Stockholm: Royal Institute of Technology (KTH).
- De Jong, Nivja H., Margarita P. Steinel, Arjen F. Florijn, Rob Schoonen & Jan H. Hulstijn. 2012. Facets of speaking proficiency. *Studies in Second Language Acquisition* 34. 5-34.
- De Jong, Nivja H., Rachel Groenhout, Rob Schoonen & Jan H. Hulstijn. 2015. Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics* 36(2). 223 - 243.
- De Jong, Nivja H., Margarita P. Steinel, Arjen F. Florijn, Rob Schoonen & Jan H. Hulstijn. 2013. Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics* 34. 893-916.
- Derwing, Tracy M., Murray J. Munro, Ron I. Thomson & Marian J. Rossiter. 2009. The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition* 31(04). 533-57.
- Foster, Pauline, Alan Tonkyn & Gillian Wigglesworth. 2000. Measuring spoken language: A unit for all reasons. *Applied Linguistics* 21(3). 354-75.
- Goldman-Eisler, Frieda. (1958). Speech production and the predictability of words in context. *Quarterly Journal of Experimental Psychology* 10, 96-106.
- Hartsuiker, Robert J. & Lies Notebaert. 2010. Lexical access problems lead to disfluencies in speech. *Experimental psychology* 57(3). 169.
- Housen, Alex & Folkert Kuiken. 2009. Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics* 30(4). 461-73.
- Hox, Joop. 2010. *Multilevel analysis: Techniques and applications*. New York: Routledge.
- Iwashita, Noriko, Annie Brown, Tim Mcnamara & Sally O'Hagan. 2008. Assessed Levels of Second Language Speaking Proficiency: How Distinct?. *Applied linguistics* 29(1). 24-49.
- Kircher, Tilo T., Michael J. Brammer, Willem Levelt, Mathias Bartels & Philip K. McGuire. 2004. Pausing for thought: engagement of left temporal cortex during pauses in speech. *NeuroImage* 21(1). 84-90.
- Maclay, Howard & Osgood, Charles E. (1959). Hesitation phenomena in spontaneous English speech. *Word* 15. 19-44.
- MacWhinney, Brian. 2000. *The CHILDES Project: Tools for analyzing talk*. Mahwah, NJ: Erlbaum.

- Pawley, Andrew & Frances H. Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (eds.), *Language and communication*, 191-226. London: Longman.
- Pinheiro, José C. & Douglas M. Bates. 2000. *Mixed-Effects models in S and S-PLUS*. New-York: Springer Verlag.
- Riazantseva, Anastasia. 2001. Second language proficiency and pausing. *Studies in Second Language Acquisition* 23. 497-526.
- Riggenbach, Heidi. 1991. Toward an Understanding of Fluency: A Microanalysis of Nonnative Speaker Conversations. *Discourse Processes* 14(4). 423-41.
- Schmidt, Richard. 1992. Psychological mechanisms underlying second language fluency. *Studies in second language acquisition* 14. 357-.
- Skehan, Peter & Pauline Foster. 2007. Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing Research. In Siska Van Daele, Alex Housen, Folkert Kuiken, Michel Pierrard & Ineke Vedder (eds.), *Complexity, accuracy, and fluency in second language use, learning, and teaching.*, 207-226. Brussels: University of Brussels Press.
- Swerts, Marc. 1998. Filled pauses as markers of discourse structure. *Journal of pragmatics* 30(4). 485-96.
- Tavakoli, Parvaneh. 2011. Pausing patterns: differences between L2 learners and native speakers. *ELT journal* 65(1). 71-9.
- Towell, Richard, Roger Hawkins & Nives Bazergui. 1996. The development of fluency in advanced learners of French. *Applied linguistics* 17(1). 84-119.
- Zareva, Alla, Paula Schwanenflugel & Yordanka Nikolova. 2005. Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition* 27(4). 567-95.

# Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency

PARVANEH TAVAKOLI

## *Abstract*

*The study reported in this paper challenges current models of measuring second language fluency by comparing monologic versus dialogic task performance, and providing a novel insight into the measurement of the interactive aspects of dialogic performance. The data that constitute 35 monologic and dialogic task performances from second language learners were coded using a battery of established measures known to tap different aspects of fluency, and subjected to statistical analysis to test for overlaps or differences. Interactive aspects of fluency in dialogue, e.g. interruptions, overlap and unclaimed between turn pauses were also investigated to compare with common measures of monologic speech. While the results confirm previous research findings suggesting that performance is in general statistically more fluent in a dialogue in terms of speed, length of pause and repair measures, they indicate that performances in the two modes are not different in terms of number and location of pauses. The analysis of the dialogues indicates that the decisions researchers make about measuring the interactive aspects of fluency would have an impact on the outcome of measurements of fluency. These findings highlight the need for developing a more systematic and reliable approach to measuring second language (L2) fluency.*

Key words: L2 fluency; monologue; dialogue

## **1. Introduction**

The extensive work in task-based language teaching research investigating the development of second language (L2) ability in instructed settings suggests that complexity, accuracy and fluency are three principal dimensions that encapsulate language proficiency (see Housen and Kuiken 2009 for a full account). While this body of research recognizes fluency as being an essential component of communicative language ability, as well as an important performance descriptor and a key indicator of L2 development (de Jong et al. 2012; Housen and Kuiken 2009; Kahng 2014; Skehan 2014), a review of the literature in this area highlights three important limitations: a) there are gaps in our understanding of L2 fluency as a construct (Kahng 2014;

Prefontaine 2013), b) the findings of fluency research often display mixed results due to the lack of a systematic approach to measuring fluency (Kormos 2006; Skehan 2014), and c) there are concerns about operationalizing and measuring fluency validly and reliably (Housen and Kuiken 2009; Housen et al. 2012). At a theoretical level, research in this area has led to new developments about defining the construct of fluency (Foster 2013; Segalowitz 2010) and offering a more in-depth understanding of how fluency operates in L1 and L2 (de Jong et al. 2012). However, it is important to note that L2 fluency has hardly been researched on its own, since most studies examine fluency as one of several aspects of L2 performance, e.g. accuracy and syntactic and lexical complexity, within a general construct of L2 proficiency (e.g. Kormos and Denes 2004; Skehan and Foster 1996). Acknowledging a separate agenda for researching the kind of discourse-level issues involved in interactional competence e.g. in language classroom settings (Walsh 2013), it is noted that within the cognitive framework of researching L2 speech, L2 fluency has largely been investigated in monologic mode, with limited attention to measuring L2 fluency in task types that involve interaction between speakers. It is also possible to argue that current approaches to conceptualizing fluency are congruent with Levelt's (1989) widely accepted three-stage model of speech production, *Conceptualization*, *Formulation* and *Articulation*, in which language processing and production is defined on the basis of a monologic perspective to performance. Although recent research has shed light on a number of significant aspects of defining and measuring fluent monologic task performance (Kormos and Denes 2004; Segalowitz 2010), little systematic research has been done to discover the way fluent interaction and effective communication can be defined in interactive tasks, or what similarities and differences distinguish fluency of performance elicited by monologic versus dialogic tasks. This lack of understanding and the inconsistency of measurement limit the reliability of models we currently use to discuss core theoretical issues of speech planning and retrieval of linguistic knowledge for speech in real time. This study is therefore an attempt to provide a more in-depth understanding of the nature of fluent performance across the two modes.

## **2. Defining fluency**

One of the earliest definitions of fluency frequently cited is Fillmore (1979) who defined fluency as "the ability to talk at length with few pauses; the ability to fill time with talk; the ability to talk in coherent and semantically dense sentences; the ability to have appropriate things to say in a wide range of contexts; and the ability to be creative and imaginative in the language use (Fillmore 1979: 51). Fillmore's definition, although

proposed for L1 fluency, underlined the complex and multifaceted nature of fluency and highlighted the main factors and processes that make fluent speech possible. Further research in this area, e.g. Freed (2000) suggested that fluency was made up of different characteristics of speech ranging from its psychological manifestations, reflections on underlying speech-planning and thinking processes, to speech production, hesitation phenomena, and temporal dimensions of speech. Segalowitz (2000: 202) called for researchers to distinguish between cognitive aspects, i.e. “the efficiency of the operation of the cognitive mechanisms underlying performance” and performance aspects of fluency, i.e. “the observable speech, fluidity and accuracy of the original performance”. In a more recent publication, Segalowitz (2010) proposed that L2 fluency comprises three distinct but inter-related concepts: Cognitive, utterance and perceived fluency. While cognitive fluency, in this framework, is concerned with mobilizing and integrating the underlying cognitive processes involved in language production (Segalowitz 2010: 48), utterance fluency refers to the measurable aspects of fluency such as speed, pausing and hesitation, and perceived fluency represents the inferences listeners make about someone’s cognitive fluency based on their perceptions of how fluent the speaker is. From a research perspective, it is intriguing to see that definitions of fluency, e.g. Fillmore’s, Freed’s and Segalowitz’s, have conceptualized the construct of fluency as a characteristic of monologic speech without considering or discussing its representation and operation during speech when two or more speakers interact with one another. Therefore a key question the current study seeks to answer is whether the same aspects of fluency unequivocally characterize fluent speech in monologues and dialogues.

### *2.1. Studying fluency in monologic and dialogic task performance*

Monologue and dialogue are two frequently used modes of oral language in both real life and pedagogic contexts. Although it has been argued that dialogues, given their interactive nature, represent language more authentically and naturally (Guillot 1999; Van Lier 2004), research in SLA has predominantly focused on measuring monologic performance elicited by tasks such as oral narratives (Skehan and Foster 1996; Tavakoli, 2011), short talk (de Jong and Perfetti 2011), and answering-machine message leaving tasks (Mehnert 1998). Frequent use of monologues in L2 fluency research can be attributed to a number of factors including a) the degree of control associated with a monologic task performance (i.e. simpler pragmatic demands for speech planning), b) predictability of the outcome of the performance, and c) clarity and ease of the procedures for measuring language produced in a monologic task. On the other hand, measuring fluency in dialogic tasks can prove difficult not only because of the complex

pragmatics involved in dialogue, leading to a less controlled and less predictable nature of performance in this mode, but more importantly because of the difficulty associated with measuring the interactive aspects of dialogues, e.g. overlap, unclaimed between-turn pauses, and the interdependence of the interlocutors' performances.

A limited number of studies have used dialogic tasks to investigate L2 fluency, and only few have examined the differences in the same speakers' fluency when they perform a monologue and a dialogue. Michel (2011), in a between-participant design, examining the effects of task complexity, i.e. "the level of challenge that a task is likely to contain" (Skehan 1998: 134), and interaction on L2 performance, used the same decision making task in two modes: a) a monologic answering machine message leaving task, and b) a dialogic telephone conversation task. The results of her study indicated that the dialogic mode elicited language of higher fluency in terms of speed, pausing and repair measures. The differences between the two modes of performance for all measures of fluency in Michel's study (2011) reached statistically significant levels, with noticeable effect sizes observed for repair and pausing behaviour. The surprisingly non-significant results from the effects of task complexity on fluency can perhaps be explained in terms of how task complexity was operationalized in this study, i.e. since both tasks involved a comparable degree of reasoning and required a decision to be made they did not vary adequately in terms of complexity. With regards to the effects of mode, Michel (2011) argues that speakers may find dialogues cognitively less demanding to perform, not because they find interactive dialogue easier to engage with fluently in terms of pragmatic or task complexity, but because speakers can use the interlocutor's turn to plan their own subsequent performance.

In a study focusing on the development and measurement of fluency in monologue and dialogue, Witton-Davies (2014) used a picture story retelling monologue and a discussion dialogue to investigate the development of fluency in L2 speakers over a four-year period. His findings supported previous research and confirmed that performance in dialogues was consistently more fluent than that in monologues, with higher speech rates, less pausing and fewer repair words being the key characteristics of dialogic performance. Neither of the above studies explored whether the choice and operationalizations of fluency measures in a dialogue had an impact on the results; this is what the current study aims to shed light on.

## *2.2. Characteristics of a dialogue*

While a monologue involves production of sequences by one speaker, a dialogue is “prototypically a joint enterprise involving more than one person” (Cameron 2001: 87), with the speakers taking turns to talk. Edwards (2008) reports that categories that make a dialogue different from a monologue include between-turn pauses, interruptions by the second speaker, and simultaneous talk. Turn-taking seems to be of central importance in a dialogue (Cameron 2001; Edwards 2008) since for a conversation to work in an ideal manner at any single moment one speaker’s talk is followed by a short silence before the next speaker takes the turn to speak. However, real life dialogues are normally far from ideal in terms of the turn taking principles. In the case of simultaneous talk, i.e. overlap, one speaker will normally win the floor and therefore the other becomes silent (Cameron 2001). And after a period of silence, normally one of the speakers breaks the silence. Turn taking is not planned in advance in a normal dialogue; rather, it develops when the speakers engage in the conversation (Cameron 2001; Wilson and Zimmerman 1986). Sacks et al. (1974) proposed a 3-step procedure for turn taking in English in which a) current speaker chooses the next speaker; b) next speaker(s) self-select themselves; and c) current speaker may continue with their turn after the silence. Although it seems simple and straight forward, the structured procedure may not be observed in everyday conversations all the time. Highlighting turn taking as one of the most salient features of social interaction, Wilson and Zimmerman (1986), among others, argued that it should not be viewed as a simple exchange of stimulus and response. Rather, turn taking is fundamentally a collaborative activity that develops in a less structured manner. Research in conversation analysis has shown that a key mechanism in the organization of turns in a conversation is the ability to anticipate the moment of completion of a current speaker’s turn, what is known as *projection* (Lerner 2003; Schegloff 2000; Schegloff 2001). Whereas previous research (e.g. Caspers 2003) emphasised the role of intonation in projection, De Ruiter et al. (2006) argue that knowledge of lexicosyntactic content, i.e. lexical and syntactic characteristics, of an utterance is necessary and perhaps sufficient for both predicting projection and regulating conversational turn taking. Such findings imply that second language learners whose L2 knowledge of lexicosyntax is not yet adequately developed may find it difficult to anticipate projection. It is necessary to note that turn-taking and projection are reported to be culturally shaped and determined, and therefore it may vary across different discourse communities (Cameron 2001; De Ruiter 2006; Edwards 2008). The data analysis (Section 4.3) will examine turn taking patterns, interruptions and overlap speaking time in the dialogue data.

### 2.3. *Measuring fluency*

In an attempt to create a more systematic approach to measuring fluency, Skehan (2003), and Tavakoli and Skehan (2005) suggested that fluency should be measured with regard to its three main characteristics: a) speed fluency, i.e. speed with which speech is performed, b) breakdown fluency, the pauses and silences that break down the flow of speech, and c) repair fluency, hesitations, repetitions and reformulations that are used to repair speech during the production process. Following from this, Skehan (2014) suggests that when measuring fluency composite measures that blend *speed* and *flow* of speech, e.g. phonation time and length of run should also be considered. Recent research findings suggest that some measures of fluency are internally related and, if not chosen carefully, one measure may overlap with others (Kormos 2006; Skehan 2014; Tavakoli and Skehan 2005). Identifying the best measures of fluency that can reliably encapsulate L2 utterance fluency and minimize the possible overlap between different measures, Witton-Davies (2014) and Mora and Valls-Ferrer (2012) suggest that pause length, pause frequency, pause location, mean length of run, speech and articulation rates, phonation time ratio, and a selection of repair measures are the most reliable measures of utterance fluency. Prefontaine (2013) reports that mean length of run and average pause time are two measures of utterance fluency that most strongly relate to self-perceptions of fluency. Kahng (2014: 810) reports that speech rate and mean length of run are strongly associated with both L2 oral proficiency and perceived fluency, whereas articulation rate and repair measures are not. It is beyond the scope of this article to reflect on how these studies illuminate all aspects of the speech planning process as well as performance in real time, but we can pick out that certain measures are crucial in understanding what is going on when speakers are engaging in dialogue and how we can reliably measure their speech fluency. In order to fulfil the aims of this study, most relevant of the fluency measures identified from previous research and some novel measures were selected:

#### Speed

5. Articulation rate: mean number of syllables per minute divided by mean amount of phonation time (excluding pauses)
6. Speech rate: mean number of syllables per minute divided by total time (including pauses)

#### Break down

7. Mean length of pauses per 60 seconds
8. Mean number of pauses per 60 seconds (clause-internal versus clause-external)

#### Repair

9. Repair measures: mean number of partial or complete repetitions, hesitations, false starts and reformulations
10. Mean number of filled pauses, e.g. *em* and *er*

### Composite

11. Mean length of run: the mean number of syllables between two pauses<sup>5</sup>
12. Phonation time ratio: time taken to perform the task (excluding pauses)

Dialogue only measures (not previously investigated)

13. Number of turns and number of interruptions

A key contribution of this study is using dialogue-only measures which will not only provide an insight into aspects of dialogic performance, but allow for a comparison of fluency across both modes.

### 3. Research questions

The following research questions guided the study.

1. Does L2 speakers' fluency remain the same in monologic and dialogic task performance? If not, what are the differences between the two modes of performance?
2. Which aspects of fluency in dialogic task performance are affected by the way measurement of fluency, in terms of turns, pauses and overlaps, is operationalized?

### 4. Methodology

#### 4.1. *Participants, procedures and tasks*

The participants were 35 EAP students enrolled on a pre-session course at a university in the UK. They were at B2 level (CEFR) and were placed on their course based on their IELTS Score (5.0 or 5.5). They were aged between 22 and 35, and had a range of diverse L1s including Arabic, Chinese, Kurdish, Russian and Thai. The participants had been on their EAP courses for four weeks when the experiment took place. For the purpose of the study, they performed a monologue and a dialogue in one of their speaking classes. The choice of the tasks was guided by three main criteria: a) the tasks were in line with the course objectives, i.e. improving the learners' speaking and listening ability, b) the instructors considered the task as interesting, relevant and at the right level, and c) the task types were familiar but the topics had not been covered in the course before so that any possible practice effect can be avoided. The monologue was a retelling of a

---

<sup>5</sup> It should be noted that following from de Jong et al., (2012) a pause is an unfilled silence of longer than 0.25 a second.

recent personal shopping experience, for which they had 1 minute to plan and 1 minute to perform the task. The dialogue involved a discussion task that required the participants to present the case for or against a particular topic e.g. which is better: watching a movie at home or in the cinema. They had 1 minute to plan and 3 minutes to perform the task. The tasks, planning time and time on task were piloted with a different group of learners before the experiment took place<sup>6</sup>.

#### 4.2. *Data analysis*

Task performances were digitally recorded, transcribed and coded for a range of fluency measures (Section 2.3 above). While many of the participants spoke longer than one minute for each task, for the purpose of the analysis, all temporal measures were based on the first 60 seconds of their performance beginning when they actually started speaking. PRAAT (Boersma and Weenink 2013) software was used to measure temporal aspects of fluency e.g. phonation time, length of pause and articulation rate. For PRAAT measurement, intra-rater reliability was used for a 10% sample of the data and coefficient measures of above 95% were achieved. For the rest of the measures, e.g. the number of filled pauses and repairs, coding was done manually with a researcher reading through the text and coding the transcripts. To ensure reliability of the coding of these measures, initially a 10% sample of the data were coded by a second researcher. This was repeated for a second and sometimes a third time until a 90% inter-rater reliability was obtained.

#### 4.3. *Measuring fluency in a dialogue*

As noted above, analyses of number of turns and number of interruptions were carried out on the dialogue data. Careful examination of the data suggests that in contrast to a monologue, in a dialogue the speakers' fluency depends at least to some extent on the interlocutor's conversational skills in using English in an international setting, e.g. turn-taking and, their willingness to communicate (Cameron 2001). Therefore, factors such as how dominant, passive, or involved (Edwards 2008; Tannen 1994) the interlocutors were, the number of turns taken by each speaker, the interruptions and overlap in

---

<sup>6</sup> It is important to note that in a within-participant design, it is very difficult to use the same topic in different modes without a practice effect. As such, task topic and task mode are typically confounded in research of this type, and therefore, the results obtained may at least be partly attributed to the interaction between topic and mode.

speaking time, and the unclaimed between-turn pauses affected different aspects of fluency in the dialogues. For the purpose of measuring fluency in the dialogues, a number of steps were taken. First, to ensure that tasks were performed interactively, any data from interlocutors who had a very unbalanced dialogue, with one speaker dominating for a long period (e.g. more than 70% of the time) or the other remaining quiet for an extended period were excluded. One minute of each speaker's performance was used for the analysis of temporal aspects, i.e. the analysis of a dialogue included two 60 seconds performances. For features such as number of turns, interruptions and repair measures the whole performance was examined. The data were checked to make sure each participant had at least two turns in the 60 seconds of their performance in a dialogue. To interpret silence, particularly in the between-turn pauses, the context in which it had occurred was also considered, e.g. whether the speaker paused to look for a lexical item or to signal the end of their turn (Kurzon 2013).

In terms of interruptions, overall there were not many interruptions in the data. Most participants did not interrupt their interlocutor frequently (mean=4); only three participants interrupted their interlocutors more frequently (mean=13). A qualitative examination of the data suggested that interruptions were mainly a sign of 'high involvement' (Tannen 1994) rather than dominance in the dialogue. A large number of these interruptions (70%) were non-lexical filled pauses (mhm) or short back channels like, 'yeah'. The rest of the interruptions (30%) were either lexical phrases like 'you're right,' or the beginning of a new turn such as 'but when you...'. All the lexical interruptions or short back channels were considered as overlapping speaking time which was considered to belong to both participants, and as such were included in the measurement of fluency for both speakers.

Kurzon (2013) argues that there are three types of conversational silence in informal situations: a) a short silence or pause where the speaker does not respond immediately, b) a voluntary silence where one of the speakers is asked a question but s/he intentionally keeps quiet, and c) a silence in a multiparty conversation where one of the participants chooses to remain silent while others are conversing. A qualitative examination of the data suggested that the unclaimed between-turn pauses did not belong to any of Kurzon's (2013) categories. These were typically long pauses after one speaker came to the end of their utterance and remained silent, while the other speaker also kept quiet because they were either uncertain if the first speaker's turn was complete, or not prepared to speak. Given that it was difficult to attribute such pauses to individual speakers, one way to deal with the unclaimed between-turn pauses was to exclude them from the analysis. A second option, deemed suitable for this study, was to

divide the pauses equally between the two speakers<sup>7</sup>. The data analysis below includes data from both ways of measuring these pauses.

## 5. Results

In order to see whether there were significant differences between fluency measures in monologic and dialogic task performance, a number of t-tests were run to compare the participants' fluency elicited by the monologue and dialogue tasks, and Cohen's (1998) definitions of effect size used. Table 1 shows the results of the t-tests for these comparisons. The measures shown here, as noted in the Methodology Section above, are identified by the relevant literature as the core measures tapping into utterance fluency.

*Table 1. Results of t-tests comparing fluency measures in monologic and dialogic performance (between-turn pauses excluded)*

| <b>Temporal Measures</b>          | <b>Monologue<br/>mean (SD)</b> | <b>Dialogue<br/>mean (SD)</b> | <b>T</b> | <b>P</b> | <b>Cohen<br/>d</b> |
|-----------------------------------|--------------------------------|-------------------------------|----------|----------|--------------------|
| Articulation rate                 | 192.65<br>(23.96)              | 225.08<br>(26.77)             | 6.81     | .001*    | 1.27               |
| Speech rate                       | 143.25<br>(23.33)              | 175.23<br>(22.16)             | 8.40     | .001*    | 1.37               |
| Mean length of pause (in seconds) | .67 (.15)                      | .51 (.08)                     | 6.35     | .001*    | 1.33               |
| Number of pauses clause-internal  | 13.85 (5.27)                   | 13.46 (6.68)                  | .29      | .773     | .06                |
| Number of pauses clause-external  | 8.86 (3.26)                    | 8.37 (3.47)                   | .73      | .471     | .15                |

<sup>7</sup> One may argue that a third option was to use the principles of Conversation Analysis to analyse the data to determine who the unclaimed pauses could be attributed to. However, given the lack of a shared L1 among the participants and the possibility of different conversational norms and practices among them, this option was ruled out.

|                      |              |              |      |       |     |
|----------------------|--------------|--------------|------|-------|-----|
| Repairs              | 7.34 (3.60)  | 5.48 (4.11)  | 2.19 | .04   | .48 |
| Filled pauses        | 12.80 (4.80) | 15.88 (8.82) | 2.01 | .05   | .43 |
| Mean length of run   | 6.62 (2.61)  | 7.95 (2.74)  | 2.99 | .005* | .50 |
| Phonation time ratio | 74.48 (6.93) | 80.37 (8.17) | 3.82 | .001* | .78 |

*All temporal measures are per 60 seconds of each speaker's talk.*

The results indicate that there were significant differences between many of the fluency measures in the two modes of task performance favouring dialogic speech with the participants producing longer runs ( $t=2.99$ ;  $p= .005$ ;  $d=.50$ ), shorter pauses ( $t=6.35$ ;  $p= .001$ ;  $d=1.33$ ), higher phonation time ratios ( $t=3.82$ ;  $p= .001$ ;  $d=.78$ ), and faster articulation rate ( $t=6.81$ ;  $p= .001$ ;  $d=1.27$ ) and speech rate ( $t=8.40$ ;  $p= .001$ ;  $d=1.37$ ) in dialogic task performance. All these significantly different measures showed medium to large effect sizes (Cohen 1998). The participants also produced more filled pauses ( $t=2.01$ ;  $p= .05$ ;  $d=.43$ ) and fewer repair measures ( $t=2.19$ ;  $p= .04$ ;  $d= .48$ ) in the dialogic task with noticeable effect sizes (Cohen 1998). However, these two  $p$  values should be interpreted with care because when a Bonferroni adjusted alpha level is considered a significance level of  $p < .005$  is set, suggesting  $p$  values of .04 and .05 should not be taken as statistically significant in this context. As for number of pauses, although there were more pauses both in the middle of clauses and at clause boundaries in monologic task performance, the differences between the two modes were generally small and negligible. Apart from pausing, the results clearly support existing findings that speakers are more fluent in dialogue than in monologue, using standard measures.

As discussed above, there are a number of measurement decisions that can affect the results of fluency measures in dialogic task performance. The first important aspect that can affect measurement of fluency is the way the between-turn pauses are operationalized, i.e. a) whether the pauses are included in the measurement of different aspects of fluency, and b) if so, who is responsible for the pauses between two speakers' turns. This is an aspect of fluency measurement that has not been discussed in fluency studies before. In the analysis presented in Table 1 above, the between-turn pauses were excluded from the measurement. This is to say, all the silent pauses made between the different turns were not attributed to either of the two speakers engaged in the dialogue.

The current study was concerned that the significant differences observed in different fluency measures between the two modes could (at least partly) be explained by

excluding the between-turn pauses. As such, alternative analyses were needed in which a) the pauses were included in the new measurement, and b) the pauses were equally divided between the two speakers. Consequently, a second set of measurements was used to calculate all temporal aspects of fluency with the pauses divided between the two speakers and included in the measurement. For example, if there was a two-second unclaimed pause between two speakers' turns, each speaker was credited for one second of the pause. It is important to note that when these pauses are included in the measurement, mean length of run and articulation rate will not be affected as they do not involve pause duration or pause frequency. Given that the between-turn pauses happen at clause boundary, the number of clause-internal pauses will not change either. Hence, these measures are excluded from the next set of t-tests. Table 2 demonstrates the differences between the temporal measures of fluency where pauses were excluded from the analysis as shown in Table 1 above compared with the new measures in which they are included in the analysis and divided between the speakers.

*Table 2. Results of t-tests comparing fluency measures in dialogic performance with the between-turn pauses included and excluded*

| Temporal Measures                | Dialogue measures when pauses excluded (as in Table 1 above)<br>mean (SD) | Dialogue measures when pauses included and divided between speakers<br>mean (SD) | <i>T</i> | <i>P</i> | <i>Cohen d</i> |
|----------------------------------|---|--|----------|----------|----------------|
| Speech rate                      | 175.23 (22.16)  | 170.62 (22.16)   | 5.45     | .001*    | .21            |
| Mean length of pause (in second) | .51 (.08)   | .54 (.10)  | 3.47     | .001*    | .33            |
| Mean number of pauses            | 21.83 (8.19)  | 23.54 (7.93)   | 6.48     | .001*    | .21            |
| Number of pauses clause-external | 8.37 (3.47)   | 10.09 (3.72)   | 6.48     | .001*    | .48            |
| Phonation time ratio             | 80.37 (8.17)  | 76.65 (7.72)   | 5.56     | .001*    | .47            |

*All temporal measures are per 60 seconds of each speaker's talk.*

The results indicate that when the between-turn pauses are included in the analyses and divided between the two speakers, performance in dialogic tasks, previously seen as more fluent on most measures than monologic performance, became significantly poorer, or less fluent, compared to dialogic scores in the previous analysis – i.e. when unclaimed pauses between the two speakers' turns were excluded. This implies that the way fluency measures are defined and operationalised has had significant effects on the results obtained from fluency measures. This may mean that the main differences achieved between monologic and dialogic task performance (Table 1 above) might have in effect been caused by the way they are measured. Therefore, a new set of t-tests were run to find out whether with this new measurement, i.e. including the between-turn pauses in the analysis, the significant differences between fluency in monologic and dialogic modes still persisted.

*Table 3. Results of t-tests comparing fluency measures in monologic and dialogic performance (between-turn pauses divided between the speakers and included in the analysis)*

| Temporal Measures                | Monologue<br>mean (SD) | Dialogue<br>mean (SD) | <i>T</i> | <i>P</i> | <i>Cohen d</i> |
|----------------------------------|------------------------|-----------------------|----------|----------|----------------|
| Articulation rate                | 192.65 (23.96)         | 225.08<br>(26.77)     | 6.81     | .001*    | 1.28           |
| Speech rate                      | 143.25 (23.33)         | 170.62<br>(22.16)     | 7.61     | .001*    | 1.20           |
| Mean length of pause (in second) | .67 (.15)              | .54 (.10)             | 4.97     | .001*    | 1.02           |
| Number of pauses clause-internal | 13.85 (5.27)           | 13.46 (6.68)          | .29      | .773     | .06            |
| Number of pauses clause-external | 8.86 (3.26)            | 10.09 (3.72)          | 1.91     | .064     | .35            |
| Mean length of run               | 6.62 (2.61)            | 7.95 (2.74)           | 2.99     | .005*    | .50            |
| Phonation time ratio             | 74.48 (6.93)           | 76.65 (7.72)          | 1.60     | .117     | .30            |

*All temporal measures are per 60 seconds of each speaker's talk.*

The results demonstrate that even when the between-turn pauses are included in the analyses, there are still significant differences between monologic and dialogic task performance with most temporal measures demonstrating more fluent performance in a dialogue. Phonation time ratio is the only temporal measure in which the difference does not reach a statistically significant level with the new analysis ( $t=1.60$ ;  $p= .117$ ;  $d= .30$ ). Although the differences between the two modes are now larger for the number of clause-external pauses, they are not statistically different in monologic and dialogic task performance ( $t=1.91$ ;  $p= .06$ ;  $d= .35$ ). These results suggest that participants' speech in dialogues was more fluent than that in monologues for some, but not all, fluency measures.

## 6. Discussion

The study confirms findings of previous research (Michel 2011; Witton-Davies 2014) that performance in dialogic tasks is more fluent than that in monologic tasks in terms of speed, length of pause and repair measures. The analysis indicates that composite measures, e.g. mean length of run and speech rate are also higher in dialogic task performance. With regard to break-down aspect of fluency, while mean length of pause is statistically shorter in the dialogues, there is little difference between performance in the monologues and dialogues in terms of number and location of pauses. In line with the findings of previous research (Tavakoli 2011), the participants in this study paused more frequently in the middle of clauses consistently across the two modes. These findings suggest that while being engaged in a dialogue encourages speed and shortens length of pause, it has little impact on how often and where L2 speakers pause. This finding suggests that speakers use pauses to monitor their speech production process (de Jong et al. 2013; Kormos 2006; Michel 2011), and to pay attention to form in terms of accuracy of their performance (Tavakoli et al. in print).

It is possible to argue that the collaborative and interactive nature of a dialogue makes it more convenient for the L2 speakers to produce a more fluent performance. Previous research reports that the collaborative nature of a dialogue allows interlocutors to use their partner's turn to plan for their utterances (Lochbaum 1998; Webber 2008). This may mean that having "listening time" helps speakers with what Levelt (1989) considers the *conceptualisation phase* (where the preverbal message is generated) and *reformulation phase* (where the preverbal message is converted into a phonetic plan for speech) of speech production. But beyond this individual aspect of improving performance, the interactive nature of having a partner in dialogue may genuinely

encourage speakers to show greater willingness to communicate interactively, and take the interlocutor's needs into account by producing fewer hesitations and repetitions and faster speech. The frequent use of filled pauses in the dialogues, compared to those in monologues, is an interesting example of taking the partner's needs into account. More qualitative research, e.g. stimulated recall protocols, is needed to provide a better understanding of factors that encourage a more fluent speech in an interactive task such as a dialogue versus a monologic performance such as retelling of a story.

Another important finding of the study is that the decisions researchers make about the measurement of fluency in a dialogue may affect the different temporal aspects of L2 fluency. For instance, the data analysis on Table 2 demonstrates that the way between-turn pauses are defined and operationalised affected measures of phonation time ratio, speech rate, mean length of pause and number of pauses. This may mean that the reliability of such measures across the two modes depends on the way they are operationalised. Whereas previous research (Tavakoli et al. in print; Mora and Valls-Ferrer 2012; Witton-Davies 2014) suggests that pause length, pause location, mean length of run, speech and articulation rates, phonation time ratio, and repair measures are the most reliable measures of fluency in a monologue, this study indicates that the same measures may not be the most reliable representatives of fluency in a dialogue since they may be affected by the decisions about overlap speaking time and between-turn pauses. Further research is needed to investigate reliability of the measures SLA researchers employ when examining fluency in a dialogic mode.

Finally, in respect to the cognitive construct of fluency itself, de Jong et al. (2013) reported that while all measures of utterance fluency were at least to some extent related to cognitive fluency, measures such as mean length of pause were only marginally linked to cognitive fluency implying that pausing can be better explained by other factors, e.g. individual differences (de Jong et al. 2013). The findings of the current study about the differences between speakers' speaking patterns in a dialogue suggest that further research is also needed to examine the extent to which task mode and speaker's conversational skills can explain cognitive fluency.

## **7. Conclusion**

Measuring fluency in monologic and dialogic L2 performance seems to raise both theoretical and methodological issues. At a theoretical level, the findings of the current study raise the question of whether fluency represents the same construct in a monologue compared with a dialogue. In a monologue, where the sole responsibility of

speech is on one speaker, fluency is predominantly characterised by the flow of speech, its speed and a lack of dysfluency measures. In a dialogue where speakers rely on one another to produce connected speech, however, fluency is more than flow and speed, as the interlocutor's conversational skills and taking care of their needs gain significance. For this reason, the speakers' attention is not focused on flow and speed only, but it is distributed to other factors such as producing fewer repairs and more filled pauses to provide a smoother interaction with the interlocutor. This may often be demonstrated through turn-taking, overlap, negotiating meaning and other characteristics of interactive discourse. The findings of this study encourage researchers to take the important aspects of interactive speech into account not only in defining and conceptualising fluency but in modelling speech production. At a methodological level, using the same measures may not be the most reliable and effective way of measuring fluency in both modes. If careful decisions are not made about operationalisation of between-turn pauses, any measure that includes those pauses may be challenged for their consistency and reliability.

## References

- Boersma, Paul & David Weenink. 2013. *Praat: doing phonetics by computer*. Version 5.3.51, retrieved 2 June 2013 from <http://www.praat.org/>
- Cameron, Debora. 2001. *Working with spoken discourse*. London: Sage.
- Caspers, Johanneke. 2003. Local speech melody as a limiting factor in the turn-taking system in Dutch. *Journal of Phonetics* 31(2).251–76.
- de Jong, Nel & Charles Perfetti. 2011. Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning* 61(2). 533-568.
- de Jong, Nivja & Ton Wempe. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods*, 41 (2). 385 - 390.
- de Jong, Nivja, Margarita Steinel, Arjen Florijn, Rob Schoonen & Jan Hulstijn. 2012. facets of speaking proficiency. *Studies in Second Language Acquisition* 34(1). 5–34.
- de Jong, Nivja, Margarita Steinel, Arjen Florijn, Rob Schoonen & Jan Hulstijn. 2013. Linguistic skills and speaking fluency in a second language. *Applied Linguistics* 34(4). 893–916.
- de Ruiter, Jan, Holger Mitterer & Nick Enfield. 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language* 82(3), 515–535.

- Edwards, Jane. 2008. The transcription of discourse. In Deborah Schiffrin, Deborah Tannen & Heidi Hamilton (eds.), *The handbook of Discourse Analysis*, 321-348. London: Routledge.
- Fillmore, Charles. 1979. On Fluency. In Charles Fillmore, Daniel Kempler & William Wang (eds.), *Individual differences in language ability and language behavior*, 85–102. New York: Academic Press.
- Foster, Pauline. 2013. Fluency. In Carol Chapelle (ed.), *The encyclopaedia of Applied Linguistics*. London: Wiley-Blackwell.
- Freed, Barbara. 2000. Is fluency, like beauty, in the eyes (and ears) of the beholder? In Heidi Rigggenbach (ed.), *Perspectives on Fluency*, 243–265. Ann Arbor: University of Michigan Press.
- Guillot, Maureen. 1999. *Fluency and its teaching*. Clevedon: Multilingual Matters.
- Housen, Alex & Fulkert Kuiken. 2009. Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics* 30(4). 461–473.
- Housen, Alex, Fulkert Kuiken & Ineke Vedder. 2012. *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Amsterdam: John Benjamins
- Hymes, Dell. **1972. On Communicative Competence.** In John Pride and Janet Holmes (eds.), *Sociolinguistics, Selected Readings*, 269-293. Harmondsworth: Penguin.
- Khang, Jimin. 2014. Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning* 64(4): 809-854.
- Kormos, Judit, 2006. *Speech production and second language acquisition*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Kormos, Judit & Marian Dénes. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32(2). 145–164.
- Kurzon, Dennis. 2013. Analysis of silence in interaction. In Carol Chapelle (ed.), *The encyclopaedia of Applied Linguistics*. London: Wiley-Blackwell.
- Levelt, Willem. 1989. *Speaking: from intention to articulation*. Cambridge: Cambridge University Press.
- Lerner, Gene. 2003. Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society* 32(2). 177-201.
- Lochbaum, Karen. 1998. A collaborative planning modal of intentional structure. *Computational Linguistics* 24(4): 525-72.
- Mehnert, Uta. 1998. The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition* 20(1). 83–108.

- Mora, Joan & Margalida Valls-Ferrer. 2012. Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly* 46(4). 610-41.
- Prefontaine, Yvonne. 2013. Perceptions of French fluency in second language speech production. *Canadian Modern Language Review* 69(3), 324-348.
- Sacks, Harvey, Emmanuel Schegloff & Gail Jefferson. 1974. A simplest systematics for the organisation of turn taking for conversations. *Language* 50(4). 696-735.
- Schegloff, Emanuel. 2000. Overlapping talk and the organization of turn-taking for conversation. *Language in Society* 29(1).1-63.
- Schegloff, Emanuel. 2001. Accounts of conduct in interaction: Interruption, overlap, and turn-taking. In Jonathan Turner (ed.), *Handbook of sociological theory*, 287-321. New York: Kluwer Academic.
- Segalowitz, Norman. 2000. Automaticity and attentional skill in fluent performance. In Heidi Rigggenbach (ed.), *Perspectives on Fluency*, 200-219. Ann Arbor: University of Michigan Press.
- Segalowitz, Norman. 2010. *The cognitive bases of second language fluency*. New York: Routledge.
- Skehan, Peter. 1998. *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Skehan, Peter. 2003. Task-based instruction. *Language Teaching*, 36(1). 1-14.
- Skehan, Peter. 2009. Modelling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics* 30(4). 501-532
- Skehan, Peter. 2014. Limited attentional capacity, second language performance, and task-based pedagogy. In Peter Skehan (ed.), *Processing Perspectives on Task Performance*, 211-260. Amsterdam: John Benjamins.
- Tannen, Deborah. 1994. *Gender and discourse*. Oxford: Oxford University Press.
- Tavakoli, Parvaneh, Colin Campbell & Joan McCormack. In print. Development of speech fluency over a short period of time: Effects of pedagogic intervention. *TESOL Quarterly*.
- Tavakoli, Parvaneh. 2011. Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal* 65(1): 71-9.
- Tavakoli, Parvaneh & Peter Skehan. 2005. Strategic planning, task structure and performance testing. In Rod Ellis (ed.), *Planning and task performance in a second language*, 239-277. Amsterdam: John Benjamins.
- Van Lier, Leo. 2004. *The ecology and semiotics of language learning: A sociocultural perspective*. Boston: Kluwer Academic.

- Webber, Bonnie. 2008. Computational perspectives on discourse and dialogue. In Deborah Schiffrin, Deborah Tannen & Heidi Hamilton (eds.), *The Handbook of Discourse Analysis*, 798-817. London: Routledge.
- Witton-Davies, Giles. 2014. *The study of fluency and its development in monologue and dialogue*. Unpublished Doctoral Thesis, University of Lancaster.

# A qualitative analysis of perceptions of fluency in second language French

YVONNE PREFONTAINE AND JUDIT KORMOS

## *Abstract*

*In the field of second language (L2) fluency, there is a common adherence to quantitative methods to examine characteristics and features of speech. This study extends the field by reporting on an investigation that analyzed native-speaker listeners' perceptions of L2 fluency in French from a qualitative perspective. Three untrained judges rated students' performance on speech tasks varying in cognitive demand and provided justifications for their perceptions of fluency. The goal of the research was to examine the factors that affect raters' evaluations of fluency in response to three oral performances from 40 adult learners of French of varying proficiency. Qualitative analysis revealed that the main speech features that influenced native listeners' perceptions of L2 fluency were speed, rhythm, pause phenomena, self-correction and efficiency/effortlessness in word choice, but also in target-like rhythm and prosody. The results of using such qualitative methodology highlights the important role that rhythm plays in fluency judgements in syllable-timed languages such as French, a factor which has not always been given much prominence in previous L2 fluency quantitative research.*

Keywords: L2 fluency; language assessment; fluency judgements; second language learning; speech production and perception

## **1. Introduction**

While a number of cross-linguistic studies have used a systematic research approach to evaluate L2 fluency, using temporal variables and native-speaker judgements within the quantitative paradigm (Derwing et al. 2004; Freed et al. 2004; Ginther et al. 2010; Kormos and Dénes 2004; Lennon 1990; Riggenbach 1991; Towell et al. 1996), an approach to investigate perceptions of fluency using qualitative data can provide important insights to account for the complexity of how L2 fluency is perceived by native speakers. Such methods thus complement both mixed-methods research (Bosker et al. 2013; Préfontaine et al. 2015) and the aforementioned purely quantitative studies. In the present study, we extend the scope of fluency research in several directions by examining native speakers' perceptions of the fluency of L2 speech. This research proposes a novel alternative methodology to enquire into the nature of L2 fluency and advocates adopting an introspective approach to account for the complexity of L2 perceptions of fluency.

## 2. Perceptions of fluent speech production

In the second language acquisition (SLA) and language testing literature, definitions of fluency are generally linked to qualitative features of ease, naturalness and appropriateness but more often make mention of more quantitative temporal aspects, such as pause phenomena, speed and the ability to produce fluent runs of speech (Brumfit 1984,2000; Ejzenberg 2000; Fillmore 2000; Kormos 2006; Pawley and Syder 1983; Sajavaara 1987; Schmidt 1992; Segalowitz 2010). With these considerations in mind and for the purposes of this study, L2 fluency will be defined as L2 skilled performance, referring to “rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention under the temporal constraints of on-line processing” (Lennon 2000: 26). This definition is appropriate because it considers the automaticity of both speech production and perception processes between speaker and listener. With regard to *perceived fluency*, we refer to Segalowitz’s (2010: 48) definition of this within his three-part model of fluency, as the “inferences listeners make about a speaker’s cognitive fluency based on their perception of utterance fluency”. While cognitive and utterance fluency are not the subject of this article, the former refers to the listener’s “ability to efficiently mobilize and integrate the underlying cognitive processes” (p. 48) while the latter refers to the “oral features of utterances that reflect the operation of underlying cognitive processes” (p. 48).

Regardless of which definition one subscribes to, there is a common feature underlying most, but not all, characterizations of L2 fluency. There is a widely held notion of speech rate or speed being the main qualifier of L2 fluency. In other words, the faster you speak, the more fluent you are. While the intent of this study is not to review the extensive quantitative L2 fluency research using temporal variables, we argue that these measures do not take full account of the range of differences L2 speakers show in the fluency of their speech and do not offer detailed insights into the wide variety of impressions listeners hold about fluent performance.

Examining issues of L2 fluency in speech production involves the detailed analysis of quantitative and qualitative features that correspond to the intuitive and subjective perceptions of listeners. Speech perception is inherently tied to speech production because human listeners are sensitive to articulation and speech sounds and they process this information to understand language. In the context of L2 perception, rhythmic and prosodic aspects interfere with how listeners process speech (Delattre 1961; Freed 2000; Pell 2001; Wennerstrom 2000, 2001). In particular, stress patterns in the L2 speech stream play a critical role in crosslinguistic speech perception and in the encoding and decoding of sounds whereby speech perception processes dynamically adjust to L2 output. Thus, in the context of L2 fluency, it is vital to understand the interrelation between listeners’ perceptions of speech and aspects of production, with varying degrees of fluency invariably influencing communication and others’ evaluation of speech competence.

Qualitative studies of L2 speech production and perception exploring French fluency are scarce within the SLA research literature. In fact, the only study to

investigate L2 fluency in French from both a speech production and a perception perspective was conducted by Freed (2000). Seeking empirical support for the belief that students who study abroad make more fluency gains, Freed analyzed two groups of French learners, one that went to study in France and another that studied at home. In this study, six native French speakers rated 30 French language learners' speech samples for fluency on a Likert scale of one (not fluent at all) to seven (extremely fluent). The speech samples were based on data collected from oral proficiency interviews (OPIs) in a pretest and posttest format. The results show that the study abroad students increased their OPI scores between Time 1 and Time 2. The untrained raters were asked to justify their observations and to rank the importance of the features of fluency specified, namely, amount of speech, rate of speech (pruned), unfilled pauses, frequency of filled pauses, length of fluent speech runs, repairs and clusters of dysfluency. More than half of the raters chose 'rate of speech', 'smoother speech with fewer false starts', 'fewer pauses/hesitations' and 'better grammar and vocabulary' (p. 254) as crucial to their perceptions of French fluency. Freed also found that the raters sometimes did not use fluidity as the basis for their fluency evaluations, but rather individual speaker attributes, such as tone of voice, accent, confidence in speech and richness of vocabulary. Although overall fluidity was the most noticeable characteristic observed by the raters, L2 fluency ratings were also influenced by factors beyond temporal and hesitation phenomena.

The goal of our study was to compile and analyze three raters' qualitative perceptions of fluent performance produced by L2 learners of French on a set of oral narrative tasks. The research aimed to uncover the linguistic processing experience of listeners when they evaluate L2 speech. Our study sought to answer the question: What are the features of L2 learners' oral production that influence perceptions of L2 fluency in French? As the research in perceived fluency in L2 French is sparse, our study sought to contribute to the existing literature by focusing on perceived fluency with regard to speech features of rhythm particularly, as well as speed, pauses, lexical retrieval, self-correction and efficiency/effortlessness. The study thus offers a novel methodology for gathering qualitative data as a means of providing a framework for understanding the source of fluency perceptions in planning and encoding L2 speech.

### **3. Method**

#### *3.1. Participants*

The context for the study was a 5-week French immersion programme at a large francophone university in Québec, Canada. The L2 speaker participants were 40 volunteer undergraduate and graduate learners registered in beginning, intermediate and advanced French courses. There were approximately 13 participants per level, ranging in age from 18 to 69 years ( $M = 26$  years,  $SD = 10.57$ ). The participants comprised 26 Canadian, 13 American and one British student, of whom 21 were female. The sample was not homogenous in order to allow for sufficient variation and to

characterize the different skill levels in spoken French. The participants were all native speakers of English and varied in their exposure to French language study and the francophone world. They had an average of six years of French instruction in a regular classroom setting with the exception of 10 participants who had attended for an average of nine years in a French immersion setting in another Canadian province outside Québec. While there are several different options of French immersion programmes available in Canada, the 10 participants aforementioned started in Kindergarten and completed in Grade 12.

The rater participants were three native speakers and French language instructors from the same university who were recruited to judge L2 speech production qualitatively. Although the raters had many years of experience in teaching French to non-francophone students, none had previously been involved in any L2 fluency rating projects. The three raters had no contact with the speaker participants inside or outside the classroom and all the L2 speech production was effectively anonymous to the raters.

### *3.2. Instruments*

The study design sought to employ different task types and degrees of task difficulty by varying the cognitive processing load, as Segalowitz (2010) suggests as L2 fluency varies according to task. Thus, the L2 speaker participants were asked to respond to three narrative speech tasks ranging in task complexity, demand and scope. In the first task, participants narrated a story based on six random pictures. The second task, a story retell, entailed retelling a story based on a short text in English about a horseback riding accident. In the final task, participants narrated a story based on an 11-frame cartoon strip presented in chronological order (for more details on the tasks see Préfontaine and Kormos (2015)).

Next, the three raters listened to each of 120 speech performances and gave their written qualitative impressions in which they described the features that most influenced their perceptions of L2 fluency in French. In the qualitative research conducted, the raters were intentionally untrained and were not provided with a definition of fluency to serve as a guide. This procedure was implemented in order to avoid imposing a particular self-fulfilling construct of L2 fluency on the raters. Rather, they were informed of the overall goal which was to reveal what native speakers “perceive in the real world as a listener” when they hear L2 spoken French and what personally influences their perceptions from both qualitative and quantitative perspectives. This open-ended approach allowed the raters to make their own judgments about what constitutes L2 fluency in French, while still providing considerable qualitative detail.

### *3.3. Procedure*

During a one-hour face-to-face data-collection meeting, the speaker participants completed three different narrative speech tasks, for each of which they were allotted

three minutes of planning time. Their speech production lasted from two to five minutes. In order to manipulate the task effects on L2 production, the three tasks were administered in a counter-balanced design to control for task-order effects.

The speech-rating project was conducted online using Google Drive. Each participant's three speech productions were uploaded in a randomized order to ensure that the raters refrained from rating the same student equally across tasks. The raters provided written qualitative comments describing the fluency features that most influenced their evaluations of perceived fluency. All the qualitative comments were provided in an Excel spreadsheet in French and translated into English by the first author.

### 3.4. *Data analyses*

Impressions of fluency were analyzed to identify the speech features that most influenced their evaluations of L2 fluency. Following Rubin and Rubin's (2005) guidelines, we took a miner approach to code, extrapolate and analyze the qualitative data.

## 4. Results and discussion

### 4.1. *Speech features influencing perceptions of L2 fluency*

Table 2 shows the major themes that emerged from the qualitative analysis. Although a number of important issues pertaining to L2 speech perception were identified, such as the expressivity and grammatical competence of L2 speakers and their native-like use of oral discourse features, in our detailed analysis, we focus on the speech features of speed, pauses, lexical retrieval, self-correction, efficiency/effortlessness, and rhythm. These features thus provide data to compare with previous studies and also signal the influence of perception issues arising from French prosody specifically. An examination of the themes indicates that a combination of quantitative and qualitative fluency features appears to be at work when native speakers evaluate the fluency of L2 speech. As the data reveal, listeners' perceptions are determined by a number of dimensions, not all of which are merely temporal.

*Table 2. L2 Perception Fluency Themes, Descriptions, Frequency and Raters' Comments across Tasks*

| Theme                               | Description   | Frequency |
|-------------------------------------|---|-----------|
| Speed                               | Rapidity or rate of speech  | 81        |
| Pause                               | Temporary interruption to the stream of speech  | 172       |
| Phenomena                           |   |           |
| Lexical Retrieval                   | Accessing words or expressions in the mental lexicon  | 60        |
| Self-correction                     | Perceived deficiencies in one's own language output (Dörnyei and Kormos, 1999)  | 48        |
| Efficiency/Effortlessness           | Reference to speaking ease or difficulty and underlying speech planning and processing efficiency in L2 communication | 406       |
| Rhythm                              | The regular, patterned beat of stressed and unstressed syllables and pauses in an utterance.                          | 201       |
| Expressivity/Psychological state    | Expressivity or inner psychological state of the interlocutor conveyed in the voice                                   | 190       |
| Grammatical competence              | Reference to the structural and syntax rules that govern a language   | 140       |
| Native-like oral discourse features | Native-like speech manifestations in spoken discourse   | 136       |

#### 4.2. Speech rate

Previous cross-linguistic research has shown the importance of speech rate and mean length of run as predictors of L2 fluency and as a means to differentiate between proficiency levels (Bosker et al. 2013; Cucchiarini et al. 2002; Derwing et al. 2004; Freed 2000; Freed et al., 2004; Ginther et al. 2010; Iwashita et al. 2008; Kormos and Dénes 2004; Lennon 1990; Préfontaine 2013; Préfontaine et al. 2015; Riggenbach 1991; Towell et al. 1996). In these studies faster speech rate and longer length of run was consistently found to be related to higher fluency scores and levels of proficiency. Compatible with these findings, the qualitative data revealed that speed is a salient quality of speech perception, but it also showed that relationship between perceived fluency and speed might be more complex than indicated in previous quantitative research.

The speed of delivery (you can hear a machine gun) and uncertain pronunciation make the message almost completely incomprehensible. She is not aware that her message will not be understood. (Rater 1, Participant 1)

While speech rate is an important feature of L2 fluency, for these listeners, the faster the speech rate, the lower the comprehension (see also Schwab and Grosjean 2004). The connection between speech rate and raters' impressions of L2 fluency is further exemplified below, where we see that speaking too slowly is not without its problems either:

Her slow speech puts me to sleep. This candidate wants to perform well but at the expense of a normal speech rate. (Rater 3, Participant 6)

From a speech-perception perspective, the raters report slow speech as problematic because it does not catch the listener's attention. These comments are in line with Lennon (2000), who argues that "a good touchstone of acceptable fluency is the degree to which listener attention is held" (p. 34). In a study of ESL learners, Munro and Derwing (2001: 455) report that "the relationship between speaking rate and other L2 dimensions of speech may be curvilinear, rather than linear". In other words, optimal comprehensibility is related to a moderate speech rate. Clark (2002) argues that the most effective messages are delivered 'on time', i.e. with respect to the temporality requirements of the listener.

Moreover, perceptions of fluency are formed not only in relation to speed, but also based on the number of words between pauses. For example:

Sometimes his word choice managed to impress me (*il s'est rendu* instead of *il est allé*) 'he travelled' instead of 'he went', and the same for his speech rate. He tells a part of the story pretty quickly, pauses, then another part, then pauses. The information relayed between pauses is fairly quick and the words seem to come easily. (Rater 1, Participant 17)

The student's speech is not really spirited: it is staccato (3-4 words, a pause; 3-4 words, a pause), ums ... a lot of ums ... that allow her to find the time to invent a sequel. I noticed a long pause in which she seeks the word 'ring', I think. She chose to break her rhythm of narration to try to meet a criterion for correctness of vocabulary. (Rater 2, Participant 34)

As mentioned previously, research has also confirmed the importance of mean length of runs measured quantitatively (Freed et al. 2004; Raupach 1980, 1987; Towell 2002; Towell et al. 1996). It is noteworthy that our data also show how listeners qualitatively describe speech features that are quantitative in nature. Furthermore, the comments reveal that the temporal measures of speech rate and fluent runs are closely observed by the raters, and that they often consider these measures jointly rather than in isolation.

#### 4.3. *Pause phenomena*

Whether in L1 or L2, pausing is an observable and natural occurrence in spontaneous speech and fulfils an important linguistic processing and social function (Clark and Fox Tree 2002). Depending on their place, length and frequency in the L2, pauses can be seen as psycholinguistic indicators of planning processes and speech encoding difficulties (Kormos 2006). The raters' observations in our study with regard to pausing can be divided into two major groups: comments related to filled pauses and unfilled pauses.

In relation to filled pauses, the raters identified the use of the French filler "euh" as contributing to more native-like language use and perceptions of fluency, for example:

Of course, this students hesitates, makes a lot of 'uh' sounds, but he often adds a conjunction (*et, euh, mais, alors, donc*) 'and, um, but, so, therefore' to ensure the link with the preceding sentence. His hesitations seem normal because francophones use the same filler trick. (Rater 2, Participant 9)

The use of fillers such as '*euh, enfin, c'est à dire, bon, alors* etc.' in learners' speech production is highly idiomatic and enhances the degree of L2 fluency because these same fillers are associated with L1 use (Raupach 1980). As demonstrated by the comments, the use of French fillers leaves a favourable impression on L1 listeners. In relation to problem-solving mechanisms, the use of fillers as a communication strategy can compensate for resource deficits in L2 processing (Dörnyei and Kormos 1998). As this rater comments, actual speech planning can be inferred:

I think she is trying to form sentences in her head before saying them, which is why she hesitates so much. (Rater 1, Participant 33)

Disfluent speech, as indicated by excessive pausing, has previously been reported as being one of the major impediments to L2 intelligibility and a source of negative perceptions of French speech performance (Olynyk et al. 1990). The data reveal that hesitations, especially at inappropriate junctures, tend to be viewed unfavourably:

The discourse organization is ok and I can follow the story quite well, but some of the pauses are not in the right place. (Rater 2, Participant 19)

Previous L2 fluency perception research in English has also reported that frequent speech hesitations, especially in the middle of a clause rather than at the end, sound disfluent to

native speakers (Ejzenberg 2000; Pawley and Syder 2000; Riggensbach 1991; Wennerstrom, 2001).

As regards unfilled pauses, the data show that in some cases unfilled pauses are seen as natural and acceptable to the listener:

The pauses sound natural ... like he is looking for his thoughts rather than searching for words or checking grammar. (Rater 3, Participant 17)

The raters' comments reveal that they are sensitive not only to the location of pauses but also to their purpose. If pauses are used by L2 speakers to plan their messages and if they are placed at appropriate junctions, they are not evaluated as signs of dysfluency.

Nevertheless, there were also comments which reflected the raters' uncertainty about the speaker's intent behind pausing. In the quote below the rater describes her own internal debate as she tried to evaluate the purpose of pausing.

I don't know if this student is interrupting his narrative to search for words in French or to organize his thoughts. (Rater 3, Participant 6)

This indicates that raters may not always be able to judge whether pauses are associated with difficulties in lexical retrieval or if, rather, the speaker is in the process of conceptualizing the storyline. The data presented here about raters' perceptions of L2 learners' pausing behaviour might elucidate why the findings of previous quantitative studies on the role of the frequency of pauses in fluency judgements are contradictory. For example, Kormos and Dénes (2004) found no link between fluency ratings and the frequency of unfilled pauses in L2 English, whereas Lennon's (1990) and Foster and Skehan's (1999) studies demonstrate the existence of such a relationship.

#### 4.4. *Lexical retrieval*

Closely related to pause phenomena is the issue of lexical retrieval in speech processing. As Lennon (2000) emphasizes, efficient lexical retrieval is one of the most important factors in L2 fluency. In the data collected, the raters commented on topics pertaining to the ease and difficulty of lexical retrieval and the communication strategies used to overcome problems with lexical access. First, the raters' reports include some comments about their perceptions with regard to the ease of lexical retrieval:

The very few hesitations sounded natural, as if he were catching his breath. It didn't sound as if he was looking for words or didn't know what to say. (Rater 3, Participant 37)

As the excerpt shows, inherent to perceptions of lexical retrieval is the notion of pausing. Therefore, not only ease of lexical retrieval but also apparent stability in lexical knowledge in language processing is paramount to perceptions of L2 fluency.

Second, comments in relation to reduced skills in word retrieval generally pertain to overall difficulty in lexical encoding:

She is looking for words and good structures but unfortunately this makes her errors even more noticeable! (Rater 2, Participant 18)

Thus, obvious searching for words and expressions does not go unnoticed and marks their unavailability in the mental lexicon (Levelt 1989). In these examples, difficulties with efficient lexical retrieval might have been managed by deploying lexical problem-solving strategies (Dörnyei and Kormos 1998; Poulisse 1993). The raters' qualitative comments also highlight the importance of reverting to communication strategies as an alternate means to convey the intended message:

She is searching for words and structures but does not use strategies to circumvent these difficulties. (Rater 3, Participant 35)

Whether for self-correction or as fillers or paraphrasing, using successful communication strategies was very much appreciated by the raters:

I really appreciated his strategy for getting around the words 'to land'. He uses a paraphrase, and although it contains errors, it illustrates well the action of landing a plane (*aller en bas et mettre son avion la où il doit aller*) 'go down and put the plane there where it must go'. (Rater 1, Participant 7)

#### 4.5 Self-corrections

The inclusion of self-corrections in this study was psycholinguistically motivated, because they are often regarded as a marker of dysfluency (Tavakoli and Skehan 2005). Self-corrections, a type of problem-solving mechanism related to one's own output, fall within the realm of communication strategies (CS) (Dörnyei and Kormos 1998). As the term implies, self-corrections are self-initiated and are deployed to help the learner repair a problem arising in their own speech-production processes (Kormos 2000). Three main themes relating to self-corrections are identified in the raters' comments. First, when perceiving L2 fluency in French, the raters reacted favourably to learners who made an effort to self-correct in their oral performances, whether this was effective or not, for example:

In terms of self-corrections, yes, he makes a lot, but they are not the right ones (*nous avons découvert? nous avons découru?*) and even he seems to doubt his own corrections. He is conscious of his grammatical errors and the fact that he puts so much effort into correcting them makes him instantly likeable. (Rater 2, Participant 21)

The raters' comments indicate that native speaker listeners are sensitive to self-correction efforts and appreciate them in L2 spoken French.

Second, the raters detected oral performances in which no effort to self-correct was made. As a consequence, perceptions of L2 fluency in French were often unfavourable:

Pauses are strangely long. She stocks up on ideas then outputs it all out, leaving no room for self-corrections. This student could have easily corrected « *il était soif* » 'he

was thirsty' and « *il faisait de l'éclairage* » – 'there was lightning'. (Rater 1, Participant 5)

Thus, self-corrections are not only valued and appreciated, but were also expected by the native-speaker raters. Previous research shows that by raising awareness of communication strategies such as self-corrections, learners not only smooth out 'trouble spots', but also improve their confidence and self-efficacy and control their speech performance better in oral testing (Dörnyei 1995; Lam 2006; Nakatani 2005, 2010).

Third, the findings also seem to suggest that self-corrections are not a specific L2 marker, and may not negatively impact on perceptions of fluency. In fact, in several cases, the data indicate the view that self-correction sounds more native-like:

There are false starts, reformulations to reorganize ideas, and self-corrections, and they are all fine. We do the same in our first language. (Rater 2, Participant 38)

These findings are consistent with the results of Lennon (1990), who found that the ability to self-correct is an important aspect of fluency because it resembles native-speaker performance. Other studies have also demonstrated that the presence of repairs per se is not indicative of a lack of fluency (Freed 2000; Olynyk et al. 1990).

Although there exist four major types of self-correction (see the taxonomy Dörnyei and Kormos 1998), the raters did not make specific comments pertaining to them. Rather, they generally acknowledged the fact that L2 speakers who initiated repairs were more favourably perceived. Therefore the use of a number of repairs as a dysfluency marker may be problematic in terms of the operationalization of fluency. As the results suggest, a higher frequency of repairs might not be indicative that speakers are perceived to be less fluent. Rather, repairs seem to support perceptions of fluency.

#### 4.6. *Efficiency and effortlessness*

The raters' comments consistently highlight the fine balance that must be struck between speech rate, rhythm, pausing and lexical retrieval, as these are central to perceptions of efficiency and effortlessness in French. In the context of the raters' observations, the psycholinguistic processes of speech planning and encoding are particularly salient because their efficiency, or whether they operate easily and effortlessly, seem to have a direct impact on listeners' perceptions of L2 automaticity as demonstrated by the qualitative perceptions.

With regard to perceived efficiency and effortlessness, four kinds of pattern emerge from the qualitative data. First, a key notion qualifying L2 efficiency and effortlessness in French is continuity, meaning the ability to 'keep going' or to 'hold the floor', for example:

There was continuity ... and the whole story was coherent from start to finish. I wanted to keep listening. He did more than share a story, he shared an experience. (Rater 3, Participant 36)

In the above example, the rater describes the speaker's remarkable ease of linguistic expression in delivering L2 speech. As we have seen, speakers who are evaluated favourably

are able to deliver longer messages with efficiency and without effort and are successful at catching the listener's attention. These findings are reminiscent of Freed (2000) where raters noted 'ease', 'confidence in speech' and 'comfort in the ability to converse' as speech qualities representative of higher fluency ratings.

Second, efficiency and effortlessness are represented in the raters' comments in terms of overall spontaneous language processing, for example:

Her anecdote was clear and she was able to include all the pictures in her story. She didn't hesitate, nor correct herself. She was quick to synthesize the information. She was a competent speaker. She sounded spontaneous. (Rater 3, Participant 25)

From the raters' comments it appears that spontaneity in spoken discourse is the speech feature underlying perceptions of effortless language processing. The data also illustrate the opposite effect of spontaneity on speech perception, despite apparent fluency features in the learners' speech:

She's fluent, but lacks spontaneity ... sounds as if she's looking for the right words rather than just speaking. (Rater 1, Participant 30)

For these raters, the emphasis on a lack of spontaneity is perceived as disengagement on the part of the speaker. Thus different degrees of efficiency and naturalness in linguistic processing in the speaker's oral performances are detected in listeners' perceptions.

Third, the listeners also attend to naturalness as another beacon of efficiency and effortlessness, for example:

Very fluent speaker. She's clearly very comfortable speaking and using French. She speaks almost like a native. I felt she was sitting in front of me, she was not just a recording. (Rater 3, Participant 43)

In these references, naturalness is suggestive of authenticity and consistent with giving one's full and undivided attention to the person or matter at hand.

Finally, excerpts referring to the overall organization of speech were another feature the raters considered when reflecting on efficiency and effortlessness, for example:

While the student may be able to speak French, there was no story as such, just a string of elements put together in a narrative which lacked relevance. He was just filling time and his story was incoherent. (Rater 3, Participant 13)

According to our data, demonstrating spontaneity, naturalness and organization of speech, or coherence seem to be critical to perceptions of L2 efficiency and effortlessness. Interestingly, these three qualifiers often appear as descriptors on various qualitative L2 speaking-ability scales. For example, the Common European Framework (CEF) (Council of Europe 2001) includes the following descriptors of spoken fluency: "Can express him/herself

fluently and spontaneously, almost effortlessly” (CEF Table 3, pp. 28–29). The Canadian Language Benchmarks (CLB) (2012) also take a similar stance by placing emphasis on learners’ ability to “give descriptions in coherent narratives” (p. 73). The American Council on the Teaching of Foreign Languages (ACTFL) (2012) proficiency guidelines qualify a distinguished level speaker as “able to use language skillfully, and with accuracy, efficiency, and effectiveness” (p. 4).

#### 4.7 *Rhythm*

Rhythm is generally understood as the regular, patterned beat of stressed and unstressed syllables and pauses in an utterance. The three raters consistently displayed sensitivity to this specific acoustic quality of speech from different perspectives. A regular rhythm seems to have considerable impact on perceived fluency in general, in terms of catching and maintaining the listener’s attention:

This student’s speech sample really interested me, not her story, but specifically her rhythm and lack of hesitation. (Rater 2, Participant 31)

Second, from a cross-linguistic perspective, it is widely acknowledged that the rhythmic patterning in English and French differ substantially (Abercrombie 1967; Pike 1945; Vaissière 1991; Walker 2001). Contrary to the stress-timed language of English, French is often described as a syllable-timed language (Abercrombie 1967; Vaissière 1991; Walker 2001). In French, all syllables within one rhythmic group have approximately the same duration, excluding the final syllable, which gets lengthened but without an increase in loudness (Wenk and Wioland 1982). In English, the accented syllable is towards the beginning of the rhythmic group and has variable lengthening, but there is an increase in loudness. This contrast in the duration of sounds is a major contributor to the impression of musicality and perception of rhythm in French as there are no stress marks on words, but rather on the rhythmic group (Price 2005). The substantial rhythmic difference between English and French and the acoustic targetlike correctness of the speakers’ rhythm was a noticeable feature, as this rater explains:

Her rhythm is not French; she’s telling her story with an English rhythm, but using French. (Rater 3, Participant 5)

### 5. Conclusion and implications

Understanding fluency within the context of L2 speech production and perception is a critical challenge facing language assessment and second language acquisition research. The goal of this study was to demonstrate the extent to which certain factors impact on speech perception and affect raters’ qualitative assessment of fluency in L2 French. All of the accounts chosen have a common goal: they try to explain how the auditory system of native-speaker listeners perceives speech output and how they try to match it to the input stored in their internal language representation of French. The data suggest that a fine

balance must be struck between speed, pausing, lexical retrieval, self-correction, efficiency/effortlessness, and particularly rhythm, to qualify as a fluent L2 French speaker. However, as the excerpts from the qualitative comments indicate, these speech features and concepts are inherently intertwined and cannot easily be distinguished from each other.

While numerous comments refer to speed of delivery, in general it is mentioned less often than rhythm, indicating that it might be secondary to rhythm in L2 French. In principle, according to the raters' reactions, an L2 speaker is considered fluent when they can combine all these features to speak easily, relatively quickly and with pauses at appropriate junctures. While all these factors together weave an intricate pattern in the fabric of L2 fluency, the speech features that were most frequently commented on by the raters in this dataset were speech rhythm, efficiency and effortlessness. This may be due to the fact that speech rhythms and efficiency/effortlessness are easily perceived features, and given the L2 French immersion context to which the raters are accustomed, they may be more conscious of these speech characteristics in L2 learners. Nonetheless, speech rhythm appears to be an overarching characteristic of fluent L2 speech in French. This one very pertinent comment summarizes well the essence of the powerful effect of rhythm as the underlying quality that must be achieved for favourable perceptions of fluency in French:

This student borrows words from the language of Molière and sings them with the music of the English language. The melody of French spoken by a francophone is very different, not only in the melody, but also on the stress placed on the word. She speaks like an anglophone.

The conclusion here is that the language of Molière must not be spoken as if it is the language of Shakespeare. In other words, it is not enough simply to produce French utterances, learners also need to learn the rhythm of the language. Like in music, speech rhythm in French is structured and anticipatory. In this regard, teaching students what constitutes a French rhythm could facilitate both the timing of speech production and its perception. A further conclusion from the qualitative analysis might also be that rhythm in French is a continuum, and that special attention must be paid to the durational properties of syllables.

## References

- Abercrombie, D. 1967. *Elements of general phonetics*. Chicago: Aldine.
- American Council on the Teaching of Foreign Languages. 2012. ACTFL Proficiency Guidelines. Alexandria, VA.
- Bosker, H., Pinget, A., Quené, H., Sanders, T. and De Jong, N. 2013. What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing* 30 (2).159–175.
- Brumfit, C. 1984/2000. Accuracy and fluency: The basic polarity. In H. Rigganbach (ed.), *Perspectives on fluency*, 61–73. Ann Arbor: University of Michigan Press.
- Citizenship and Immigration Canada. 2012. Canadian Language Benchmarks: English as a second language for adults. Ottawa, Ontario.
- Clark, H. H. 2002. Speaking in time. *Speech Communication*, 36 (1). 5–13.
- Clark, H. H. and Fox Tree, J. 2002. Using uh and um in spontaneous speaking. *Cognition* 84 (1). 73–111.

- Council of Europe. 2001. *Common European Framework of reference of languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Cucchiaroni, C., Strik, H. and Boves, L. 2002. Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America* 111 (6). 2862–2873.
- Delattre, P. 1961. La leçon d'intonation de Simone de Beauvoir, Étude d'intonation déclarative comparée. *The French Review*, 35 (1). 59–67.
- Derwing, T., Rossiter, M., Munro, M. and Thomson, R. 2004. Second language fluency: Judgments on different tasks. *Language Learning* 54 (3). 655–679.
- Dörnyei, Z. 1995. On the teachability of communication strategies. *TESOL Quarterly* 29 (1). 55–85.
- Dörnyei, Z., and Kormos, J. 1998. Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in Second Language Acquisition* 20 (3). 349–385.
- Ejzenberg, R. 2000. The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (ed.), *Perspectives on fluency*. 287–313. Ann Arbor: University of Michigan Press.
- Fillmore, C. 2000. On fluency. In H. Riggenbach (ed.), *Perspectives on fluency*. 43–60. Ann Arbor, MI: University of Michigan Press.
- Foster, P., and Skehan, P. 1999. The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research* 3 (2). 215–247.
- Freed, B. 2000. Is fluency, like beauty, in the eyes and ears of the beholder? In H. Riggenbach (ed.), *Perspectives on fluency*. 243–265. Ann Arbor, MI: University of Michigan Press.
- Freed, B., Segalowitz, N. and Dewey, D. 2004. Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition* 26 (2). 275–301.
- Ginther, A., Dimova, S. and Yang, R. 2010. Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing* 27 (3). 379–399.
- Iwashita, N., Brown, A., McNamara, T. and O'Hagan, S. 2008. Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics* 29 (1). 24–49.
- Kormos, J. 2000. The role of attention in monitoring second language speech production. *Language Learning* 50 (2). 343–384.
- Kormos, J. 2006. *Speech production and second language acquisition*. Mahwah: Lawrence Erlbaum.
- Kormos, J. and Dénes, M. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32 (2). 145–164.
- Lam, W. 2006. Gauging the effects of ESL oral communication strategy teaching: A multi-method approach. *Electronic Journal of Foreign Language Teaching* 3 (2). 142–157.
- Lennon, P. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning* 40 (3). 387–417.
- Lennon, P. 2000. The lexical element in spoken second language fluency. In H. Riggenbach (ed.), *Perspectives on fluency*. 25–42. Ann Arbor: University of Michigan Press.
- Levelt, W. J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

- Munro, M. J., and Derwing, T. M. 2001. Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition* 23 (4). 451–468.
- Nakatani, Y. 2005. The effects of awareness-raising training on oral communication strategy use. *The Modern Language Journal* 89 (1). 76–91.
- Nakatani, Y. 2010. Identifying strategies that facilitate EFL learners' oral communication: A classroom study using multiple data collection procedures. *The Modern Language Journal* 94 (1). 116–136.
- Olynyk, M., D'Anglejan, A. and Sankoff, D. 1990. A quantitative and qualitative analysis of speech markers in the native and second language speech of bilinguals. In R. Scarcella, R. Andersen, and S. Krashen (eds.), *Developing communicative competence in a second language*. 139–155. Rowley: Newbury House.
- Pawley, A., and Syder, F. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. Richards and R. Schmidt (eds.), *Language and communication*. 191–226. New York: Longman.
- Pawley, A. and Syder, F. 2000. The One-Clause-at-a-Time Hypothesis. In H. Rigganbach (ed.), *Perspectives on fluency*. 163–199. Ann Arbor: University of Michigan Press.
- Pell, M. 2001. Influence of emotion and focus location on prosody in matched statements and questions. *The Journal of the Acoustical Society of America* 109 (4) 1668–1680.
- Pike, K. 1945. *Intonation of American English*. Ann Arbor: University of Michigan Press.
- Poullisse, N. 1993. A theoretical account of lexical communication strategies. In R. S. Weltens and B (eds.), *The bilingual lexicon*. 157–189. Amsterdam: John Benjamins.
- Price, G. 2005. *An introduction to French pronunciation*. Malden: Blackwell.
- Préfontaine, Y. 2013. Perceptions of French fluency in second language speech production. *Canadian Modern Language Review* 69 (3). 324–348.
- Préfontaine, Y. and Kormos, J. 2015. The relationship between task difficulty and second language fluency in French: A mixed methods approach. *The Modern Language Journal* 99 (1). 96–112.
- Préfontaine, Y., Kormos, J. and Johnson, D. 2015. How do utterance measures predict raters' perceptions of fluency in French as a second language? *Language Testing*, 1–21. doi:10.1177/0265532215579530
- Raupach, M. 1980. Temporal variables in first and second language speech production. In H. W. Dechert and M. Raupach (eds.), *Temporal variables in speech*. 263–270. Mouton: The Hague.
- Raupach, M. 1987. Procedural learning in advanced learners of a foreign language. In J. A. Coleman and R. Towell (eds.), *The advanced language learner*. 123–155. London: CILT.
- Rigganbach, H. 1991. Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes* 14 (4). 423–441.
- Rubin, H. J., and Rubin, I. 2005. *Qualitative interviewing: The art of hearing data* 2nd ed.. Thousand Oaks: Sage.
- Sajavaara, K. 1987. Second language speech production: Factors affecting fluency. In H. Dechert and M. Raupach (eds.), *Psycholinguistic models of production*. 45–65. Norwood: Ablex.
- Schmidt, R. 1992. Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition* 14 (3). 357–385.

- Schwab, S. and Grosjean, F. 2004. La perception du débit en langue seconde. *Phonetica*, 61 (1). 84–94.
- Segalowitz, N. 2010. *Cognitive bases of second language fluency*. New York: Routledge.
- Tavakoli, P. and Skehan, P. 2005. Strategic planning, task structure and performance testing. In R. Ellis (ed.), *Planning and task performance in a L2*. 239–277. Amsterdam: John Benjamins.
- Towell, R. 2002. Relative degrees of fluency: A comparative case study of advanced learners of French. *International Review of Applied Linguistics in Language Teaching* 40 (1). 117–150.
- Towell, R., Hawkins, R., and Bazergui, N. 1996. The development of fluency in advanced learners of French. *Applied Linguistics* 17(1), 84–119.
- Vaissière, J. 1991. Rhythm, accentuation and final lengthening in French. In J. Sundberg, L. Nord, and R. Carlson (eds.), *Music, language, speech and brain*. 108–120. Stockholm: Macmillan Press.
- Walker, D. 2001. *French sound structure*. Calgary: University of Calgary Press.
- Wenk, B. and Wioland, F. 1982. Is French really syllable–timed? *Journal of Phonetics*, 10 (2). 193–216.
- Wennerstrom, A. 2000. The role of intonation in second language fluency. In H. Riggenbach Ed., *Perspectives on fluency*. 102–127. Ann Arbor: University of Michigan Press.
- Wennerstrom, A. 2001. *The music of everyday speech: Prosody and discourse analysis*. Oxford: Oxford University Press.

Clare Wright and Parvaneh Tavakoli  
University of Reading

Introduction to special issue: New directions and developments in defining, analyzing and measuring L2 speech fluency

## 2. Segalowitz

Second language fluency and its underlying cognitive and social determinants

Norman Segalowitz  
Concordia University (Montreal)  
&  
Queensland University of Technology (Brisbane)

Contact: Norman Segalowitz,  
Psychology Department, Concordia University,  
Montreal, Quebec Canada H4B 1R6  
<norman.segalowitz@concordia.ca>  
+1 514.848.2424 ext. 2239

## 3. Ladders and Snakes in Second Language Fluency

Peter Skehan<sup>8</sup>, Pauline Foster<sup>2</sup>, and Sabrina Shum<sup>9</sup>  
Peter Skehan<sup>10</sup>, Pauline Foster<sup>2</sup>, and Sabrina Shum<sup>11</sup>

3. Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency

Nivja H. de Jong  
Utrecht University

Address: Trans 10, 3512 JK Utrecht, the Netherlands  
Email: n.dejong@uu.nl

---

<sup>8</sup> St. Mary's University, Twickenham

<sup>9</sup> Chinese University of Hong Kong

<sup>10</sup> St. Mary's University, Twickenham

<sup>11</sup> Chinese University of Hong Kong

5 Yvonne Prefontaine and Judit Kormos

A qualitative analysis of perceptions of fluency in seco  
IRAL Special Issue in L2 Fluency

| <b>Contributors and affiliations</b>   | <b>Title of the paper</b>   |
|--|---|
| Wright and Tavakoli<br>University of Reading   | Introduction to special issue: New directions and developments in defining, analyzing and measuring L2 speech fluency |
| Norman Segalowitz, Concordia University and Queensland University of Technology                                | Second language fluency and its underlying cognitive and social determinants  |
| Foster, Pauline, Peter Skehan, St Mary's University College, London, & Sabrina Shum, Hong Kong City University | Influences on second language fluency: Insights from first language performance                                       |
| de Jong, Nivja, Utrecht University   | Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency                         |
| Tavakoli, Parvaneh, University of Reading  | Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency                   |
| Prefontaine, Yvonne, University of Illinois at Urbana Champaign, and Judit Kormos, Lancaster University        | A qualitative analysis of perceptions of fluency in second language French  |

nd language French