Bilingualism: Language and Cognition

cambridge.org/bil

L2 exposure modulates the scope of planning during first and second language production

Annie C. Gilbert[1,2,3], Maxime Cousineau-Perusse[1] and Debra Titone[1,3]

[1]Department of Psychology, McGill University, Montréal, Canada; [2]School of Communication Sciences and Disorders, McGill University, Montréal, Canada and [3]Center for Research on Brain, Language, and Music, Montréal, Canada

Research Article

Cite this article: Gilbert AC, Cousineau-Perusse M, Titone D (2020). L2 exposure modulates the scope of planning during first and second language production. *Bilingualism: Language and Cognition* **23**, 1093–1105. https://doi.org/10.1017/S1366728920000115

Received: 22 September 2017
Revised: 7 January 2020
Accepted: 17 January 2020
First published online: 27 February 2020

Keywords:
scope of planning; speech production; bilingualism; second language exposure; individual differences; bilingual experience

Address for correspondence:
Annie C. Gilbert, E-mail: annie.c.gilbert@mail.mcgill.ca

Abstract

The psycholinguistic literature suggests that the length of a to-be-spoken phrase impacts the scope of speech planning, as reflected by different patterns of speech onset latencies. However, it is unclear whether such findings extend to first and second language (L1, L2) speech planning. Here, the same bilingual adults produced multi-phrase numerical equations (i.e., with natural break points) and single-phrase numbers (without natural break points) in their L1 and L2. For single-phrase utterances, both L1 and L2 were affected by L2 exposure. For multi-phrase utterances, L1 scope of planning was similar to what has been previously reported for monolinguals; however, L2 scope of planning exhibited variable patterns as a function of individual differences in L2 exposure. Thus, the scope of planning among bilinguals varies as a function of the complexity of their utterances: specifically, by whether people are speaking in their L1 or L2, and bilingual language experience.

Introduction

Our ability to speak involves multiple cognitive processes that rapidly unfold in a cascaded manner. These include planning the meaning of the message we wish to communicate (i.e., semantic information), accessing and selecting the specific words we will use (i.e., lexical forms), the grammatical constructions involved (i.e., syntactic information), and the pronunciation or articulation of what we will produce (i.e., phonological information; e.g., Levelt, 1989; Pivneva, Palmer & Titone, 2012). Here, we investigate how far in advance bilingual speakers plan these components when given a stimulus to be named: that is, their SCOPE OF PLANNING. In some cases, the scope of planning could be long, such as if we plan all components of an entire utterance before starting to speak (reflected by relatively longer speech onset times). In other cases, the scope of planning could be short, including only a few components before starting to speak. For example, if we plan only the initial portion of an utterance before beginning to speak, with the intention to plan the rest as we go (reflected by relatively shorter speech onset times).

Past work on monolingual speech production suggests that the scope of planning varies for different kinds of linguistic information (e.g., semantics vs. phonology) as well as the complexity and length of what we intend to produce. For example, speakers usually begin with the meaning of what they want to say: that is, semantics. Accordingly, semantic planning happens earlier than other linguistic elements (i.e., phonology) and, moreover, semantic representations remain activated throughout production (Meyer, 1996; Meyer, Sleiderink & Levelt, 1998; Meyer & van der Meulen, 2000; Michel Lange, Cheneval, Python & Laganaro, 2017; Oppermann, Jescheniak & Schriefers, 2010; Smith & Wheeldon, 1999, 2004; Yang & Yang, 2008). With respect to complex multi-word utterances, Smith and Wheeldon (1999) demonstrated that the first part of an utterance is planned in greater semantic detail than the rest of the utterance (i.e., its initial phrase, word, or CHUNK; see Finch, 2000 for details). Smith and Wheeldon (2004) further demonstrated that semantic planning can even occur immediately for the entire utterance, suggesting that the scope of planning can be quite long for semantic information.

Phonological information, the focus of this paper, is also planned during speech production. This is reflected by the impact of phonological length or complexity of a to-be-produced item on speech onset latencies. For example, Meyer, Belke, Häcker and Mortensen (2007) found that speech onset latency during single word production tracked word length – longer words were associated with longer speech onset latencies. However, subsequent findings for single word production did not support these findings. For instance, Damian, Bowers, Stadthagen-Gonzalez and Spalek (2010) found no effect of word length on speech onset latency in a single-word speech production task. Such inconsistencies may have arisen because of uncontrolled confounds with word length, such as familiarity or frequency (Alario, Ferrand, Laganaro, New, Frauenfelder & Segui, 2004).

© Cambridge University Press 2020

Moreover, single word production alone may be problematic for comprehensively evaluating the scope of phonological planning, as there is much less to plan and produce compared to more natural speech production situations involving series of words. Consistent with this view, production for multi-phrase utterances has shown robust length effects. For example, Smith and Wheeldon (1999) compared speech onset latencies for two-phrase utterances that had an initial simple phrase and a later complex phrase, or vice versa, while controlling for overall utterance complexity (e.g., complex-simple sentence: *The dog and the foot move above the kite*, vs. simple-complex sentence: *the dog moves above the foot and the kite*). Here, speech onset latencies tracked the phonological length of the first phrase, with simple-complex sentences yielding shorter speech onset latencies than complex-simple sentences, even for sentences consisting of the same words. Thus, the structure of an utterance has an impact on speech planning over and above its semantic content or the specific words to be produced (see also Martin, Crowther, Knight, Tamborello Ii and Yang (2010) for consistent results from a similar task.)

Other work suggests that the scope of phonological speech planning is not fixed but rather varies adaptively as a function of global contextual demands. For example, Griffin (2003) found longer speech onset latencies when the first word of a two-word pair was short (one syllable) vs. longer (two to five syllables). While these results superficially contradict Smith and Wheeldon (1999), the short items in Griffin's task were much shorter than the "simple" phrases of previous studies, which may have differentially impacted the scope of planning. Interestingly, this effect disappeared when transition words, such as *next to,* were inserted between the two words of the pair, a manipulation that made the first phrase longer and gave speakers more time to plan the second word following speech onset (Griffin, 2003). Eye movement data acquired simultaneously supported this conclusion by showing that participants spent more time looking at the second picture prior to speaking when the initial item was short but not long. However, once they started to speak, participants spent less time looking at the second item during the production of the first short items, presumably because they had already started planning the second word before initiating speech.

Thus, semantic speech planning is initiated earlier and extends later in time than phonological planning. It is also dynamically affected by the complexity of specific to-be-produced items. Phonological planning, in contrast, occurs on a rolling basis, involving the planning of smaller speech chunks, the size of which can be adapted to fit current communicative demands. However, this general account derives almost entirely from experiments involving people who speak only one language (i.e., monolinguals; Allum & Wheeldon, 2007; Brown-Schmidt & Konopka, 2015; Costa & Caramazza, 2002; Damian & Dumay, 2007; Ferreira, 1991; Ferreira & Swets, 2002; Holmes, 1988; Lindsley, 1975; Martin et al., 2010; Martin, Miller & Vu, 2004; Meyer, 1996; Meyer et al., 2007; Meyer et al., 1998; Meyer & van der Meulen, 2000; Oppermann et al., 2010; Smith & Wheeldon, 1999, 2004; Swets, Jacovina & Gerrig, 2014; Wagner, Jescheniak & Schriefers, 2010; Wheeldon & Lahiri, 1997, 2002; Zhao, Alario & Yang, 2015). Thus, a gap in the literature is whether conclusions about speech planning would apply to people who speak more than one language (i.e., bilinguals).

To fill this gap, it is crucial to first consider the ways in which people producing L1 or L2 speech may differ and how these differences might modulate the scope of planning. One important difference is that bilinguals typically have greater current L1 vs. L2 exposure (along with its associated impact on overall L2 proficiency, Chakraborty & Goffman, 2011; Thordardottir, 2011; Thordardottir & Brandeker, 2013; Vermeer, 2001). This, in turn, can influence speech planning strategies by either shortening or lengthening the scope of planning (as reflected by speech onset latency), or making it more or less flexible depending on the complexity of what is to be produced. Because speakers would likely have an easier time producing L1 vs. L2 speech for a variety of reasons (e.g., increased ease of lexical-semantic and phonological activation of words), they may be more likely to begin speaking before having planned an entire utterance in their L1 than their L2, indicating a shorter scope of planning in their L1 vs. L2. Also relevant are individual differences in variables such as the amount of current L2 exposure or the historical consistency of their L2 exposure (i.e., constant or variable L2 exposure levels throughout the lifespan).

A second difference between L1 and L2 production is the potential for cross-language activation and interference, which is more likely for L2 vs. L1 speech planning as a function of L2 exposure (De Groot, 2011; Jacobs, Fricke & Kroll, 2016; Kroll & Bialystok, 2013). Cross-language activation in natural production might manifest in terms of accented L2 pronunciations, word choice, or syntactic formulations. Further, the L2 could also influence the L1, particularly if L2 fluency is high (De Groot, 2011). Thus, when producing two or more languages, managing cross-language activation and selecting language-appropriate conceptual, semantic, lexical, grammatical, and phonological representations are an ongoing demand that may also vary with factors such as L2 exposure (e.g., Pivneva et al., 2012). For these reasons, cross-language interference could also affect the scope of L1 vs. L2 speech planning in a variety of ways that include a general slowing of the different components of speech production leading to longer speech onset latencies in general (e.g., Hermans, Bongaerts, De Bot & Schreuder, 1998; Hermans, Ormel, van Besselaar & van Hell, 2011).

Finally, a third difference between L1 and L2 production may be bilinguals' subjective confidence when communicating in a particular language (see Clément, Gardner & Smythe, 1977, 1980 for studies on francophones learning English in Montreal). For example, to the extent that a bilingual individual is more confident retrieving words and communicating in their L1 vs. their L2, their scope of planning might be shorter during L1 production but longer during L2 production. Consequently, speakers might err on the side of caution by planning their entire utterance at once prior to speaking, yielding very long speech onset latencies, or try to maximize their speech fluency by starting to speak before having planned an entire utterance, yielding shorter speech onset latencies. This difference between L1 and L2 speech planning might also vary as a function of individual differences in factors such as L2 exposure. Accordingly, bilinguals with high vs. low L2 exposure might be more confident producing speech in their L2 (because of reduced L1 on L2 interference effects) but be slightly less confident in their L1 (because of increased L2 on L1 interference effects).

The present study

Here, we investigated the scope of planning in bilingual adults when they produced utterances in both their L1 and L2. Given the work cited above, we had three main expectations: L2 and

https://doi.org/10.1017/S1366728920000115 Published online by Cambridge University Press

L1 speech planning would differ, speakers' scope of planning would vary adaptively alongside individual differences in L2 experience (especially for the L2, although also for L1 when L2 experience is high, De Groot, 2011), and speakers would strategically differ in L1 and L2 (i.e., they would either be cautious, as indicated by a positive relationship between phrase length and onset latencies, or adaptive, as indicated by an inverse relationship between phrase length and onset latencies).

To test these expectations, we used a production task where the same bilingual adults produced both multi-phrase utterances (numerical equations) and single-phrase utterances (isolated numbers) of comparable lengths, in both their L1 and L2. These two types of stimuli allowed the observation of phrase length effects on the scope of planning over and above overall effects of stimuli length. We chose numerical stimuli instead of words and sentences because the former were easier to control cross-linguistically compared to the latter in terms of frequency and semantic complexity (see Gollan, Montoya, Cera & Sandoval, 2008; and Gollan, Slattery, Goldenberg, Van Assche, Duyck & Rayner, 2011, for a description of factors that influence word frequency effects in L2.). We also carefully selected participants because of our use of numerical materials. Specifically, previous research on number processing and mathematical stimuli has shown that numbers are not processed the same way in the L1 and L2 (Prior, Katz, Mahajna & Ribinsten, 2015; Salillas & Wicha, 2012; Spelke & Tsivkin, 2001; Van Rinsveld, Brunner, Landerl, Schiltz & Ugen, 2015; Van Rinsveld, Dricot, Guillaume, Rossion & Schiltz, 2017; Van Rinsveld, Schiltz, Brunner, Landerl & Ugen, 2016). Thus, all participants were selected to have the same L1 as well as a similar linguistic background during childhood, to ensure their knowledge of numbers in their L1 was comparable and entrenched.

Following past work reviewed above (e.g., Griffin, 2003, Smith and Wheeldon, 1999, Martin et al., 2010), we assessed the scope of planning of multi-phrase utterances by evaluating whether speech onset latencies tracked the length of the first or second phrase to be produced. If speech onset latencies were systematically LONGER as a function of increased length of the first but not the second phrase, we would conclude that the scope of planning was consistently short (i.e., speakers planned only the initial phrase of the utterance before initiating speech). Conversely, if speech onset latencies were systematically LONGER as a function of increased length of the second phrase (irrespective of the first phrase), we would conclude that the scope of planning was consistently long, including the entire utterance. Other systematic patterns would be interpreted as a sign that speakers adaptively modulated their scope of planning to fit the specific demands of the utterance (for example, if speech onset latencies were systematically SHORTER as a function of increased length of either the first or second item). Such pattern would indicate that speakers were not restricted to one scope of planning (short or long) for all utterances, but instead used the optimal scope of planning given the to-be-produced utterance. Based on single word production experiment (Alario et al., 2004; Meyer et al., 2007), the scope of planning of single phrase utterances is expected to include the entire word, with longer words leading to longer speech onset latencies.

Given this logic, our predictions for L1 speech were as follows. To the extent that speakers were proficient and confident using their L1, their scope of planning for multi-phrase utterances should mirror what has been found previously for monolingual speakers producing multi-word utterances. That is, speech onset latency should vary with the length of the first phrase to be produced, indicating that speakers use a smaller scope of planning, planning only the first phrase before initiating speech, and trusting that they can incrementally plan the second phrase after they begin to speak. This pattern should hold unless the first phrase is extremely short (i.e., one syllable as in Griffin, 2003), in which case speakers would modulate their scope of planning to also include the second phrase, thus leading to overall longer speech onset latencies (Griffin, 2003). In contrast, L1 speech onset latencies for single phrase utterances should consistently increase as length of the one and only phrase to be produced increased, perhaps as a function of L2 exposure. Such pattern would indicate that the scope of planning cannot be adaptively modulated if the utterance to produce consists of only one phrase (i.e., does not present a clear breaking point).

However, our predictions for L2 speech were less clear. Given the ways that L2 speech production might differ from L1 speech (i.e., being less fluent, creating more cross-language competition, reducing speaker confidence), there are at least two possible predictions that may be distinguished. If scope of planning is primarily affected by L2 proficiency/fluency, then speakers will likely have a reduced scope of planning for the L2 relative to the L1. As a consequence, only the length of the first phrase should have an impact on speech onset latencies. Alternatively, to the extent that reduced confidence in speaking in an L2 has an impact on the scope of planning, L2 speakers might be more cautious or risk-averse when producing L2 speech, and thus refrain from speaking until they have planned out the entire utterance. As a consequence, length of the second phrase should have an impact on speech onset latencies. Finally, to the extent that L2 speakers were highly fluent and comfortable in their L2 because of high L2 exposure, they may exhibit a more adaptive pattern, similar to L1 speech. In contrast, L2 speakers' scope of planning for single phrase utterances should consistently increase as length of the one and only phrase to be produced increased, likely again as a function of L2 exposure.

Method

Participants

Eighteen bilingual adults were recruited through the McGill University psychology subject pool, and from the Université de Montréal, ranging in age from 19 to 33 years (mean = 23, SD = 4.21). Data from one participant were removed due to error rates greater than 3.5 standard deviations from the group mean. Remaining participants were either L1 French speakers (n = 13), or simultaneous bilingual speakers who learned French from birth and started acquiring English as an L2 before the age of 3; n = 4). All participants self-reported French as their L1 and their usual language of communication. Participants were also screened for any perceptual, speech, and learning impairments.

Information regarding language use, preferences, and proficiency were collected using a modified version of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumstein & Kaushanskaya, 2007), which provided estimates of participants' L2 proficiency, language dominance, and L2 exposure (i.e., proportion of time spent in L2 on a daily basis; see Table 1). Participants were highly proficient in English (L2) as indicated by their self-reported SPEAKING, FLUENCY, and OVERALL L2 proficiency. An L1 and L2 semantic judgement task was also

https://doi.org/10.1017/S1366728920000115 Published online by Cambridge University Press

Table 1. Self-reported and objective measures of language preference and proficiency of participants.

	M	*SD*	*Min*	*Max*
Preference to speak L1 (% time choose to speak L1 over L2) *	69.41	25.67	25[+]	100
Relative word retrieval efficiency (1–5) **	1.94	1.09	1	5
Current French (L1) exposure (% time spent)	55.50	17.63	30	90
Current English (L2) exposure (% time spent)	42.50	17.51	10	70
Self-reported L2 proficiency ratings (1–7) ***				
Speaking	5.50	0.89	4	7
Fluency	5.18	1.42	2	7
Overall	5.53	0.87	4	7
Semantic judgement task				
French				
Accuracy (proportion correct)	0.95	0.02	0.90	0.98
RT (ms)	645	117	519	1012
English				
Accuracy (proportion correct)	0.95	0.03	0.89	1
RT (ms)	631	135	500	1080
English / French ratio				
Accuracy	1	0.03	0.95	1.04
RT	0.97	0.10	0.82	1.13

**"When choosing a language to speak with a person who is equally fluent in all your languages, what percentage of time would you choose to speak French?"*
***"How easy is it for you to find the words you want to use, when speaking normally, in French compared to English.* (Scale of 1 to 5, where 1 = easier in French, and 5 = easier in English)"
****"Please rate your linguistic ability in English according to a 1 to 7 scale, where 1 = limited, and 7 = native-like."*
[+]The participant who reported the lowest *Preference for speaking in their L1* (25%) also reported being exposed to his L2 only 30% of the time. His low score probably indicates a desire to practice his L2 when given the opportunity more than his language preference *per se*. The second lowest score was 30%.

used to obtain a more objective measure of language dominance. During this task, participants decided as quickly and accurately as possible whether a series of nouns had living (cat) or non-living (car) referents. Language dominance was estimated by comparing the accuracy and response time (RT) to their L1 and their L2, (i.e., the English-L2 score divided by the French-L1 score, where scores > 1 indicate English dominance and scores < 1 indicate French dominance). Paired-sample t-tests comparing accuracy and RT scores across languages yielded no significant difference between L1 and L2 scores [accuracy: $t = 0.094$, $MSE = 0.006$, $p = 0.926$; RT: $t = 0.843$, $MSE = 16.732$, $p = 0.412$], further suggesting that our participants were equally proficient in their L2 and L1. Information regarding language of schooling was acquired for each level separately (elementary, high school, Cegep, university) to ensure that participants initially learned to process numbers in their L1 and to estimate the stability of their L2 exposure level.

Despite their overall high L2 proficiency, 15 out of 17 participants showed a clear preference for their L1, suggesting lower confidence in their L2. Specifically, these speakers indicated that they prefer to speak their L1 when given the choice between their L1 or L2, and subjectively find word retrieval to be easier in French (L1) than in English (L2). Hence, regardless of their high objective L2 proficiency, these participants behaved and saw themselves as French-L1 dominant speakers. The other two participants showed signs of reverse preference, reporting it easier to retrieve word in English-L2 than in French-L1. These participants were kept in the sample based on the fact that they attended primary school in French, meaning that they likely learned to process numbers in French and still report French as their usual language of communication.

To further examine how the different language variables related to each other and to L2 exposure, we computed Spearman Rho correlations (Table 2). Many of the self-reported L2 proficiency ratings (*Speaking*, *Fluency*, and *Overall*), and language dominance scores (*Preference to speak L1, Relative word retrieval efficiency, and Semantic judgement ratio*) intercorrelated, and the *Overall* self-reported proficiency ratings correlated with *Current L2 exposure*. Given this later correlation, we decided not to include proficiency estimates as predictors in statistical models including L2 exposure estimates to avoid model collinearities (Baayen, 2008; Tabachnick & Fidell, 2007).

Stimuli

The to-be-produced stimuli consisted of isolated numbers and simple mathematical equations. Isolated numbers were carefully selected in terms of the number of syllables to be produced (from one to eight syllables) in French or English. Five monosyllabic digits were selected for each language and were combined to create longer numbers that systematically varied in terms of the overall number of syllables to be produced. For example, from the English monosyllabic digit "4", we derived the bisyllabic "14", the trisyllabic "24", and so forth. Five to-be-produced numbers were thus created for each length, in both languages (see Table 3; with the exception of five- and six-syllable long numbers in French that were lengthened in two different ways).

https://doi.org/10.1017/S1366728920000115 Published online by Cambridge University Press

Table 2. Correlations between self-reported and objective measures of language dominance and L2 proficiency of participants.

				Self-reported L2 proficiency ratings		
	Current English (L2) exposure (% time spent)	Preference to speak L1	Relative word retrieval efficiency	Speaking	Fluency	Overall
Preference to speak L1	−0.113					
Relative word retrieval efficiency	0.419	−0.584*				
Self-reported L2 proficiency						
Speaking	0.305	−0.35	0.429			
Fluency	0.355	−0.43	0.62*	0.858**		
Overall	0.633**	−0.325	0.538*	0.754**	0.745**	
Semantic judgement task ratio						
Accuracy	−0.132	−0.056	−0.264	0.296	0.078	0.251
RT	−0.003	0.735**	−0.571*	−0.336	−0.529*	−0.452

*p < .05, **p < .01

Table 3. Isolated numbers used for French (L1) and English (L2) trials as a function of their length (in syllables).

	Numbers used in	
Number length	French (L1) trials	English (L2) trials
1 syll	*2, 4, 5, 7, 9.*	*1, 2, 4, 5, 9.*
2 syll	*17, 19, 24, 32, 35.*	*14, 15, 19, 20, 30.*
3 syll	*42, 45, 54, 67, 69*	*24, 32, 35, 51, 59.*
4 syll*	*77, 79, 82, 84, 85.*	*102, 105, 201, 204, 209.*
5 syll	*242, 245, 254, 267, 269, 2042, 2045, 2049, 2054, 2057.*	
6 syll*	*277, 279, 282, 284, 285, 2077, 2079, 2082, 2084, 2085.*	
7 syll	*2242, 2245, 2254, 2267, 2269.*	
8 syll*	*2277, 2279, 2282, 2284, 2285.*	

*Number length reflects the classic vigesimal way to refer to these numbers in French (77 = *soixante-dix-sept, 85 = quatre-vingt-cinq*…) and not the regularized decimal version used in some francophone communities (77 = *septante-sept*, 85 = *octante-cinq*…)

Each number was presented twice for a total of ten trials of each number length, except for the shortest number length of French (L1) that were presented three times each, for a total of 15 trials. Longer numbers (5 or more syllables) were removed from the English (L2) task when we had determined after testing 6 participants that it increased participant fatigue during testing. All data were analyzed without these longer trials. In total, the reported analyses relied on 85 French (L1) trials and 40 English (L2) trials per participant. French trials were presented among 98 filler trials belonging to a different experiment (no fillers in the English blocks).

Equations consisted of pairings of isolated numbers of different lengths (from one to four syllables) separated by a mathematical operator (either addition or multiplication). In total, five trials were created per phrase length pairing (as presented in Table 4). Each participant saw half the equations with a "+" and half with a "x" (counterbalanced across participants). It is unclear if the operator ("+" or "x") will be processed as part of the first or second phrase, or completely independently. However, we made certain that it was consistently one syllable long (using *plus* and *times*, and not *minus* or *divided by*), and would not create any confound across materials. Here and throughout, we use the term "phrase" in a broad sense to refer to speech planning units regardless of their linguistic content or structure. Longer phrases (4 syllables) were removed from the English (L2) task after the first 6 participants because we observed participant fatigue, and all data were analyzed without these longer trials. In total, the reported analyses relied on 80 French (L1) trials and 45 English (L2) trials per participant.

Procedure

French and English trials were recorded at different times, during the same experimental session. The session included the two tasks presented in this paper (isolated numbers and equations), as well as two other tasks (not reported here). Task order was fixed across languages (first the isolated numbers, then the short equations, followed by two unrelated tasks). Participants performed all

https://doi.org/10.1017/S1366728920000115 Published online by Cambridge University Press

Table 4. Number pairings with regards to first and second number length in both French (L1) and English (L2). Equations are presented with a "+" to simplify the display.

		Second number length						
		1 syll		2 syll		3 syll		4 syll
		French (L1)	English (L2)	French (L1)	English (L2)	French (L1)	English (L2)	French (L1)
First number length	1 syll	2 + 7	1 + 2	2 + 19	1 + 30	2 + 45	1 + 24	2 + 85
		4 + 5	2 + 4	4 + 17	2 + 15	4 + 67	2 + 59	4 + 79
		5 + 2	4 + 9	5 + 32	4 + 19	5 + 69	4 + 32	5 + 82
		7 + 9	5 + 1	7 + 24	5 + 20	7 + 42	5 + 51	7 + 84
		9 + 4	9 + 5	9 + 35	9 + 14	9 + 54	9 + 35	9 + 77
	2 syll	17 + 4	14 + 9	17 + 35	14 + 19	17 + 54	14 + 35	17 + 82
		19 + 2	15 + 2	19 + 32	15 + 20	19 + 42	15 + 32	19 + 85
		24 + 7	19 + 4	24 + 17	19 + 14	24 + 45	19 + 24	24 + 77
		32 + 5	20 + 5	32 + 19	20 + 30	32 + 69	20 + 51	32 + 84
		35 + 9	30 + 1	35 + 24	30 + 15	35 + 67	30 + 59	35 + 79
	3 syll	42 + 7	24 + 1	42 + 19	24 + 19	42 + 45	24 + 35	42 + 77
		45 + 2	32 + 4	45 + 24	32 + 15	45 + 67	32 + 51	45 + 79
		54 + 9	35 + 9	54 + 17	35 + 14	54 + 69	35 + 59	54 + 82
		67 + 4	51 + 5	67 + 35	51 + 20	67 + 42	51 + 24	67 + 85
		69 + 5	59 + 2	69 + 32	59 + 30	69 + 54	59 + 32	69 + 84
	4 syll	77 + 9		77 + 19		77 + 42		77 + 84
		79 + 5		79 + 35		79 + 45		79 + 82
		82 + 4		82 + 17		82 + 54		82 + 77
		84 + 7		84 + 35		84 + 69		84 + 85
		85 + 2		85 + 24		85 + 67		85 + 79

tasks in French (their L1) first, and then in English (their L2). To avoid potential cross-language switch costs caused by changing the language of production, French and English sessions were separated by five short cognitive tasks and a health and language history questionnaire. The entire testing session lasted about two hours.

These experiments took place at McGill (an English-speaking University): thus, precautions were taken to ensure that participants were in a "French-L1" mode during data collection for the French trials. Upon arrival to the lab, participants were greeted in their L1 (French) by a native French speaker (experimenter), and all subsequent interactions were kept in French as much as possible. The language of subject-researcher interactions was switched to the participant's L2 (English) only after the questionnaire and executive control tasks were completed.

Prior to initiating the present task, participants were instructed to use the monosyllabic names of mathematical operations "*plus*" and "*times*" instead of the longer "*added to*" and "*multiplied by*" (French: "*fois*", and not "*multiplié par*"). They engaged in a practice session, which they could repeat as needed. Different stimuli were used in the practice session to avoid affecting the main task. When the experimenter determined that the participant understood all instructions, the main task began. Each trial began with a 50-millisecond fixation cross, followed by a stimulus (isolated number or short equation), both positioned in the center of a screen. Speech onset latencies were recorded using the voice-key of an SRBOX response box (*Psychology Software Tools, Inc.*), running under E-prime. Participants' verbal responses were also recorded using a digital recorder for further verification of potential voice key malfunctions and naming accuracy. The display visual remained onscreen even after the voice-key was triggered, and the next trial was initiated by a participant mouse click.

Statistical analysis

Speech onset latencies were analysed using a series of linear mixed effect (LME) models, testing each language separately, in two steps. The first step tested the overall impact of independent variables (phrase length) on the dependent variable (speech onset latencies) above and beyond individual differences. These comprised our core models, which allowed us to draw general conclusions about L1 vs. L2 production that generalized over all participants. The core models evaluated the fixed effect of phrase length in number of syllables (for isolated numbers, total phrase length; for equations, length of first and second phrase in equations, as well as their interaction – as scaled continuous variables). The random structure used in the core models took into account the trial produced (i.e., items, intercept only) and participants (intercept and slope adjustments for phrase length in isolated

https://doi.org/10.1017/S1366728920000115 Published online by Cambridge University Press

Table 5. Number of trials removed from the analysis as a function of exclusion criteria, type of stimuli and language of production.

	Single-phrase utterances (isolated numbers)		Multi-phrase utterances (equations)	
Exclusion criteria	French-L1	English-L2	French-L1	English-L2
Errors	24	4	34	30
Outlier speech onset values*	5	4	6	1
Voice key malfunction	27	9	7	6
Other technical malfunction	3	0	0	7

*Speech onset values < 150 ms or > 2000 ms.

number trials or the interaction between first and second phrase length in equations trials; e.g., [Speech onset latency ~ scale (Length of 1st phrase)* scale(Length of 2nd phrase) + (1 + Length of 1st phrase* Length of 2nd phrase | Participant) + (1 | Equation-item)]). When models did not converge, we followed the procedure outlined in Barr, Levy, Scheepers and Tily (2013), whereby factors were removed from random slope adjustments for participants until the model converged.

The second step evaluated how individual differences in the bilingual experience modulated the effects of the independent variables observed in the core models. These models tested the effect of phrase length as well as individual differences variables associated with current L2 exposure (as a scaled continuous variable), and interactions between these factors. The random structure used in these models took into account the trial produced (i.e., items, intercept only) and participants (intercept only; e.g., [Speech onset latency ~ scale(Length of 1st phrase)* scale (Length of 2nd phrase)* scale(proportion time spent in L2) + (1 | Participant) + (1 | Equation-item)]).

Finally, we examined L1 and L2 patterns of results for stimuli of comparable length. These models tested the main effect of language of production (as a two-level deviation-coded categorical variable: "French" vs. "English") on speech onset latencies above and beyond the effects of stimuli characteristics and individual differences previously observed. The random structure used in these models took into account the trial produced (i.e., items, intercept only) and participants (intercept only; e.g., [Speech onset latency ~ scale(Length of 1st phrase) * scale(Length of 2nd phrase) * scale(proportion time spent in L2) + Language of production + (1 | Participant) + (1 | Equation-item)]). All LME models were implemented in RStudio version 3.2.4 (R Development Core Team, 2010), using the lme4 library, version 1.1-7 (Bates, Maechler, Bolker & Walker, 2014) and *p*-value estimates were obtained using the lmerTest package version 2.0-29 (Kuznetsova, Brockhoff & Christensen, 2015). We used the effects package version 4.0-0 to extract partial effects of interest (Fox, 2003), and partial effects plots were generated using ggplot2 version 2.1.0 (Wickham, 2009). All figures represent fitted data.

Results

Data collection for 1 participant was interrupted prior to the English testing block. Data for the remaining participants (17 in French trials, 16 in English trials) contained very few errors, and only a small number of trials were removed from the analyses due to outlier values, voice key malfunctions or other malfunctions (see Table 5 for details). After data trimming, analyses were performed on 2009 isolated number trials (1386 French-L1 trials out of 1445, and 623 English-L2 trials out of 640 trials) and 1989 equation trials (1313 French-L1 trials out of 1360, and 676 English-L2 trials out of 720 trials). Table 6 presents mean speech onset latencies for isolated number trials as a function of number length for both languages while Table 7 presents mean speech onset latencies for equation trials as a function of first and second phrase length for both languages. The following section presents the results of statistical analyses performed on trials from each language separately.

Table 6. Mean speech onset latency and standard error (in ms) as a function of isolated number length in French (L1) and English (L2) trials.

	French trials		English trials		**Conditions overall**	
Number length	*M*	*SE*	*M*	*SE*	*M*	*SE*
1 syll	569	*19*	574	*18*	**572**	***13***
2 syll	599	*18*	626	*22*	**612**	***14***
3 syll	603	*22*	648	*28*	**625**	***18***
4 syll	641	*26*	746	*39*	**692**	***25***
5 syll	688	*32*				
6 syll	710	*28*				
7 syll	684	*33*				
8 syll	712	*37*				
Conditions overall	**651**	***10***	**648**	***16***		

Speech onset latencies to French-L1 trials

Single-phrase utterances (isolated numbers)

The core LME model for French-L1 isolated number trials yielded a significant main effect of number length (in syllables) [$b = 62.98$, $SE = 12.52$, $t = 5.032$, $p < 0.0001$], with longer numbers patterning with longer planning latencies. The individual differences model revealed an interaction between number length and current L2 exposure [$b = 13{,}716$, $SE = 3.925$, $t = 3.495$, $p < 0.0005$], indicating that as current L2 exposure increased, so did the effect of number length, both patterning with slower speech planning altogether (see Figure 1). The model also yielded a main effect of number length [$b = 64.053$, $SE = 9.142$, $t = 7.006$, $p < 0.0000001$], but no main effect of L2 exposure [$b = 22.445$, $SE = 24.481$, $t = 0.917$, $p = 0.375$].

https://doi.org/10.1017/S1366728920000115 Published online by Cambridge University Press

Table 7. Mean speech onset latencies and standard error (in ms) as a function of equations' first and second phrase length in French (L1) and English (L2) trials.

		Length of second phrase									
		1 syll		2 syll		3 syll		4 syll		**Conditions overall**	
	Length of first phrase	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	***M***	***SE***
French trials	1 syll	624	*33*	636	*34*	644	*31*	623	*28*	**632**	***29***
	2 syll	625	*26*	669	*28*	640	*31*	648	*31*	**645**	***28***
	3 syll	587	*27*	580	*28*	610	*29*	608	*29*	**596**	***26***
	4 syll	589	*28*	613	*31*	609	*37*	587	*28*	**599**	***29***
Conditions overall		**606**	***26***	**624**	***27***	**626**	***30***	**616**	***27***		
English trials	1 syll	716	*39*	745	*45*	749	*41*			**737**	***40***
	2 syll	705	*49*	692	*36*	704	*30*			**700**	***36***
	3 syll	742	*49*	748	*41*	696	*28*			**729**	***34***
Conditions overall		**721**	***44***	**728**	***37***	**717**	***30***				

Multi-phrase utterances (equations)

The core LME model for French-L1 equation trials revealed a significant main effect of first phrase length [$b = -16.9697$, $SE = 5.3515$, $t = -3.171$, $p = 0.005$], with speech onset latencies significantly DECREASING as first phrase length increased. The model revealed no main effect of second phrase length [$b = 3.5422$, $SE = 3.9551$, $t = 0.896$, $p = 0.3722$], and no interaction among factors [$b = 0.9123$, $SE = 3.9605$, $t = 0.23$, $p = 0.81819$]. The individual differences model revealed no interaction between current L2 exposure and either first or second phrase length [$b < -3.6$, $SE > 3.7$, $t < -1$, $p > 0.3$], and no 3-way interaction [$b = -3.578$, $SE = 3.719$, $t = -0.962$, $p = 0.336$]. (See Appendix A for complete model descriptions and outputs.) Thus, irrespective of L2 exposure, all participants showed the same speech production pattern in their production of multi-phrase utterances in their L1.

Speech onset latencies to English trials

Single-phrase utterances (isolated numbers)

The core LME model for English-L2 isolated number trials revealed a main effect of number length on speech onset latencies [$b = 59.06$, $SE = 12.07$, $t = 4.895$, $p < 0.0001$], with longer numbers patterning with longer speech onset latencies[1]. The individual differences model revealed an interaction between number length and current L2 exposure [$b = 17.29$, $SE = 6.557$, $t = 2.637$, $p = 0.0086$], with increased current L2 exposure patterning with a greater effect of number length, leading to slower speech planning altogether (see Figure 2). This pattern of effects mirrors the one observed in French trials, with increases in current L2 exposure predicting longer speech onset latencies in general. The model also yielded a significant main effect of number length [$b = 60.978$, $SE = 8.882$, $t = 6.865$, $p < 0.000001$], but no main effect of current L2 exposure [$b = 32.36$, $SE = 22.253$, $t = 1.454$, $p = 0.17$].

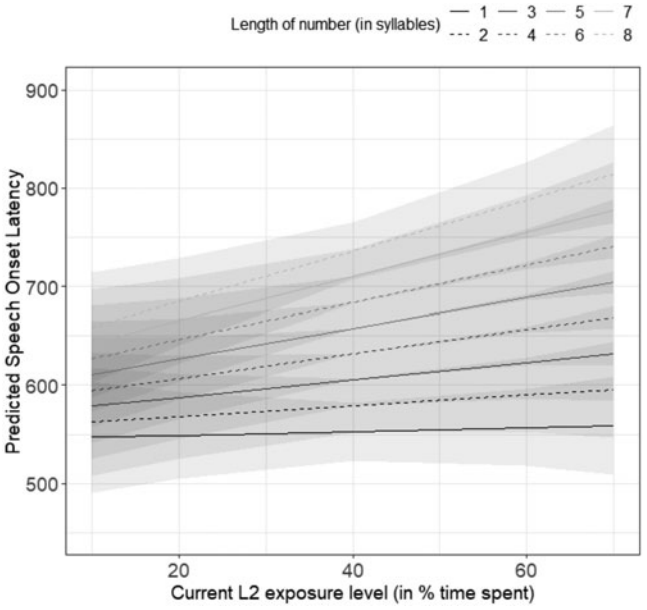


Fig. 1. Interaction between length of the number to produce and current L2 exposure on speech onset latencies for the French (L1) single-phrase utterances. Shaded area represents standard error.

Multi-phrase utterances (equations)

The core LME model for English-L2 equation trials yielded no main effects for the length of either the first or second phrase [$b < -4.95$, $SE > 9.85$, $t < -0.5$, $p > 0.6$], and no interaction between factors [$b = -11.494$, $SE = 9.694$, $t = -1.186$, $p = 0.254$][1]. However, the individual differences model revealed a significant interaction between L2 exposure and the length of the second phrase [$b = -16.745$, $SE = 7.197$, $t = -2.327$, $p = 0.0203$] (see Figure 3), with speech onset latencies increasing as the length of the second phrase increased among speakers with lower levels

[1]A deviation coded categorical variable was added to the core models to test for the potential impact of having been exposed to longer English trials (first 6 participants). The model revealed no significant main effect of task version [$b = 51.4182$, $SE = 48.1433$, $t = 1.068$, $p = 0.3036$] on the scope of planning for single-phrase English-L2 trials, nor did the task version interact with number length [$b = -0.7735$, $SE = 22.4929$, $t = -0.034$, $p = 0.9731$]. Similar results were also observed for multi-phrase English-L2 trials with no significant main effect of task version [$b = 108.7842$, $SE = 68.45$, $t = 1.589$, $p = 0.134$] and no interaction between task version and item lengths. [$b < -26.811$, $SE > 19.173$, $t < 1.4$, $p > 0.1$]. Hence, data from all participants can be analyzed within the same models.

https://doi.org/10.1017/S1366728920000115 Published online by Cambridge University Press

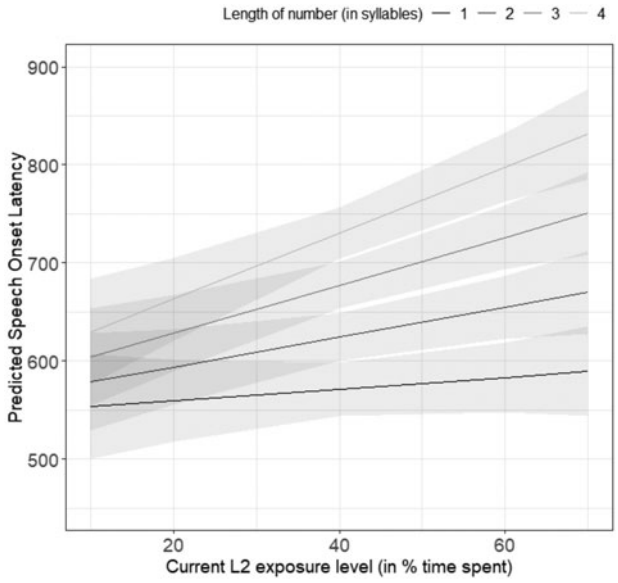


Fig. 2. Interaction between length of the number to produce and current L2 exposure on speech onset latencies to English (L2) single-phrase utterances. Shaded area represents standard error.

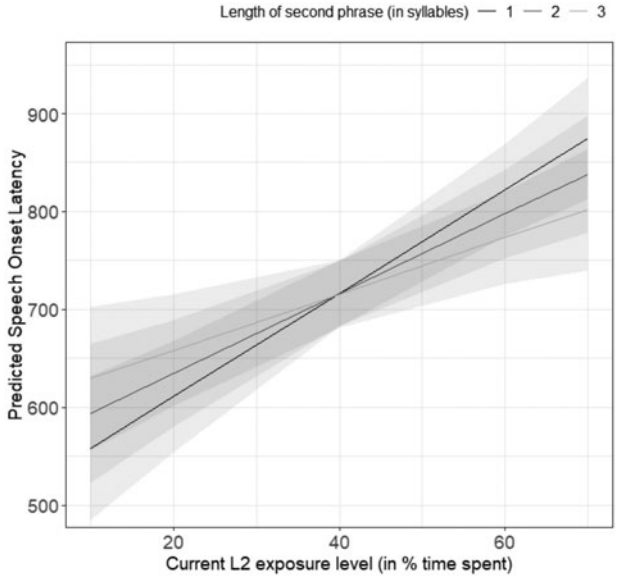


Fig. 3. Interaction between L2 exposure and second phrase length on speech onset latencies to English (L2) multi-phrase utterances. Shaded area represents standard error.

of current L2 exposure, but decreasing among speakers with the highest levels of current L2 exposure. The model also yielded a main effect of L2 exposure [$b = 69.699$, $SE = 32.006$, $t = 2.178$, $p = 0.0485$], but no interaction with the length of the first phrase, and no 3-way interaction between factors [$b < -11$, $SE > 7.2$, $t < -1.6$, $p > 0.1$]).

Comparison of French-L1 and English-L2 trials

The LME models comparing speech onset latencies to French-L1 and English-L2 trials of comparable length (single-phrase utterance: 1 to 4 syllables, multi-phrase utterances: 1 to 3 syllables per phrase), above and beyond current L2 exposure effects, yielded main effects of trial language for both single-phrase utterances [$b = -101.469$, $SE = 17.778$, $t = -5.708$, $p < 0.00001$] and multi-phrase utterances [$b = -91.891$, $SE = 10.330$, $t = -8.896$, $p < 0.00001$]. Accordingly, French-L1 trials were consistently produced with shorter speech onset latencies than English-L2 trials of comparable length.

General discussion

Our goal was to investigate the scope of L1 and L2 speech planning for the same bilingual adults, and whether it was modulated by individual differences in L2 experience and stimulus characteristics. The results from single-phrase utterances revealed a significant effect of trial language on speech onset latencies, suggesting that L2 speech planning is not equivalent to L1 speech planning (French-L1 trials being associated with shorter speech onset latencies than English-L2 trials.) The results also revealed that both the length (in syllables) of the to-be-produced number and individual differences in L2 experience were predictive of speech onset latencies in L1 and L2. Specifically, longer numbers patterned with longer speech onset latencies in both languages, and higher L2 exposure levels patterned with greater number length effects and slower speech onset latencies overall in both languages. In other words, stimulus length and individual differences in bilingual experience both predicted speech onset latency modulations, despite the fact that French-L1 trials were systematically produced with shorter speech onset latencies than English-L2 trials. Interestingly, one might have expected an opposite effect of current L2 exposure on L2 speech onset latencies, with increased L2 exposure leading to shorter L2 speech onset latencies instead of longer latencies as observed in the present data. Furthermore, the observation of current L2 exposure effects on L1 as well as L2 planning might be possible thanks to the generally high L2 proficiency level of the participants, high enough for L2 to influence L1 (De Groot, 2011).

Taken together, these overall patterns of data suggest that both L1 and L2 productions can be affected by language experience, regardless of the inherent differences between L1 and L2 planning. However, isolated numbers consist of single-phrase utterances, which do not allow the observation of scope of planning modulations *per se*. Consequently, we next turned our attention to the planning and production of the equation trials that consisted of multi-phrase utterances.

During the production of French-L1 multi-phrase utterances, speech onset latencies varied as a function of the length of the first but not the second phrase, suggesting that participants favored a short scope of planning in their L1 including only the first phrase of the utterance. However, speech onset latencies DECREASED as phrase length increased, suggesting that speakers adaptively modulated their scope of L1 planning. Accordingly, speakers would be more likely to initiate speech earlier if the first phrase was longer (possibly allowing speakers to finish planning the second phrase during its production), than when the first phrase was shorter (suggesting that they might have at least started to plan both phrases before initiating speech).

This suggests that the default scope of planning in L1 is short, including only the first phrase to produce, but that it can be adaptively modulated (extended) to fit the demands of specific utterances (see Griffin, 2003 for similar patterns of results). Interestingly, speech onset latencies of French-L1 equations could not be predicted by current L2 exposure levels, suggesting

https://doi.org/10.1017/S1366728920000115 Published online by Cambridge University Press

that speakers can adaptively modulate their scope of L1 planning to circumvent the effects of individual differences in L2 experience (as observed in single-phrase utterances).

In contrast, speech onset latencies to English-L2 multi-phrase utterances did not pattern with either phrase lengths or by their interaction. Nonetheless, current L2 exposure interacted with the length of the second phrase (not the first phrase), such that increasing the length of the second phrase predicted opposite effects for speakers with higher and lower levels of current L2 exposure. Accordingly, increasing the length of the second phrase led to longer latencies for speakers with lower L2 exposure and led to shorter latencies for speakers with higher L2 exposure (regardless of their overall substantially longer speech onset latencies). This interaction between SECOND phrase length and L2 exposure suggests that the default scope of planning in L2 is longer, here including both phrases to produce. It also suggests that the scope of L2 planning can, in certain circumstances, be adaptively modulated to circumvent the effects of increased L2 exposure and fit the demands of specific utterances (i.e., a speaker might decide to initiate speech before completely planning the second phrase, especially if the second phrase is long). Such pattern of results contrasts with the L1 results obtained from the same participants, supporting the idea that L2 speech planning is not equivalent to L1 speech planning. This conclusion is further bolstered by an LME model showing a main effect of language of production on equation trials, above and beyond stimuli characteristics and current L2 exposure effects.

Interestingly, as in the case of single-phrase utterances, one might have expected an opposite effect of current L2 exposure on speech onset latencies to L2 multi-phrase utterances. Namely, more exposure to the L2 was generally expected to make speakers more confident in their L2 abilities as well as more proficient in L2, which in turn might lead to shorter L2 speech onset latencies (instead of longer latencies as observed in the present data). To help constrain our interpretation of this counter-intuitive L2 exposure effect, we conducted a follow-up analysis taking participants' historical consistency of L2 exposure into account.

Impact of historical consistency of L2 exposure.

Because of Montreal's heterogeneous and dynamic linguistic environment, the participants in our sample differed not only in terms of current L2 exposure, but also in terms of historical consistency of their L2 exposure. For instance, some participants reported attending university in the same language as their previous school setting (either both in French, or both in English), which suggests that their current level of L2 exposure might be consistent with previously experienced L2 exposure levels (hereafter "historically consistent" participants). In contrast, some participants reported now attending university in English, while their previous school setting was in French, leading to a more recent increase in their current L2 exposure (hereafter "recent L2 increase" participants). Of note, a participant having experienced a recent increase in L2 exposure might still have relatively low current L2 exposure (i.e., might have gone from limited to moderate L2 exposure). Although independent sample t-tests revealed no differences across groups ("historically consistent" vs "recent L2 increase") with regard to the L2 proficiency, exposure, and dominance scores presented in Table 1, it is possible that this difference affected participants' scope of L2 planning. (See Appendix B for participants' details.)

Adding historical consistency of L2 exposure to the individual differences model revealed two significant 3-way interactions (between historical consistency of L2 exposure, current L2 exposure and FIRST PHRASE length and between historical consistency of L2 exposure, current L2 exposure and SECOND PHRASE length) but not a significant 4-way interaction. To interpret these interactions, we reanalyzed each group of participants separately. Of note, these follow-up analyses involve a very small number of participants per model (9 in the historically consistent group, 7 in the recent L2 increase group), and are only used to constrain interpretation of the three-way interactions (See Appendix B for complete model outputs).

The model for the recent L2 increase group revealed that they generally showed the same pattern found at the entire group level described above, with the second-phrase length having a different impact on speakers with higher vs lower levels of current L2 exposure. This suggests that speakers from the recent L2 increase group used a longer scope of planning in L2, taking into account the characteristics of the second phrase at the initial stages of speech planning. In contrast, the model for participants from the historically consistent group yielded only a main effect of first phrase length with speech onset latencies increasing alongside length of the first phrase. This suggests that speakers from the historically consistent group favored a shorter scope of planning that mainly included the first phrase of the utterance.

Therefore, we speculate that the counter-intuitive effect of current L2 exposure can potentially be explained in part by variations in the historical consistency of L2 exposure. One could also speculate that the overly long speech onset latencies for the recent L2 increase group are a sign that they may be less likely to keep their languages separate during production, thus leading them to be more susceptible to co-activation of number representations more generally. Alternatively, the recent L2 increase group may simply use a qualitatively different planning strategy, in which they take longer to initiate speech but take less time to articulate their productions. However, exploratory analyses of utterance durations for a subset of four participants revealed no such trade off between speech onset latency and utterance duration. (See Appendix B, Figure 3).

Interestingly, these two different patterns of results match the two main expectations stated in the *Introduction*. Namely, we expected that if scope of planning is primarily affected by L2 proficiency/fluency, then speakers would likely have a reduced scope of planning in L2, leading to effects of first phrase length on speech onset latencies but not second phrase length. We observed this pattern in speakers from the historically consistent group. However, unlike in L1 productions, the positive direction of the effect suggests that these speakers did not adaptively modulate their scope of planning to include the second phrase in cases where the first phrase was short. Instead, they show signs of consistently planning only the first phrase of the utterance, which suggests that their speech planning is less flexible in L2 compared to L1.

We also expected that if scope of planning is instead affected by reduced confidence in speaking in an L2, then speakers might be more cautious when producing L2, planning the entire utterance before initiating speech, leading to effects of second phrase length on speech onset latencies but not first phrase length. We observed this pattern in speakers from the recent increase group. Finally, we also expected that if higher levels of L2 exposure contributed to higher fluency and/or confidence in L2 speakers, then speakers with higher levels of current L2

https://doi.org/10.1017/S1366728920000115 Published online by Cambridge University Press

exposure may exhibit more adaptive L2 scope of planning compared to speakers who are exposed to lower current L2 exposure. This pattern of results was observed in speakers from the recent increase group, but not in speakers from the historically consistent group.

Similar follow-up analyses were also applied to speech onset latencies to English-L2 single-phrase utterances where higher levels of current exposure the L2 are associated with larger phrase-length effects. As observed for L2 multi-phrase utterances, speakers from the recent L2 increase group showed the same counter-intuitive pattern found at the entire group level (higher current L2 exposure levels patterning with greater number length effects), while current L2 exposure had no significant impact on the speech onset latencies of historically consistent speakers (only a main effect of number length).

Taken together, these results suggest that the scope of L2 planning is highly dependent on factors relating to individual differences in bilingual experience and specific stimuli characteristics. In addition to the central results about L1 and L2 scope of planning in bilingual speakers, the present study also yielded interesting information regarding the impact of utterance structure on scope of planning and bilingual speech production in general. We now address these in greater detail.

The impact of utterance structure on scope of planning

Although we observed SHORTER speech onset latencies for French L1 trials with a LONGER initial phrase, we did not observe a length threshold that could determine whether a speaker would initiate speech after planning only the first phrase of an utterance or not. Instead, we observed a continuous shortening of mean speech onset latencies as the first phrase got longer, indicating that THE PROBABILITY of a speaker initiating speech before having planned the second phrase increased as the first phrase got longer. These probabilities of a speaker modulating their planning strategy were also not significantly affected by current L2 exposure levels.

In addition, the fact that L2 exposure had a different impact on French-L1 multi-phrase utterances and single-phrase utterances, (regardless of their comparable total lengths) suggests that an utterance's internal structure has a significant impact on bilinguals' ability to modulate their scope of planning. This observation was confirmed by a repeated measures ANOVA on stimuli of overlapping length that revealed a significant interaction between stimuli length and number of phrases [$F(5,80) = 13.334$, $p < 0.0001$, $\eta^2 = .455$], with speech onset latencies being longer for single-phrase utterances than for segmentable multi-phrase utterances. Thus, multi-phrase utterances allowed speakers experiencing greater planning difficulties to modulate their speech planning strategy in order to alleviate said difficulty. Participants did not have the same opportunity with single-phrase utterances that had to be planned at once, even if they exceeded the limits of optimal serial working memory (Cowan, 2005).

Although isolated numbers did not provide a clear breaking point, the grand averages of Table 6 suggest that some speakers might have spontaneously broken their planning of the longest numbers into smaller chunks anyway. Given that the seven- and eight- syllable long numbers were built by adding "2000" (Fr.: *deux mille*, i.e., 2 syllables) in front of the five- and six-syllable long numbers, some speakers, particularly those with less efficient serial working memory, might have seized this opportunity to segment their planning into two chunks (planning the "2000" by itself, followed by the rest of the number). Such behavior might explain why the average speech onset latencies for the longest numbers do not follow the steady increase observed across shorter number length.

Implications for bilingualism

We interpret the totality of data presented here as consistent with models of bilingual language production that posit the constant parallel activation of both languages (Green, 1998; Kroll & Bialystok, 2013; Kroll, Bobb & Wodniecka, 2006). Indeed, most models generally agree that both languages are constantly active regardless of the target language to produce (at least to a certain degree), and that speakers must recruit their executive functions to manage the interference caused by the co-activation of both languages (Pivneva et al., 2012) – see also Kroll, Bobb and Hoshino (2014), for a review of the hypotheses regarding the involvement of different cognitive functions in L1-L2 language management. The results from the present study add to this body of work by examining the variables that can modulate the intensity or manageability of cross-language interference. The results not only demonstrate that individual differences in L2 exposure play an important role in planning simple L1 and L2 utterances, but also that the impact of L2 exposure may change over time.

Further, the counter-intuitive effect of L2 exposure (where increased current L2 exposure did not systematically benefit L2 production) suggests that, for increased L2 exposure levels to have a beneficial effect on L2 production, they have to be maintained over a certain period of time. Hence, a recent increase in L2 exposure can in fact be temporarily detrimental to L2 production given the increased cross-language co-activation it might occasion. An open question, therefore, is exactly how long such increased L2 exposure must be maintained before participants begin to experience improvements in L2 speech planning. This time-frame presumably varies across speakers in terms of different variables related to the bilingual experience like the age of first exposure to L2, time spent in L2 at different moments of their lives, or their general language learning abilities. The beneficial or detrimental effect of increased L2 exposure might also be related to the speakers' overall executive functions, which get recruited to manage this cross-language interference. Therefore, one could hypothesize that speakers with better executive function might be more efficient at managing increases in their L2 exposure levels compared to others.

Conclusion

In sum, the results of this study support several conclusions about bilingual speech planning. First and foremost, L2 speech planning is not equivalent to L1 speech planning, and should therefore be investigated in its own right. Second, speakers' bilingual experience has a significant impact on both L2 and L1 speech planning. Finally, speakers treat stimulus characteristics (e.g., phrase length) differently when speaking in L1 and L2. Future research should therefore take into account both the specifics of the speakers' bilingual experience and stimuli characteristics, as well as the cognitive control abilities that get recruited by speakers to manage their bilingual experience as well as stimuli-level variables.

Finally, it is important to emphasize that our use of number stimuli offers the advantage of being highly overlearned and practiced in both the L1 and L2, compared to content words, the latter

https://doi.org/10.1017/S1366728920000115 Published online by Cambridge University Press

of which will be more affected by L1/L2 differences in lexical-semantic retrieval, syntactic processing, and potentially other psycholinguistic factors. Thus, the results presented here offer a characterization of what is maximally possible with respect to the scope of phonological planning during L2 and L1 speech production. We hope that such information will prove important for grounding the results of future work that investigates similar issues using words that have more content, and consequently, also incorporate the simultaneous demands of semantic planning.

Supplementary Material. Supplementary material can be found online at https://doi.org/10.1017/S1366728920000115.

Acknowledgements. This work was supported by a postdoctoral fellowship awarded to A. C. Gilbert by the Fonds Québécois de Recherche sur la Société et la Culture, and by an NSERC Discovery Grant awarded to D. Titone.

References

Alario, FX, Ferrand, L, Laganaro, M, New, B, Frauenfelder, UH and Segui, J (2004) Predictors of picture naming speed. *Behavior research methods, instruments, & computers* **36**, 140–155. doi:10.3758/BF03195559

Allum, PH and Wheeldon, LR (2007) Planning scope in spoken sentence production: the role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **33**, 791–810. doi:10.1037/0278-7393.33.4.791

Baayen, RH (2008) *Analyzing linguistic data : a practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Barr, DJ, Levy, R, Scheepers, C and Tily, HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* **68**, 255–278. doi:10.1016/j.jml.2012.11.001

Bates, D, Maechler, M, Bolker, B and Walker, S (2014) lme4: Linear mixed-effects models using Eigen and S4. (Version R package version 1.1-6). Retrieved from http://CRAN.R-project.org/package=lme4

Brown-Schmidt, S and Konopka, AE (2015) Processes of incremental message planning during conversation. *Psychonomic bulletin & review* **22**, 833–843. doi:10.3758/s13423-014-0714-2

Chakraborty, R and Goffman, L (2011) Production of Lexical Stress in Non-Native Speakers of American English: Kinematic Correlates of Stress and Transfer. *Journal of Speech, Language and Hearing Research* **54**, 821–835A. doi:10.1044/1092-4388(2010/09-0018)

Clément, R, Gardner, RC and Smythe, PC (1977) Motivational variables in second language acquisition: A study of Francophones learning English. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement* **9**, 123–133. doi:10.1037/h0081614

Clément, R, Gardner, RC and Smythe, PC (1980) Social and individual factors in second language acquisition. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement* **12**, 293–302. doi:10.1037/h0081081

Costa, A and Caramazza, A (2002) The Production of Noun Phrases in English and Spanish: Implications for the Scope of Phonological Encoding in Speech Production. *Journal of Memory and Language* **46**, 178–198. doi:10.1006/jmla.2001.2804

Cowan, N (2005) *Working memory capacity*. New York: Psychology Press.

Damian, MF, Bowers, JS, Stadthagen-Gonzalez, H and Spalek, K (2010) Does word length affect speech onset latencies when producing single words? *Journal of Experimental Psychology: Learning, Memory, and Cognition* **36**, 892–905. doi:10.1037/a0019446

Damian, MF and Dumay, N (2007) Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language* **57**, 195–209. doi:10.1016/j.jml.2006.11.001

De Groot, AMB (2011) *Language and cognition in bilinguals and multilinguals: An introduction*. New York, NY: Psychology Press.

Ferreira, F (1991) Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language* **30**, 210–233. doi:10.1016/0749-596X(91)90004-4

Ferreira, F and Swets, B (2002) How Incremental Is Language Production? Evidence from the Production of Utterances Requiring the Computation of Arithmetic Sums. *Journal of Memory and Language* **46**, 57–84. doi:10.1006/jmla.2001.2797

Finch, G (2000) *Linguistic Terms and Concepts*. UK: Macmillan Education.

Fox, J (2003) Effect Displays in R for Generalised Linear Models. *Journal of Statistical Software* **8**, 1–27. doi:10.18637/jss.v008.i15

Gollan, TH, Montoya, RI, Cera, C and Sandoval, TC (2008) More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language* **58**, 787–814. doi:10.1016/j.jml.2007.07.001

Gollan, TH, Slattery, TJ, Goldenberg, D, Van Assche, E, Duyck, W and Rayner, K (2011) Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General* **140**, 186–209. doi:10.1037/a0022256

Green, DW (1998) Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* **1**, 67–81. doi:10.1017/S1366728998000133

Griffin, ZM (2003) A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic bulletin & review* **10**, 603–609. doi:10.3758/bf03196521

Hermans, D, Bongaerts, T, De Bot, K and Schreuder, R (1998) Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition* **1**, 213–229. doi:10.1017/S1366728998000364

Hermans, D, Ormel, E, van Besselaar, R and van Hell, J (2011) Lexical activation in bilinguals' speech production is dynamic: How language ambiguous words can affect cross-language activation. *Language and Cognitive Processes* **26**, 1687–1709. doi:10.1080/01690965.2010.530411

Holmes, VM (1988) Hesitations and sentence planning. *Language and Cognitive Processes* **3**, 323–361. doi:10.1080/01690968808402093

Jacobs, A, Fricke, M and Kroll, JF (2016) Cross-Language Activation Begins During Speech Planning and Extends Into Second Language Speech. *Language Learning* **66**, 324–353. doi:10.1111/lang.12148

Kroll, JF and Bialystok, E (2013) Understanding the Consequences of Bilingualism for Language Processing and Cognition. *Journal of Cognitive Psychology* **25**, 497–514. doi:10.1080/20445911.2013.799170

Kroll, JF, Bobb, SC and Hoshino, N (2014) Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current directions in psychological science* **23**, 159–163. doi:10.1177/0963721414528511

Kroll, JF, Bobb, SC and Wodniecka, Z (2006) Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition* **9**, 119–135. doi:10.1017/S1366728906002483

Kuznetsova, A, Brockhoff, PB and Christensen, RHB (2015) lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-29.

Levelt, WJM (1989) *Speaking: From intention to articulation*. Cambridge: MIT Press.

Lindsley, JR (1975) Producing simple utterances: How far ahead do we plan? *Cognitive Psychology* **7**, 1–19. doi:10.1016/0010-0285(75)90002-X

Marian, V, Blumstein, SE and Kaushanskaya, M (2007) The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research* **50**, 940–967. doi:10.1044/1092-4388(2007/067)

Martin, RC, Crowther, JE, Knight, M, Tamborello II, FP and Yang, C.-L. (2010) Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition* **116**, 177–192. doi:10.1016/j.cognition.2010.04.010

Martin, RC, Miller, M and Vu, H (2004) Lexical-semantic retention and speech production:further evidence from normal and brain-damaged participants for a phrasal scope of planning. *Cognitive Neuropsychology* **21**, 625–644. doi:10.1080/02643290342000302

Meyer, AS (1996) Lexical Access in Phrase and Sentence Production: Results from Picture–Word Interference Experiments. *Journal of Memory and Language* **35**, 477–496. doi:10.1006/jmla.1996.0026

Meyer, AS, Belke, E, Häcker, C and Mortensen, L (2007) Use of word length information in utterance planning. *Journal of Memory and Language* **57**, 210–231. doi:10.1016/j.jml.2006.10.005

Meyer, AS, Sleiderink, AM and Levelt, WJ (1998) Viewing and naming objects: eye movements during noun phrase production. *Cognition* **66**, B25–833. doi:10.1016/S0010-0277(98)00009-2

https://doi.org/10.1017/S1366728920000115 Published online by Cambridge University Press

Meyer, AS and van der Meulen, FF (2000) Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic bulletin & review* **7**, 314–319. doi:10.3758/BF03212987

Michel Lange, V, Cheneval, PP, Python, G and Laganaro, M (2017) Contextual phonological errors and omission of obligatory liaison as a window into a reduced span of phonological encoding. *Aphasiology* **31**, 201–220. doi:10.1080/02687038.2016.1176121

Oppermann, F, Jescheniak, JD and Schriefers, H (2010) Phonological advance planning in sentence production. *Journal of Memory and Language* **63**, 526–540. doi:10.1016/j.jml.2010.07.004

Pivneva, I, Palmer, C and Titone, D (2012) Inhibitory control and l2 proficiency modulate bilingual language production: evidence from spontaneous monologue and dialogue speech. *Frontiers in Psychology* **3**, 57. doi:10.3389/fpsyg.2012.00057

Prior, A, Katz, M, Mahajna, I and Ribinsten, O (2015) Number word structure in first and second language influences arithmetic skills. *Frontiers in Psychology* **6**(266). doi:10.3389/fpsyg.2015.00266

R Development Core Team. (2010) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org

Salillas, E and Wicha, NY (2012) Early learning shapes the memory networks for arithmetic: Evidence from brain potentials in bilinguals. *Psychological science* **23**, 745–755. doi:10.1177/0956797612446347

Smith, M and Wheeldon, L (1999) High level processing scope in spoken sentence production. *Cognition* **73**, 205–246. doi:10.1016/S0010-0277(99)00053-0

Smith, M and Wheeldon, L (2004) Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **30**, 675–686. doi:10.1037/0278-7393.30.3.675

Spelke, E and Tsivkin, S (2001) Language and number: A bilingual training study. *Cognition* **78**, 45–88. doi:10.1016/S0010-0277(00)00108-6

Swets, B, Jacovina, ME and Gerrig, RJ (2014) Individual differences in the scope of speech planning: evidence from eye-movements. *Language and Cognition* **6**, 12–44. doi:10.1017/langcog.2013.5

Tabachnick, BG and Fidell, LS (2007) *Using multivariate statistics* (5th ed. ed.). Boston: Pearson/Allyn & Bacon.

Thordardottir, E (2011) The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism* **15**, 426–445. doi:10.1177/1367006911403202

Thordardottir, E and Brandeker, M (2013) The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders* **46**, 1–16. doi:10.1016/j.jcomdis.2012.08.002

Van Rinsveld, A, Brunner, M, Landerl, K, Schiltz, C and Ugen, S (2015) The relation between language and arithmetic in bilinguals: Insights from different stages of language acquisition. *Frontiers in Psychology* **6**(265). doi:10.3389/fpsyg.2015.00265

Van Rinsveld, A, Dricot, L, Guillaume, M, Rossion, B and Schiltz, C (2017) Mental arithmetic in the bilingual brain: Language matters. *Neuropsychologia* **101**, 17–29. doi:10.1016/j.neuropsychologia.2017.05.009

Van Rinsveld, A, Schiltz, C, Brunner, M, Landerl, K and Ugen, S (2016) Solving arithmetic problems in first and second language: Does the language context matter? *Learning and Instruction* **42**, 72–82. doi:10.1016/j.lerninstruc.2016.01.003

Vermeer, A (2001) Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics* **22**, 217–234. doi:10.1017/s0142716401002041

Wagner, V, Jescheniak, JD and Schriefers, H (2010) On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **36**, 423–440. doi:10.1037/a0018619

Wheeldon, LR and Lahiri, A (1997) Prosodic units in speech production. *Journal of Memory and Language* **37**, 356–381. doi:10.1006/jmla.1997.2517

Wheeldon, LR and Lahiri, A (2002) The minimal unit of phonological encoding: prosodic or lexical word. *Cognition* **85**, B31–B41. doi:10.1016/S0010-0277(02)00103-8

Wickham, H (2009) *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.

Yang, J.-C. and Yang, Y.-F. (2008) Horizontal Flow of Semantic and Phonological Information in Chinese Spoken Sentence Production. *Language and speech* **51**, 267–284. doi:10.1177/0023830908098543

Zhao, L.-M., Alario, FX and Yang, Y.-F. (2015) Grammatical planning scope in sentence production: Further evidence for the functional phrase hypothesis. *Applied Psycholinguistics* **36**, 1059–1075. doi:10.1017/S0142716414000046

https://doi.org/10.1017/S1366728920000115 Published online by Cambridge University Press