Bilingualism: Language and Cognition

cambridge.org/bil

Planning scope in second language sentence production: Do bilingual speakers plan ahead more in L1 than in L2?

Felicity F. Frinsel[1] and Robert J. Hartsuiker[2]

[1]Cornell University, Department of Psychology, Ithaca, NY, USA and [2]Ghent University, Department of Experimental Psychology, Ghent, Belgium

Research Article

Cite this article: Frinsel FF, Hartsuiker RJ (2023). Planning scope in second language sentence production: Do bilingual speakers plan ahead more in L1 than in L2? *Bilingualism: Language and Cognition* **26**, 684–694. https://doi.org/10.1017/S1366728923000068

Received: 2 August 2021
Revised: 14 December 2022
Accepted: 25 January 2023
First published online: 15 February 2023

Keywords:
Planning scope; bilingualism; sentence processing; speech production

Address for correspondence:
Felicity F. Frinsel
Department of Psychology,
Cornell University
Ithaca, New York 14853
Email: fff26@cornell.edu

Abstract

Language production is incremental in nature; we tend to plan linguistic chunks prior to articulating the first word of the utterance. Researchers have acquired knowledge about how far ahead sentences are generally planned, but mostly in monolinguals or the speaker's first language (Allum & Wheeldon, 2007; Martin, Crowther, Knight, Tamborello II, & Yang, 2010; Wagner, Jescheniak, & Schriefers, 2010). It is unclear whether the scope of planning is the same in bilinguals, or the speaker's second language. Here, we examined planning scope in Dutch–English bilinguals' sentence production using a paradigm that elicits descriptions of short animations. Analyses of speech onset times and articulation durations suggest that, on the surface, bilinguals have comparable planning scope in L1 and L2. However, in their L2, bilinguals extended their articulation duration, suggesting that they committed early to the initial noun phrase, but produced it more slowly to buy time to plan the next noun phrase.

Introduction

It seems as though producing a sentence is rather easy. However, the speaker has to go through several complex cognitive processes in order to convert thoughts into speech. That is, as soon as the speaker has a conceptual representation of what they want to convey, lexical representations – or lemmas – are accessed from the mental lexicon and ordered according to the grammatical rules of the language being used (grammatical encoding). Once the lemmas are accessed, sounds are retrieved and organized to generate an articulatory plan (phonological encoding). Subsequently, articulatory movements are executed, resulting in overt speech (Bock & Levelt, 1994; Dell, Chang, & Griffin, 1999; Levelt, 1989). These stages are reflected in models of language production that generally consist of three overarching levels of processing: a message level, a grammatical encoding level, and a phonological encoding level.

Most speech production models suggest that the preparation for an utterance happens in an incremental manner (e.g., Levelt, 1989), meaning that speakers do not wait until they have planned the entire sentence before they start pronouncing the first word. Previous studies on monolingual speech production have aimed to establish the extent to which sentences are planned in advance, also referred to as planning scope. The goal of the present paper is to determine the planning scope in bilingual speakers' L1 and L2.

In search of the minimal unit of planning, some authors have argued that speech planning occurs in a word-by-word fashion (e.g., Bonin, Malardier, Méot, & Fayol, 2006; Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998; Zhao & Yang, 2016), whereas others have argued for a phrasal scope of planning (Allum & Wheeldon, 2007; Lee, Brown-Schmidt, & Watson, 2013; Martin et al., 2010; Smith & Wheeldon, 1999). However, a growing number of studies have shown that the extent to which sentences are planned in advance is not fixed but varies for the different levels of encoding (e.g., Konopka & Meyer, 2014; Meyer, 1996; Zhao & Yang, 2016) and seems to depend on a myriad of external (Ferreira & Swets, 2002; Konopka, 2012; Krivokapic, 2012; Swets, Jacovina, & Gerrig, 2013; Wagner et al., 2010) and internal factors (Bishop & Intlekofer, 2020; Swets, Jacovina, & Gerrig, 2014; Swets, Fuchs, Krivokapić, & Petrone, 2021).

For example, Griffin (2003) found that gaze durations prior to speech onset were extended for the second noun in a two-word pair when the first noun was monosyllabic rather than multisyllabic. Thus, when the first word was monosyllabic, extended speech onset latencies were observed, suggesting that both nouns were planned simultaneously. However, when adding "next to" between the two nouns this reversed word-length effect was no longer found. Hence, she concluded that the scope of planning was adapted as a function of available preparation time for subsequent words.

In a study by Meyer (1996) participants were asked to describe two picture displays using the sentence structure "The A is next to the B" (e.g., the mug is next to the nail). At the same time, an auditory distractor was presented that was semantically (e.g., cup for mug and screw for nail) or phonologically (e.g., bug for mug and mail for nail) related to one of the target nouns. When the distracter noun was semantically related to either the first or second target noun, an increase in speech onset time was observed. However, in case of phonological relatedness, speech onset time was only affected when the distracter and the first target noun were related. These findings indicate that the lexical representations of the nouns A and B may be retrieved before the onset of speech, but the phonological structure is not.

Additionally, Wagner and colleagues (2010) replicated the effects found by Meyer (1996) but hypothesized that if speakers adapt to an increase in cognitive load by readjusting the scope of advance planning (i.e., narrowing down the scope of advance planning with enhanced cognitive load), there should be semantic interference for the first noun only. After increasing cognitive load by adding color information (e.g., "the red frog is next to the blue cup"), the interference effect observed earlier in the processing of simple noun phrases was enhanced, and this interference effect was larger for the first noun than the second noun. A second experiment showed that when the cognitive load on production was manipulated by having the participants produce sentences that alternated with respect to color information (i.e., sentences without color information and sentences with color information), the scope of advance planning became smaller and the proportion of incrementally produced utterances increased. That is, when required to switch between utterance formats, advance planning for the second noun phrase was no longer found in sentences with and without color information. In a third experiment participants produced simple sentences while retaining a set of items (either digits or adjectives) in working memory. Interestingly, although working memory load did affect speech production (increased working memory load resulted in longer naming latencies), there was no reduced scope of advance planning when participants were presented with a concurrent verbal working memory task. Thus, the cognitive load manipulation that required a switch between utterance formats affected the scope of planning differently than the cognitive load manipulation (i.e., dual task) that was not directly related to speech planning.

The effects of visuospatial and verbal working memory load on the scope of planning was further explored by Klaus, Mädebach, Oppermann, and Jescheniak (2017). They found that the lexical representations of both nouns were activated prior to speech onset, regardless of working memory load condition. However, at the phonological level, both nouns were activated within the visuospatial working memory load condition, but no object-related distractor effect was observed in the verbal working memory load condition. Thus, a reduced planning scope was only found on the phonological level as a function of the type of working memory load (for more direct evidence that planning scope is linked to working memory, see Bishop & Intlekofer, 2020; Klaus & Schriefers, 2018; Swets et al., 2014).

Moreover, Oppermann, Jescheniak, and Schriefers (2010) illustrated that the scope of planning on the phonological level can be affected by familiarity with the sentence structure. The syntactic structure of the to-be-produced sentence depended on the lead-in fragment: in German, the lead-in fragment "Man sah wie…" [one saw how…] elicits the production of a subject-object-verb (SOV) utterance, whereas the lead-in fragment "Vorhin…" [earlier …] elicits a verb-subject-object (VSO) structure instead. They found that both nouns were phonologically activated when using the former structure, whereas for the latter structure, only the first noun was found activated prior to the speech onset. Phonological planning scope was thus affected by the sentence construction, initiated by the lead-in fragment. In German, the VSO sentence construction is rather infrequent, which could possibly explain the smaller scope of planning for those sentences (Dryer, 2005). Specifically, due to its low frequency, VSO sentences could be more difficult to process, and the increase in processing demands might have contributed to a reduced advance planning scope.

However, although some studies have shown that there are cross-linguistic differences in planning scope as well (e.g., Brown-Schmidt & Konopka, 2008; Myachykov, Scheepers, Garrod, Thompson, & Fedorova, 2013; Swets et al., 2021), there are limited data available to indicate that the strategies and patterns observed in monolingual speech planning could be generalized to speech planning in bilinguals (but see Gilbert, Cousineau-Perusse, & Titone, 2020; Konopka, Meyer, & Forest, 2018; Li, Ferreira, & Gollan, 2022). Therefore, to add to the expanding list of factors that alter planning scope, we investigated whether and how the scope of planning differs with respect to L1 and L2 sentence processing in bilinguals. In this study, we refer to bilinguals as those who speak two languages, acquired successively, and are highly proficient in their non-native language.

Earlier studies on bilingualism have shown that L2 speakers (as compared to L1) are generally slower listeners (e.g., Lagrou, Hartsuiker, & Duyck, 2011), readers (e.g., Cop, Keuleers, Drieghe, & Duyck, 2015), and speakers (e.g., Broos, Duyck, & Hartsuiker, 2018; Ivanova & Costa, 2008; Sadat, Martin, Alario, & Costa, 2012), indicating greater computational effort in L2. As identifying, retrieving, planning, and producing words and phrases may be more taxing in bilinguals' L2 than in their L1 (Hanulová, Davidson, & Indefrey, 2011; Runnqvist, Strijkers, Sadat, & Costa, 2011), different planning strategies may be utilized to compensate for such heightened cognitive demands. For instance, Konopka et al. (2018) found that Dutch–English bilinguals tend to plan more hierarchically in their L2 (i.e., extensively encoding conceptual representations of a larger message prior to linguistic encoding) and more linearly in their L1 (i.e., planning messages concept-by-concept and utterances word-by-word). However, the degree to which working memory resources are demanded interacts with language proficiency in one's non-native language. That is, sentence processing expends more working memory resources when less proficient unbalanced bilinguals are using their non-dominant language (Service, Simola, Metsänheimo, & Maury, 2002). Thus, the scope of planning may also be affected by bilinguals' experience with each language. Indeed, both experience with the task and higher L2 proficiency levels led speakers to change from hierarchical planning to a more linear incremental planning in L2 (Konopka et al., 2018).

In a study by Gilbert et al. (2020) French–English bilinguals were asked to produce isolated numbers and numerical equations (e.g., four minus forty). When producing single numbers, bilinguals' speech onset latencies increased as the number of syllables in the to-be-produced number increased – both in their L1 and L2. However, for the numerical equations, the speech onset latencies decreased as the number of syllables increased for the first number in the equation, but only when produced in French. Interestingly, speech onset latencies in English increased as the length of the second number increased, but only for those with lower levels of current L2 exposure. Thus, also French–English bilinguals tend to have a planning scope that is larger in their L2 and smaller in their L1 and occurs to be modulated as a function of proficiency.

The abovementioned studies used tasks in which only simple sentences were elicited, even though planning scope has been found to vary as a function of phrase type. Therefore, the present study aims to investigate whether sentences with simple or complex initial phrases (hereafter simple-complex and complex-simple sentences) also result in different planning scopes in bilinguals and more specifically bilingual L2 production.

Earlier studies have reported longer onset latencies for complex-simple sentences (e.g., a bag and a rabbit jump next to an ax) than for simple-complex sentences (e.g., a bag jumps next to a rabbit and an ax), supporting a phrasal scope of planning. Because, if only the first word or the entire utterance was planned prior to speech onset, no such difference would have emerged as the initial word and the length of the utterance are identical between the matched sentences. Based on the literature, we hypothesized that if the planning scope is affected at the level of grammatical encoding, the speech onset times would be different for L1 and L2. That is, the complexity effect (i.e., longer onset latencies for complex-simple sentences as compared to simple-complex sentences) would be observed in bilinguals' L1 (indicating a phrasal scope of planning), but not in bilinguals' L2. Note though that we predicted the planning scope to be modulated as a function of proficiency, with more L1-like planning in L2 for those with higher L2 proficiency.

However, some studies have shown that advance planning at the phonological level is more affected (i.e., a smaller planning scope) in situations with increased processing demands than advance planning at the level of grammatical encoding. Thus, even if the complexity effect is found in bilinguals' L1 and L2, planning in the L1 and L2 could still differ on other levels. For that reason, rather than exclusively focusing on planning before the onset of speech, we used articulation durations as a dependent measure to examine any differences in strategies during speech.

Although speculative, we hypothesized that planning scope was likely affected at the level of phonological encoding if articulation durations were extended. This idea is based on earlier work on age differences in sentence processing in which the abovementioned complexity effect was observed for both young and older participants (Martin, Miller, & Vu, 2004), yet older but not younger participants looked longer at the second object after speech onset, resulting in either a pause between the first and second noun or extended articulation durations of the first noun (Mortensen, Meyer, & Humphreys, 2008). Therefore, we examined whether articulatory durations of the first noun differed across sentence types and languages and whether, in complex-simple sentences, the articulatory durations between the first and second noun differed across both languages. Again, the planning scope was expected to be modulated as a function of proficiency.

Method

Participants

Participants consisted of 36 unbalanced Dutch–English bilinguals (35 female, 1 male; age: 18–21), who were all first-year undergraduate psychology students recruited from the participant pool of Ghent University and earned one course credit for their participation. Data from eight additional participants were eliminated as a large number of target trials in one or both languages got lost due to technical issues. To control for years of language experience, four others were excluded as they were considerably older.

The Flemish university students tested here were native speakers of Dutch who had acquired English in formal school settings, but also had considerable exposure to English before school age from television, film, video games and so on (for more details see Broos, Bencivenni, Duyck, & Hartsuiker, 2021). As a measure of overall English proficiency, participants completed the *Lexical Test for Advanced Learners of English* (LexTALE; Lemhöfer & Broersma, 2012) at the end of the experimental session. The LexTALE is a self-paced lexical decision task designed to measure proficiency differences among advanced learners of English. The average score on the LexTALE was 71.5% ($SD = 8.58$, range = 56.25–92.5), indicating that they were proficient. In addition, they reported to have normal or corrected-to-normal vision.

Materials

We used a set of 112 black and white line-drawing pictures from the Severens, Van Lommel, Ratinckx, and Hartsuiker (2005) database. All pictures were familiar objects that were matched on frequency, length, number of syllables, age of acquisition, and *H*-statistic (i.e., an index of name agreement) between both languages. Fifty-eight of these pictures were used to create 45 sets of three-picture animations (hereafter triplets). The picture combinations were carefully chosen to ensure that the names of the three pictures were conceptually, orthographically, and phonologically different within and across both languages (e.g., L2: bag, rabbit, ax; L1: zak, konijn, bijl).

Simple-complex sentence constructions were required when the leftmost picture moved in a certain direction, while the other two pictures did not show any motion; complex-simple sentence constructions were elicited if both the left and the middle picture moved simultaneously in the same direction, while the third picture remained in place (as in Figure 1). Materials can be found on Open Science Framework (https://osf.io/ws7vx/).

Each triplet was presented once in the simple-complex and once in the complex-simple sentence condition. Within each sentence condition, the same pictures never occurred twice in the same screen position, and each picture combination was unique (e.g., bag, rabbit, ax; rabbit, scarf, bridge; kite, ax, duck). Each noun was thus produced four to six times throughout the experiment, but in different positions or constructions. In addition to manipulating sentence type, each movement type (jump next to, jump over, bounce towards, bounce under) occurred equally often. The order in which the movements were distributed was pseudorandomized to prevent the participants from foreseeing an upcoming movement type. Animations were made using PowerPoint software (version 15.30) and were converted into the Windows Media Video (WMV) format.

For the filler triplets, all three pictures simultaneously moved in the same direction (left, right, up, or down). Martin et al. (2010, see Experiment 3) found a smaller effect of complexity when fillers were structurally more similar to the initial noun phrase in complex-simple sentences. To bypass any effect of structural overlap on speech onset latencies, participants were instructed to produce sentences like "They all roll up" instead of "A bag and a rabbit and an ax roll up." Twenty-nine pictures were used to create 20 sets of filler triplets. Even though identical filler triplets never occurred twice within the same block, they were repeated three times over the course of the experiment. However, those triplets were always presented with different movement types (roll up/down/left/right; move up/down/left/right; bounce left/right).

https://doi.org/10.1017/S1366728923000068 Published online by Cambridge University Press

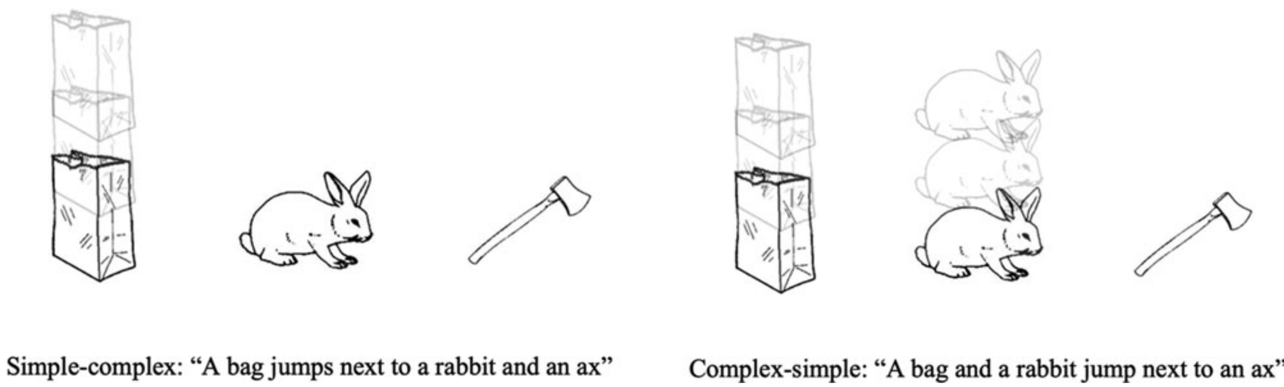

Fig. 1. Examples of animations and corresponding responses for simple-complex and complex-simple sentences. The color transition illustrates the motion of the picture(s), here moving up and down on the screen.

The experimental session was blocked in two ways. First, the experiment was divided into two language blocks. That is, after participants completed the task in one language, they did the same task again in the other language. The language condition was counterbalanced such that half of the participants started in Dutch and ended in English, whereas the other half started in English and ended in Dutch, allowing for control over variables such as familiarity and motivation. Second, within each language condition, there were six blocks of 15 target triplets and 10 filler triplets, making up a total of 90 target trials and 60 filler trials per language block. Triplets presented in Block 1 in simple-complex constructions were again presented in Block 4 in complex-simple constructions. That is, triplets in Blocks 1, 2, and 3 were repeated in Blocks 4, 5, and 6, but the sentence condition and the movement type differed. Blocks were counterbalanced across participants (i.e., half of the participants completed the blocks in a reversed order), and trials were randomized within each block.

Procedure

The experiment was run using E-prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2012). Each participant was seated in a silent room and placed in front of a computer screen. The experiment consisted of a familiarization phase, a practice phase, and an experimental phase. During the familiarization phase, participants were first informed about the movement types and sentence constructions, while they were looking at animations of geometrical figures rather than pictures of everyday objects. Next, they briefly studied the pictures and their corresponding names by themselves to ensure that participants would use the same labels, necessary for our articulation duration analyses.

Participants then completed a set of 20 practice trials, created out of the additional 24 pictures, to ensure they understood the task and could use the correct sentence constructions. In both the practice and experimental phase, a rectangular black frame was centered across the screen for 1000 ms. After its offset, three pictures were presented on the screen which began to move after 300 ms for a total duration of 2000 ms. Participants were instructed to describe the animation from left to right as soon as they could within the given timeframe of 5000 ms. The next trial started after a blank screen was presented for 2000 ms. Participants were encouraged to take a break after the third block.

After completing all six blocks, the same procedure, including the familiarization and practice phases, was repeated in the other language. Responses were recorded by the InSound function in E-prime 2.0 using the microphone on a Plantronics Audio 355 headset. The onset of the recordings was aligned with the onset of the animations, resulting in separate .wav files with a duration of 5000 ms each. The experiment took one hour to complete.

Data analysis

Speech onset time, measured as the duration between the start of the animation and the beginning of participants' speech, articulation duration of the first noun in simple-complex and complex-simple sentences, and the time between the offset of the first and the onset of the second noun as well as articulation durations of the second noun in complex-simple sentences were analyzed using Praat speech analysis software (Boersma & Weenink, 2018). The speech onset time and the articulation durations, measured in milliseconds, were manually determined based on visual inspection of the waveform.

Observations were treated as outliers if the absolute z-score was equal to or greater than 2.5. Additionally, the experimenter coded each trial for accuracy and technical errors. Data points were coded as erroneous and excluded from further analysis if the speaker: (1) used a definite instead of an indefinite article[1]; (2) repaired or repeated (part of) an utterance; (3) misnamed at least one picture within a triplet; or (4) did not respond. Utterances that were not properly recorded by E-prime 2.0 were considered technical errors. According to these criteria, 19% of data points (1217 observations) were marked as erroneous or malfunctions, and 2% of the trials (129 observations) were marked as outliers. Final analyses were based on 5134 trials.

Data analyses were carried out in R version 4.1.0 (R Core Team, 2021) and RStudio version 1.4.1106 using the packages lme4 1.1-27 (Bates, Maechler, Bolker, & Walker, 2015), emmeans 1.6.1 (Lenth, 2021), and lmerTest 3.1.3 (Kuznetsova, Brockhoff, & Christensen, 2017). To estimate main effects and interaction effects Type II Wald Chi-square tests were conducted. To interpret interactions including factors with more than three levels pairwise post-hoc comparisons were used.

[1]Participants were instructed to use an indefinite article instead of a definite article to limit variation between both language conditions. That is, in Dutch two definite articles are used (*de* and *het*) and only one indefinite article (*een*), while in English only one definite (*the*) and indefinite (*a/an*) article is used.

https://doi.org/10.1017/S1366728923000068 Published online by Cambridge University Press

Results

Before speaking

To analyze speech onset latencies, a series of linear mixed-effect models were fitted. First, we tested the overall effect of complexity (simple-complex and complex-simple) and language (L1 and L2) on speech onset latencies. Interactions were added between all fixed effects. The base model contained by-subject and by-item random intercepts. Moreover, we also included by-subject random slopes for language and complexity. By doing so, we allowed the variances for the two levels of language and complexity to be different. For the analysis of speech onset times, trials that yielded erroneous responses were excluded.

We note that including language order in the statistical model appeared to modulate the effect of language, indicated by an interaction between language order and language, but the main effects and interactions of interest remained untouched. Considering the fact that we were not testing but controlling for language order, we refrained from reporting these language order effects in this manuscript for the sake of a clear presentation of the main findings, but the model outputs can be found in the online supplementary materials on Open Science Framework.

Next, participants' L2 LexTALE scores were added to the base model to evaluate how individual differences in L2 proficiency modulated the overall effects in L2. This linear mixed-effect model contained fixed factors for complexity (simple-complex and complex-simple) and LexTALE score (scaled continuous variable). The structure of the random intercepts was the same as that in the base model, but only included by-subject random slopes for complexity.

Speech onset times

Table 1 shows the estimated mean speech onset times per language and sentence type.

The estimated group difference between complex-simple and simple-complex sentences, averaged over the levels of language, was 70 (95% CI: 51–89), meaning that the estimated mean speech onset time for complex-simple sentences was 70 milliseconds slower than the estimated mean speech onset time for simple-complex sentences. The parameter estimates for speech onset time indicated that there was an overall main effect of complexity ($\chi^2(1) = 51.14$, $p < .001$), but no interaction between language and complexity ($\chi^2(1) = .26$, $p = .612$), indicating that more time was taken to start complex-simple sentences and that the complexity effect did not differ across languages (Figure 2). However, there was a main effect of language ($\chi^2(1) = 9.58$, $p = .002$), with average speech onset times being faster in L2 (1144 ms) than speech onset times in L1 (1213 ms; $\beta = 68.14$, $SE = 22.02$, $z = 3.09$, $p = .002$). A supplemental analysis revealed that speech onset times numerically decreased across blocks in L1 ($\beta = 64.7$, $SE = 38.4$, $z = 1.69$, $p = .54$) and L2 ($\beta = 81$, $SE = 38.5$, $z = 2.1$, $p = .29$), but repetition did not affect speech onset latencies differently in L1 and L2 ($p = .06$).

The model accounting for individual differences in L2 proficiency revealed no main effect of L2 proficiency ($\chi^2(1) = 1.31$, $p = .25$) or interaction effect between L2 proficiency and complexity ($\chi^2(1) = 0.06$, $p = .80$).

While speaking

To determine whether there was a difference in planning during speech, we subsequently tested whether articulatory durations of the first noun differed across the sentence types within and between each language. For complex-simple sentences, the articulation duration between the first and second noun as well as the articulation duration of the second noun was compared across both languages.

First, we tested the overall effect of complexity and language on the articulation duration of the first noun. Interactions were added between both fixed effects. Second, the articulation duration between the offset of the first noun and the onset of the second noun in complex-simple sentences was analyzed with language as a fixed effect. Next, a linear mixed-effect model was fitted to analyze the articulation duration of the second noun in complex-simple sentences, with language as a fixed effect.

Finally, participants' L2 LexTALE scores were added to all three base models to evaluate how individual differences in L2 proficiency modulated the overall effects in L2. The first linear mixed-effect model contained fixed factors for complexity (simple-complex and complex-simple) and LexTALE score (scaled continuous variable), whereas the other two models only contained LexTALE score as a fixed factor. For all models by-subject and by-item random intercepts were included. Additionally, only for the first model by-subject random slopes for complexity were included.

Articulation duration of first noun

Table 2 presents a summary of the articulation duration measures for the first noun.

The estimated group difference between complex-simple and simple-complex sentences, averaged over the levels of language, was −11 (95% CI: −19.94 - −2.33), meaning that the estimated mean articulation duration of the first noun in complex-simple sentences was 11 ms shorter than the estimated mean articulation duration of the first noun in simple-complex sentences. The parameter estimates for articulation duration of the first noun indicated that there was an overall main effect of Complexity ($\chi^2(1) = 6.09$, $p = .013$) with more time taken to articulate the first noun in simple-complex sentences.

There was a main effect of language ($\chi^2(1) = 54.63$, $p < .001$) as the articulation duration of the first noun in L1 (353 ms) was significantly shorter compared to the articulation duration of the first noun in L2 (405 ms; $\beta = -54.28$, $SE = 7.35$, $z = -7.38$, $p < .001$). Moreover, language interacted with complexity ($\chi^2(1) = 26.29$, $p < .001$). In L2, the first noun in simple-complex sentences took more time to produce than that in complex-simple sentences ($\beta = 23.52$, $SE = 5.11$, $z = 4.61$, $p < .001$), but no such difference was observed in L1 ($\beta = 1.25$, $SE = 5.10$, $z = 0.25$, $p = .81$). Thus, the main effect of complexity appears to be driven by this language effect in which the articulation duration of the first noun in simple-complex sentences (as compared to the first noun in complex-simple sentences) was only significantly extended in L2 (see Figure 3A).

The model accounting for individual differences in L2 proficiency revealed an interaction between complexity and L2 proficiency ($\chi^2(1) = 6.80$, $p = .009$). As shown in Figure 3B, the articulation duration of the first noun was more extended in simple-complex sentences in L2 for those with lower L2 proficiency.

Articulation duration between nouns

There was a main effect of language ($\chi^2(1) = 6.47$, $p = .01$), indicating that the average articulation duration between the offset of the first noun and the onset of the second noun (i.e., 'and a

https://doi.org/10.1017/S1366728923000068 Published online by Cambridge University Press

Table 1. Estimated mean onset latencies in milliseconds, standard errors, and 95% confidence intervals (CI) as a function of sentence type for both L1 and L2.

	Language					
	L1			*L2*		
Sentence type	Mean	SE	95% CI	Mean	SE	95% CI
Simple-Complex	1179	34	1112 – 1247	1107	36	1036 – 1178
Complex-Simple	1245	35	1176 – 1314	1181	39	1104 – 1259

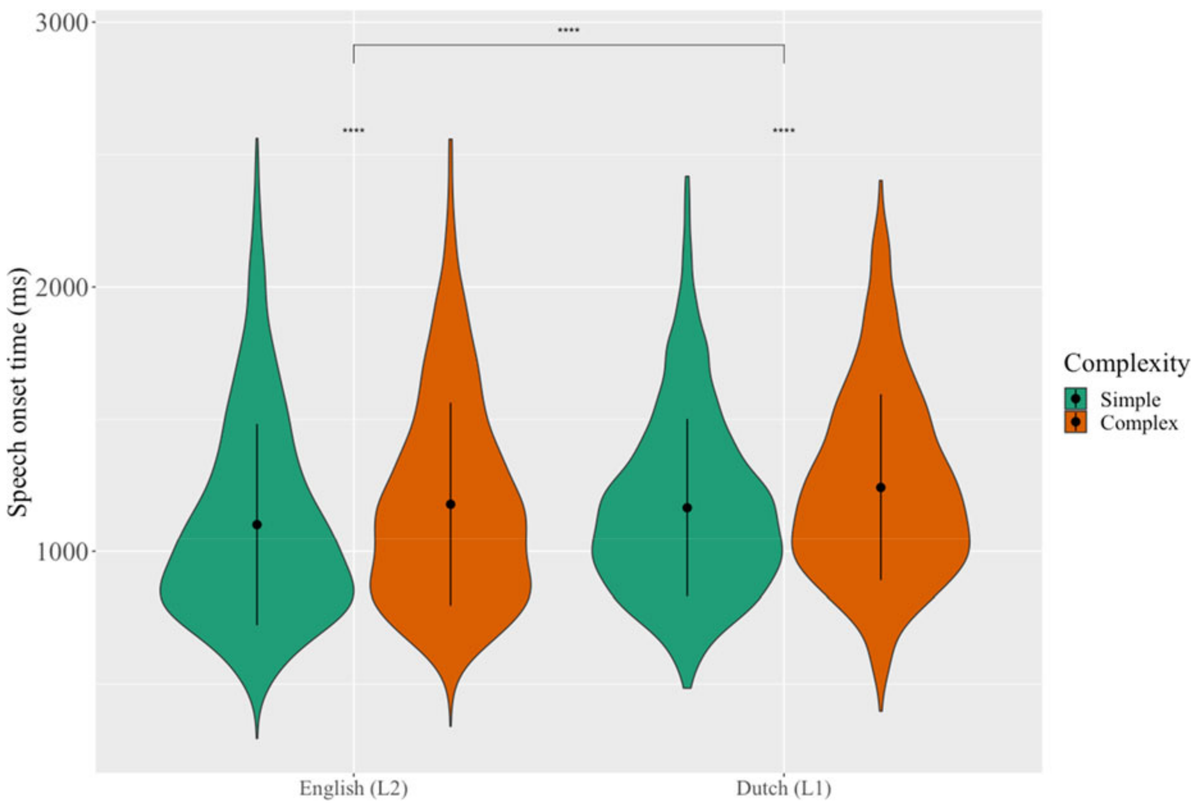

Fig. 2. Estimated speech onset times (ms) and standard deviations for L1 and L2 as a function of complexity.

(n)') in L2 (425 ms) was significantly slower compared to the average articulation duration in L1 (382 ms). The model accounting for individual differences in L2 proficiency did not show an effect of L2 proficiency ($\chi^2(1) = 3.19$, $p = .07$).

Articulation duration of the second noun

There was a main effect of language ($\chi^2(1) = 82.06$, $p < .001$), indicating that the average articulation duration of the second noun in L2 (470 ms) was significantly slower compared to the average articulation duration in L1 (364 ms). Again, the model accounting for individual differences in L2 proficiency did not show an effect of L2 proficiency ($\chi^2(1) = 1.39$, $p = .24$).

Before and during speech

The results presented above have shown that Dutch–English bilinguals started speech sooner in their L2 than in their L1, but the articulation durations during speech were extended in L2 as compared to L1. As a next step, we fitted a series of linear mixed-effects models to further explore the similarities and differences in phrasal processing in L1 and L2.

First, we compared the total duration from the onset of the animation (i.e., speech onset times) until the offset of the first noun (i.e., articulation duration of the first noun). This model contained fixed effects for complexity and language, including an interaction term, and by-subject and by-item random intercepts and by-subject random slopes for language and complexity.

Combining the speech onset times and articulation durations of the first noun revealed a main effect for complexity ($\chi^2(1) = 30.69$, $p < .001$), but no main effect for language ($\chi^2(1) = 0.37$, $p = .54$), or an interaction effect between language and complexity ($\chi^2(1) = 0.88$, $p = .35$). Thus, even though speech onset latencies were faster in L2 than in L1 and the articulation durations of the first noun were longer in L2 than L1, there was no longer a language difference when combining the two measures, but the overall complexity effect remained robust (see Table 3).

https://doi.org/10.1017/S1366728923000068 Published online by Cambridge University Press

Table 2. Estimated mean articulation durations in milliseconds, standard errors, and 95% confidence intervals (CI) of the first noun as a function of sentence type for both L1 and L2.

	Language					
	L1			*L2*		
Sentence type	Mean	SE	95% CI	Mean	SE	95% CI
Simple-Complex	352	12	327 – 375	419	13	393 – 445
Complex-Simple	353	13	328 – 378	395	13	370 – 420

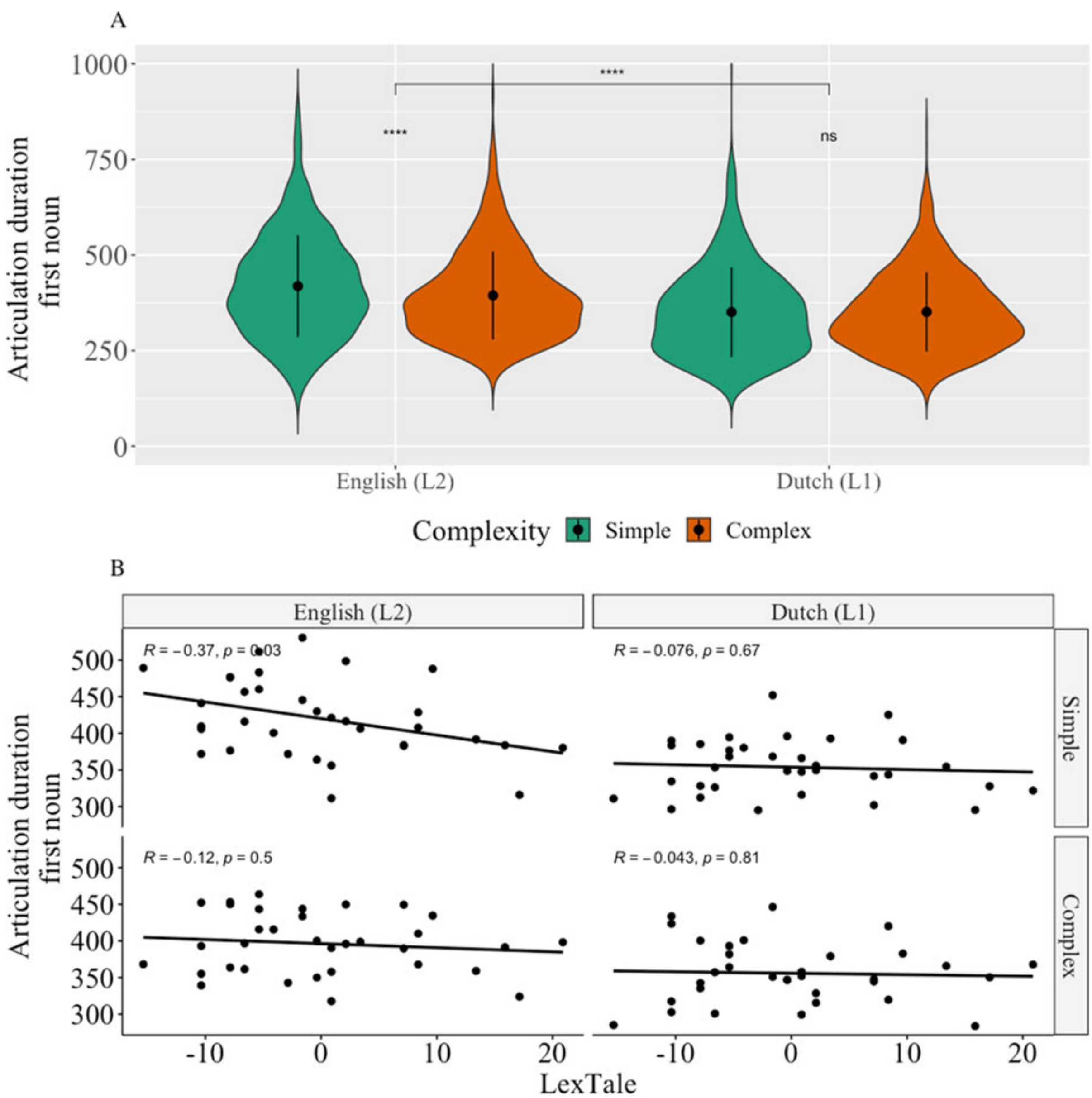

Fig. 3. A: Estimated articulation duration of the first noun and standard deviations for L1 and L2 as a function of complexity. **B:** Visualization of the effect of L2 proficiency on the average articulation duration of the first noun in L1 and L2 for simple-complex and complex-simple sentences.

Then the total duration from the onset of the animation until the onset of the second noun was compared. As only complex-simple sentences were included, this model contained language as a fixed effect. The random effect structure was the same as the previous model, but without the by-subject random slopes for complexity. There was again no main effect for language ($\chi^2(1) = 0.58$, $p = .45$), indicating that the total duration from the onset of the animation until the onset of the second noun

Table 3. Estimated mean durations in milliseconds, standard errors, and 95% confidence intervals (CI) of the speech onset time and articulation duration of the first noun combined as a function of sentence type for both L1 and L2.

	Language					
	L1			*L2*		
Sentence type	Mean	SE	95% CI	Mean	SE	95% CI
Simple-Complex	1532	36	1463 – 1602	1527	37	1455 – 1560
Complex-Simple	1600	39	1525 – 1675	1578	41	1497 – 1659

in L2 (2003 ms) was not significantly slower compared to the average duration in L1 (1980 ms).

Keeping everything in the prior model equal but changing the dependent variable, we compared the total duration from the onset of the animation until the offset of the second noun. Interestingly, this time there was a main effect for language ($\chi^2(1) = 13.65$, $p < .001$), with longer durations in L2 (2469 ms) compared to L1 (2340 ms).

General discussion

The main goal of our study was to test whether sentence planning differs between L1 and L2 of proficient but unbalanced bilinguals. It has previously been shown that for sentences with complex initial phrases, the speech onset starts later than for sentences with simple initial phrases, indicating a phrasal planning scope (e.g., Allum & Wheeldon, 2007, 2009; Martin et al., 2010; Sadat et al., 2012; Wagner et al., 2010). It has also been suggested that planning scope can vary as a function of cognitive load (e.g., Oppermann et al., 2010; Wagner et al., 2010). As processing in L2 might be cognitively more exhausting than processing in L1 for a variety of reasons, we sought to determine whether the size of planned chunks varies for Dutch–English bilinguals as a function of language. We measured speech onset times and articulatory durations in complex-simple and simple-complex sentence production in L1 and L2. Our main findings are summarized below.

We found a complexity effect: onset latencies were significantly longer for complex-simple sentences than for simple-complex sentences. This complexity effect observed in previous studies (e.g., Martin et al., 2010; Smith & Wheeldon, 1999;) was successfully replicated in both Dutch and English, with the effect being comparable in size to the ones reported in earlier research. Speech onset latencies were shorter in L2 than in L1.

Additionally, we measured the articulation duration of the first noun and found that articulatory durations of the first noun in simple-complex and complex-simple sentences were extended in L2 as compared to L1. Interestingly, the extended articulation durations of the first noun were further exaggerated in simple-complex sentences in L2, especially for those with lower L2 proficiency. The robust complexity effect and the exaggerated articulation durations for the first noun in simple-complex sentence in L2 could be indicative of a phrasal scope of planning, with exaggerated articulation durations as the following (complex) phrase needs to be processed. Moreover, in complex-simple sentences, durations between the offset of the first noun and the onset of the second noun as well as the articulation duration of the second noun were significantly extended in L2.

Although the articulation durations were generally slower in L2, speech onset times were found to be shorter. When considering the speech onset times and articulation durations together, no differences between L1 and L2 were any longer present until the onset of the second noun. As the complexity effect was robust, indicating a phrasal scope of planning on the level of grammatical encoding, the differences in speech onset times and articulation durations suggest that bilinguals may plan ahead on different levels of encoding before production. The fact that the main effect of language returned when including the articulation duration of the second noun indicates that the articulation duration of the second noun was not only slower in L2 as compared to L1, but also exaggerated. This pattern aligns with what was observed for the first noun in simple-complex sentences in L2, supporting the presumption that such exaggerated articulation durations indicate processing of the following phrase. The degree to which such strategies were applied appeared to interact with one's L2 proficiency level as extended articulation durations were found especially for those with a lower L2 proficiency level.

Speech onset times and articulation durations

As pointed out earlier, in order to acquire a more complete picture of speech planning and production, behavioral measures, such as speech onset latencies and articulation durations, should be examined collectively in order to expose any subtle differences in strategies. Based on results from earlier studies, we predicted a difference in the size of planned chunks across languages. We argued that the effect of cognitive load could result in two different patterns: a planning scope reduction, or a seemingly similar planning scope but with more planning taking place during articulation.

The complexity effect on speech onset times was of equal magnitude in L1 and L2 production, indicating a phrasal scope of planning on the level of grammatical encoding. However, we found evidence supporting the idea of planning taking place during articulation, as articulation durations were significantly extended in L2, but the speech onset times were shorter. It seems that the effect of cognitive load is thus not necessarily directly reflected in speech onset latencies but could come to light in articulation durations. Three potential explanations could be forwarded for these findings.

The first has to do with the L2 proficiency level. The participants in our study were already proficient in English, which could have masked any between- and within-speaker differences at earlier processing stages. It is possible that others who are less proficient in their second language do not only plan utterances differently during later processing stages, after articulation has already begun, but also at processing stages preceding

https://doi.org/10.1017/S1366728923000068 Published online by Cambridge University Press

articulation. In their study, Gilbert et al. (2020) found that those who were subject to a recent increase in L2 exposure took the second phrase into account (i.e., longer scope of planning as compared to L1) during the initial planning stage, indicated by longer speech onset times as a function of word length (measured in number of syllables) of the second phrase. In this group they indeed observed longer speech onset times for multi-phrase utterances for those with lower L2 exposure. However, those who were relatively consistently exposed to their L2 (comparable to the sample used in our current study) tended to only take the first phrase into account (i.e., same scope of planning as L1) during the initial planning stage, indicated by longer speech onset times as a function of word length of the first phrase. In this group no differences were found when controlling for differences in current L2 exposure.

The second explanation concerns the possible effect of visual grouping. Prior research on visual grouping has shown that recognizing an object and naming it is more difficult when that specific object is grouped together with other objects (Zhao, Alario, & Yang, 2015). As for the sentences used in this study, the first two objects were moving together in the complex-simple sentences, whereas the first object was moving alone in simple-complex sentences. Results from a study by Zhao, Paterson, and Bai (2018) revealed an interplay between visual grouping and utterance planning, indicating that speakers use multiple cues to plan utterances. One possibility is then that the extended articulation durations are the result of visual encoding of the moving scene. However, as longer articulation durations were also observed in simple-complex sentences, this explanation seems unlikely. Moreover, data by Martin et al. (2010) confirm that complexity effects are not solely due to grouping in dynamic scenes, as identical patterns were also observed when stationary displays were used instead.

The third explanation is perhaps the most interesting; when using L2, the speech onset of noun phrases starts before the noun phrase is fully encoded on, for instance, the phonological level. This strategy allows for speech onset latencies that are relatively shorter but at the same time result in longer articulatory durations as planning and articulation happen simultaneously. This latter explanation would be in line with our second prediction that increased cognitive load could result in a seemingly similar planning scope, while observing speech patterns that suggest that planning takes place partly during articulation.

As others have indeed suggested (e.g., Ferreira & Swets, 2002; Wagner et al., 2010), speakers may implicitly use strategies that alter the scope of speech planning. A situation in which someone is more likely to alter the planning scope of speech is when lexical access is constrained. For instance, when bilinguals are interacting in their non-dominant language, they may experience more difficulties with accessing lexical items than when they are communicating in their dominant language. It seems plausible that they would narrow their speech planning scope when using their L2 by starting speech production while still retrieving and planning subsequent lexical items in order to match their speech production in L1 in both rate and fluency.

Indeed, results from the current study demonstrate that the articulatory durations tend to be longer in L2 speech production compared to those in L1 speech production, whereas the speech onset time was shorter in L2 compared to that in L1. Moreover, the articulatory durations of the first noun in simple-complex sentences and the second noun in complex-simple sentences were further extended in L2, as the following phrase had to be processed. These patterns indicate that difficulties in non-native language processing and production, often linked to enhanced cognitive load, come to the surface in the form of extended articulatory durations rather than a reduced planning scope.

As a robust complexity effect was observed, it is possible that the overall timing differences in speech onset and articulation duration between L1 and L2 are reflective of differences in processing on the phonological level, with exaggerated articulation durations of the first and second noun in L2 as the following phrase needs to be encoded on the grammatical level as well. However, without the use of online methods or the implementation of distractor words, it is impossible to pinpoint exactly on which level(s) L1 and L2 sentence processing differs. Additional research is needed to confirm this hypothesis.

Proficiency and language experience

Bilinguals with higher L2 LexTALE scores showed slightly shorter overall articulation durations in L2 than bilinguals with lower scores. Thus, although the participants in our study were generally highly proficient in English, those with higher L2 proficiency scores appear to show more native-like patterns of planning during speech in L2 than those with lower proficiency scores. Yet, there was no significant effect of proficiency on the speech onset times. More research is needed, using a less homogenous sample, to further investigate the role of proficiency on the scope of planning.

Moreover, because experiences with linguistic representations are relatively fewer in either language (as bilinguals can only use one of their languages at a time) compared to that of the language that is exclusively used by monolinguals, the links between phonology, the lexicon, and semantics may be weaker in bilinguals compared to monolinguals (Gollan, Montoya, Cera, & Sandoval, 2008; Yan & Nicoladis, 2009). Future research should therefore also consider a direct comparison between monolinguals and bilinguals.

Conclusion

On the surface, bilinguals do not have a smaller planning scope in L2. Our data add to the existing literature as they replicated the complexity effect previously observed in monolingual speech planning in bilinguals' L1 and L2. However, bilinguals use strategies (i.e., extended articulation duration) to extend processing time in their L2. Between speaker comparisons suggest that the use of such strategies decreases as the L2 proficiency level increases.

Acknowledgments. This paper was supported by a Concerted Research Action (BOF13/GOA/032) from the Special Research Fund, Ghent University.

Data availability statement. Data availability: The data that support the findings of this study are openly available in Open Science Framework at https://osf.io/ws7vx/.

Competing interests declaration. Competing interests: The author(s) declare none

References

Allum PH and Wheeldon L (2007) Planning scope in spoken sentence production: The role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **33**, 791.

Allum PH and Wheeldon L (2009) Scope of lexical access in spoken sentence production: Implications for the conceptual–syntactic interface. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **35**, 1240.

Bates D, Maechler M, Bolker B and Walker S (2015) Fitting Linear Mixed Effects Models Using lme4. *Journal of Statistical Software*, **67**, 1–48.

Bishop J and Intlekofer D (2020) Lower working memory capacity is associated with shorter prosodic phrases: implications for speech production planning. In *Proceedings of 10th International Conference on Speech Prosody* (Vol. 39).

Bock K and Levelt WJ (1994) Language production: Grammatical encoding. Handbook of psycholinguistics. ed. by Morton A. Gernsbacher, 945–984.

Boersma P and Weenink D (2018) Praat: doing phonetics by computer (Version 6.0.37) [Computer program]. Retrieved April, 2018.

Bonin P, Malardier N, Méot A and Fayol M (2006) The scope of advance planning in written picture naming. *Language and Cognitive Processes*, **21**, 205–237.

Broos WP, Bencivenni A, Duyck W and Hartsuiker RJ (2021) Delayed picture naming in the first and second language. *Bilingualism: Language and Cognition*, **24**, 389–400.

Broos WPJ, Duyck W and Hartsuiker RJ (2018) Are higher-level processes delayed in second language word production? Evidence from picture naming and phoneme monitoring. *Language, Cognition and Neuroscience*, **33**, 1219–1234.

Brown-Schmidt S and Konopka A (2008) Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, **109**, 274–280.

Cop U, Keuleers E, Drieghe D and Duyck W (2015) Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, **22**, 1216–1234.

Dell GS, Chang F and Griffin ZM (1999) Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science*, **23**, 517–542.

Dryer MS (2005) Order of subject, object, and verb. In Haspelmath M, Dryer MS, Gil D and Comrie B (eds), *The world atlas of language structures*. Oxford: Oxford University Press, pp 330–333.

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.

Gilbert AC, Cousineau-Perusse M and Titone D (2020) L2 exposure modulates the scope of planning during first and second language production. *Bilingualism: Language and Cognition*, **23**, 1093–1105.

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.

Griffin ZM (2003) A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review*, **10**, 603–609.

Griffin ZM and Bock K (2000) What the eyes say about speaking. *Psychological Science*, **11**, 274–279.

Hanulová J, Davidson DJ and Indefrey P (2011) Where does the delay in L2 picture naming come from? Psycholinguistic and neurocognitive evidence on second language word production. *Language and Cognitive Processes*, **26**, 902–934.

Ivanova I and Costa A (2008) Does bilingualism hamper lexical access in speech production?. *Acta Psychologica*, **127**, 277–288.

Klaus J, Mädebach A, Oppermann F and Jescheniak JD (2017) Planning sentences while doing other things at the same time: effects of concurrent verbal and visuospatial working memory load. *The Quarterly Journal of Experimental Psychology*, **70**, 811–831.

Klaus J and Schriefers H (2018) An investigation of the role of working memory capacity and naming speed in phonological advance planning in language production. *The Mental Lexicon*, **13**, 159–185.

Konopka A (2012) Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, **66**, 143–162.

Konopka A and Meyer AS (2014) Priming sentence planning. *Cognitive Psychology*, **73**, 1–40.

Konopka A, Meyer AS and Forest TA (2018) Planning to speak in L1 and L2. *Cognitive Psychology*, **102**, 72–104.

Krivokapic J (2012) Prosodic planning in speech production. *Speech Planning and Dynamics*, 157–190

Kuznetsova A, Brockhoff PB and Christensen RHB (2017) "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software*, **82**, 1–26.

Lagrou E, Hartsuiker RJ and Duyck W (2011) Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **37**, 952.

Lee EK, Brown-Schmidt S and Watson DG (2013) Ways of looking ahead: Hierarchical planning in language production. *Cognition*, **129**, 544–562

Lemhöfer K and Broersma M (2012) Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods*, **44**, 325–343.

Lenth RV (2021) Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, **69**, 1–33.

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

Li C, Ferreira VS and Gollan TH (2022) Language control after phrasal planning: Playing Whack-a-mole with language switch costs. *Journal of Memory and Language*, **126**, 104338.

Martin RC, Crowther JE, Knight M, Tamborello II, FP and Yang CL (2010) Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition*, **116**, 177–192.

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.

Meyer AS (1996) Lexical access in phrase and sentence production: Results from picture word interference experiments. *Journal of Memory & Language*, **35**, 477–496.

Meyer AS, Sleiderink AM and Levelt WJ (1998) Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, **66**, 25–33.

Mortensen L, Meyer AS and Humphreys GW (2008) Speech planning during multiple object naming: Effects of ageing. *Quarterly Journal of Experimental Psychology*, **61**, 1217–1238.

Myachykov A, Scheepers C, Garrod S, Thompson D and Fedorova O (2013) Syntactic flexibility and competition in sentence production: The case of English and Russian. *Quarterly Journal of Experimental Psychology*, **66**, 1601–1619.

Oppermann F, Jescheniak JD and Schriefers H (2010) Phonological advance planning in sentence production. *Journal of Memory and Language*, **63**, 526–540.

Runnqvist E, Strijkers K, Sadat J and Costa A (2011) On the temporal and functional origin of L2 disadvantages in speech production: A critical review. *Frontiers in Psychology*, **2**, 379–387.

Sadat J, Martin CD, Alario FX and Costa A (2012) Characterizing the bilingual disadvantage in noun phrase production. *Journal of Psycholinguistic Research*, **41**, 159–179.

Schneider W, Eschman A and Zuccolotto A (2012) E-Prime 2.0 Professional. Pittsburgh, PA: Psychology Software Tools.

Service E, Simola M, Metsänheimo O and Maury S (2002) Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, **14**, 383–408

Severens E, Van Lommel S, Ratinckx E and Hartsuiker RJ (2005) Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica*, **119**, 159–187.

Smith M and Wheeldon L (1999) High level processing scope in spoken sentence production. *Cognition*, **73**, 205–246.

Swets B, Fuchs S, Krivokapić J and Petrone C (2021) A cross-linguistic study of individual differences in speech planning. *Frontiers in Psychology*, **12**, 1439.

Swets B, Jacovina ME and Gerrig RJ (2013) Effects of conversational pressures on speech planning. *Discourse Processes*, **50**, 23–51.

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.

Team RC (2021) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. R Foundation for Statistical Computing.

https://doi.org/10.1017/S1366728923000068 Published online by Cambridge University Press

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 Experiment Psychology: Learning, Memory, and Cognition*, **36**, 423–440.
Yan S and Nicoladis E (2009) Finding le mot juste: Differences between bilingual and monolingual children's lexical access in comprehension and production. *Bilingualism: Language and Cognition*, **12**, 323–335.
Zhao LM, Alario FX and Yang YF (2015) Grammatical planning scope in sentence production: Further evidence for the functional phrase hypothesis. *Applied Psycholinguistics*, **36**, 1059–1075.
Zhao LM and Yang YF (2016) Lexical planning in sentence production is highly incremental: Evidence from ERPs. *PloS one*, **11**, e0146359.
Zhao L, Paterson KB and Bai X (2018) Visual Grouping in Accordance with Utterance Planning Facilitates Speech Production. *Frontiers in Psychology*, **9**, 307.