



Cognitive Science 40 (2016) 1911–1940

Copyright © 2015 Cognitive Science Society, Inc. All rights reserved.

ISSN: 0364-0213 print / 1551-6709 online

DOI: 10.1111/cogs.12315

Breaking Down the Bilingual Cost in Speech Production

Jasmin Sadat,^{a,b} Clara D. Martin,^{c,d} James S. Magnuson,^{e,f}
François-Xavier Alario,^b Albert Costa^{a,g}

^a*Department of Information and Communications Technologies, Pompeu Fabra University*

^b*Laboratory of Cognitive Psychology, CNRS and Aix-Marseille University*

^c*Basque Center on Cognition, Brain and Language*

^d*IKERBASQUE, Basque Foundation for Science*

^e*Department of Psychology, University of Connecticut*

^f*Haskins Laboratories*

^g*Catalan Institution for Research and Advanced Studies, (ICREA)*

Received 29 January 2015; received in revised form 26 June 2015; accepted 19 August 2015

Abstract

Bilinguals have been shown to perform worse than monolinguals in a variety of verbal tasks. This study investigated this bilingual verbal cost in a large-scale picture-naming study conducted in Spanish. We explored how individual characteristics of the participants and the linguistic properties of the words being spoken influence this performance cost. In particular, we focused on the contributions of lexical frequency and phonological similarity across translations. The naming performance of Spanish-Catalan bilinguals speaking in their dominant and non-dominant language was compared to that of Spanish monolinguals. Single trial naming latencies were analyzed by means of linear mixed models accounting for individual effects at the participant and item level. While decreasing lexical frequency was shown to increase naming latencies in all groups, this variable by itself did not account for the bilingual cost. In turn, our results showed that the bilingual cost disappeared when naming words with high phonological similarity across translations. In short, our results show that frequency of use can play a role in the emergence of the bilingual cost, but that phonological similarity across translations should be regarded as one of the most important variables that determine the bilingual cost in speech production. Low phonological similarity across translations yields worse performance in bilinguals and promotes the bilingual cost in naming performance. The implications of our results for the effect of phonological similarity across translations within the bilingual speech production system are discussed.

Keywords: Speech production; Bilingualism; Phonological similarity; Cognates; Lexical frequency; Individual effects; Bilingual disadvantage

Correspondence should be sent to Jasmin Sadat, Laboratoire Parole et Langage, Aix-Marseille Université-CNRS UMR 7309, 5 avenue Pasteur—BP 80975, 13604 Aix-en-Provence Cedex 1, France. E-mail: jasmin.sadat@upf.edu

1. Introduction

Being able to communicate in two languages is an obvious asset, but being bilingual also has some negative consequences on linguistic performance. Compared to monolingual speakers, bilinguals show slower naming latencies, take longer to articulate, make more errors, and experience more tip-of-the-tongue states (e.g., Gollan, Fennema-Notestine, Montoya, & Jernigan, 2007; Gollan & Goldrick, 2012; Gollan, Montoya, Cera, & Sandoval, 2008; Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Gollan, Montoya, & Werner, 2002; Gollan & Silverberg, 2001; Guion, Flege, Liu, & Yeni-Komshian, 2000; Ivanova & Costa, 2008; Kohnert, Hernandez, & Bates, 1998; Mackay & Flege, 2004; Roberts, Garcia, Desrochers, & Hernandez, 2002; Sadat, Martin, Alario, & Costa, 2012). Remarkably, these costs surface not only in a bilingual's non-dominant language (L2) but even in his or her first learned and dominant language (L1; Ivanova & Costa, 2008; Sadat et al., 2012). Thus, in this study we will refer to the bilingual cost as the performance difference between monolinguals and bilinguals, speaking either in L1 or L2.

Since bilingualism or multilingualism is becoming increasingly common, it is important to understand the linguistic cost that is associated with it. Despite a growing body of research, our knowledge about the origin of this phenomenon remains rather limited (for a review see Runnqvist, Strijkers, Sadat, & Costa, 2011). Here, we will explore how the linguistic properties of the words being spoken may shed light on this issue. A reasonable starting point would be to assume simply that the main variables found to govern monolingual speech production (e.g., lexical frequency, name agreement, age of acquisition) would also govern bilingual speech production (e.g., Gollan et al., 2008).

In this article, we will explore how important variables of monolingual speech production behave in bilingual speakers. Performance differences between these two groups of speakers could be captured by a detailed characterization of the variables influencing their speech production. In particular, we will focus on lexical frequency and phonological similarity across translations (also known as cognate status). While these two variables are among the most studied variables in the context of bilingual speech processing (e.g., Costa, Caramazza, & Sebastián-Gallés, 2000; Costa, Santesteban, & Caño, 2005; Dijkstra, Grainger, & Van Heuven, 1999; Gollan et al., 2005), it remains to be established firmly whether and how they influence the bilingual cost in speech production (e.g., Bialystok, Craik, Green, & Gollan, 2009).

In what follows we outline the three primary accounts that have been proposed to account for a bilingual cost in speech production and the variables that are suggested to contribute to it. We will then describe the approach of this study, in which we will directly test the relationship between the phenomenon of the bilingual cost and the most important variables associated with it. Our final goal is to establish the relative contributions of each of these variables that have been put forward to influence monolingual speech production and to determine how they impact bilinguals.

1.1. The bilingual cost as a frequency effect

The first account of the bilingual cost that we will consider builds on the pervasive observation of frequency effects in language production (e.g., Oldfield & Wingfield, 1964, 1965). This account assumes that higher frequency use strengthens the links between a concept and its lexical representation, which in turn leads to faster word retrieval. Since bilinguals use each of their two languages less frequently than monolinguals use their only language, bilinguals will have weaker links in each of their languages (“weaker links hypothesis,” Gollan et al., 2008). As a result of this frequency lag, and everything else being equal, bilinguals will be slower in retrieving a word for production relative to monolinguals. The weaker links hypothesis makes other explicit predictions which pertain to modulations of the bilingual cost and the frequency effect. This account predicts that low-frequency words suffer most from reduced use. According to Gollan et al. (2011, p. 4), “the bilingual disadvantage is especially large for retrieval of low-frequency words, whereas little or no bilingual disadvantage is found for production of high-frequency words.” As a direct consequence, the size of the frequency effect should equally depend on language use. Decreasing language use would lead to an increase in the frequency effect in speech production of that language. By and large, then the magnitude of the frequency effect should increase from monolingual speakers to bilinguals speaking in L1 to bilinguals speaking in L2 (i.e., a similar pattern as in naming latencies; we will come back to these predictions in the General Discussion). In summary, this framework yields two main predictions regarding the way lexical frequency may impact the bilingual cost: (a) the bilingual cost is expected to be substantially reduced (to the point of disappearing) in the case of high-frequency words, and (b) the magnitude of the lexical frequency effect should increase with reduced language use, from monolinguals to bilinguals speaking in L1 to bilinguals speaking in L2.

1.2. The bilingual cost at post-lexical processing levels

The second explanation for the bilingual cost that we will consider claims that delays in bilingual language processing stem from processing stages subsequent to lexical access (Hanulová, Davidson, & Indefrey, 2011; Indefrey, 2006). Indefrey and colleagues reviewed studies comparing native and non-native speech processing on the basis of electrophysiological and hemodynamic evidence. These studies suggest that speaker group differences emerge at very late time points of speech processing, which they interpret as corresponding to phonological and articulatory processing stages. Therefore, the authors propose that the bilingual cost may stem from particularly demanding processes at the level of phonological and phonetic encoding, syllabification, and/or articulation. While this account points to a particular locus, it is less clear about the mechanisms underlying it. Thus, some additional assumptions are required to establish specific predictions from it. A possible approach to this issue would be to rely on variables reflecting processing from phonological or articulatory levels, although it may be hard to specify variables that exclusively index processing at post-lexical stages. Variables thought to influence late

stages of production such as frequency most likely also affect speech processing at earlier levels (e.g., Barry, Hirsh, Johnston, & Williams, 2001; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Gahl, 2008). Below, we shall tentatively relate the post-lexical processes considered in this account with the effects of phonological similarity within and across languages (i.e., phonological neighborhood in the spoken and unspoken language, and cognate status).

1.3. The bilingual cost as cross-language competition and control

Finally, an alternative explanation postulates that part of the bilingual cost originates from language control processes (executive control account; Abutalebi & Green, 2008; Green, 1998). When speaking, bilinguals constantly need to resolve the potential competition between their two language systems. This extra-requirement is thought to slow down bilingual lexical access compared to monolinguals, irrespective of the representational level(s) at which this competition process may occur. Recent bilingual research has linked the ability to resolve conflict between languages to a more general type of conflict resolution. It may be that a bilingual's ability of resolving non-verbal conflict relates to the extent to which he or she performs poorer on verbal tasks (e.g., Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok, Craik, & Luk, 2008; Costa, Hernandez, & Sebastián-Gallés, 2008; for a review see Hilchey & Klein, 2011; but see Alario, Ziegler, Massol, & De Cara, 2012; De Bruin, Treccani, & Della Sala, 2015; Paap & Greenberg, 2013). Under this assumption, bilingual speakers who experience less conflict in non-linguistic tasks may also be better at resolving conflict between the two language systems and thus should show less of a bilingual cost. It is thus expected that partialing out the general ability to resolve conflict from verbal performance measures would capture the part of the bilingual cost that is linked to general conflict resolution abilities. Note that this measure mainly characterizes properties related to the participants (e.g., proficiency and relative language use) rather than the linguistic material per se.¹ Although the focus of this study is to explore the influence of word-related properties, we nevertheless considered measures thought to capture individual conflict-solving performance as a covariate because of the importance this dimension has in the recent literature.

1.4. Operationalization of the relative contributions to the bilingual cost

The concurrent manipulation of different psycholinguistic variables was used to gauge the relative contributions of each of the variables that have been put forward to explain performance differences between mono- and bilingual speakers. The general rationale of our analysis approach is as follows: The bilingual cost can be observed in a plain task such as picture naming. If (part of) the bilingual cost is driven by a certain variable, then the inclusion of this variable in the analysis of performance should reduce or dissipate the bilingual cost. Note that for this rationale to work, the effect of the contributing variable should be estimated as a cost, just as bilingual cost effects are estimated as costs relative to monolingual performance (the details of this procedure are explained in the

Methods section under Data analysis). For example, the contribution of lexical frequency of the words should be assessed relative to the highest value of lexical frequency values (i.e., frequency cost) by inverting the measure of lexical frequency from highest to lowest. Note that this is simply a linear (convenience) transformation providing a mirror image that does not affect the patterns examined. If the effect of the bilingual cost is no longer present when such a predictor is included in a model, then this variable could be said to explain the performance differences between these two groups of speakers. If there is more than one contributing variable, the bilingual cost can be said to be broken down into more elementary contributions. The relationship between such contributing variable(s) and specific processing levels can then be discussed to understand their impact on bilingual speech production relative to monolinguals.

Lexical frequency is perhaps the variable that has most explicitly been related to differences in speech performance between mono- and bilinguals (e.g., Gollan et al., 2008, 2011). In our experiment, we will explore this link at a finer granularity level than it has been done previously by estimating the magnitude of frequency effects for every participant individually. This approach is motivated, among others, by one possible confound in the accounts in which the bilingual cost is interpreted as a frequency effect. Increased frequency effects are not only observed in participants with decreased language use, but also in general with slower reaction times (e.g., Balota & Ferraro, 1993, 1996; Cerella, 1985; Spieler & Balota, 2000, for showing a larger frequency effect in slower response times of older adults; but see also Baayen & Milin, 2010, for larger frequency effects in slower responding participants when age is controlled for). It may be that larger frequency effects in bilinguals compared to monolinguals are due to the mere fact that bilinguals show slower naming times (maybe for reasons unrelated to frequency), and not as a direct consequence of their experience with each of their languages (see also Gollan et al., 2008, for similar frequency effects in younger and older speakers after adjusting for age-related slowing). Below, the individual frequency effects of the participants in our study will also be assessed in relationship with their overall speed of response.

A second item-related variable that is specific to bilingual speech production and that has been shown to be very influential is cognate status. This variable can be tentatively related to the second, post-lexical, account of the bilingual cost. Cognate words are translations sharing high phonological similarity across languages like, for example, “tomato” (English) and “tomate” (Spanish), as compared to words that do not share many sounds (“pumpkin” [English] and “calabaza” [Spanish]). Several studies have reported that cognate words show a processing benefit over non-cognates in bilingual production (i.e., cognate facilitation effect; see Costa et al., 2000). This effect is most often attributed to post-lexical (e.g., phonological) and lexical processes (see Costa et al., 2005, for detailed discussion and caveats). While the amount of phonological similarity across translations (henceforth: translation similarity) is a highly influential variable in bilingual speech production, its effect has been largely neglected in the context of explaining the bilingual cost (but see Gollan & Acenas, 2004, demonstration that bilinguals experience more tip-of-the-tongue states than monolinguals unless the target word was a cognate for which performances were similar). Until now there has been no detailed description of how the

bilingual cost may vary as a function of translation similarity of the to-be-produced word. In the context of the post-lexical account, we will examine the relationship between the bilingual cost and translation similarity. Just as for lexical frequency, the contribution of this variable will also be explored at a finer granularity level than most previously published studies. In addition, we will also consider the possible influence of phonological similarity beyond translations, relative to other words in the mental lexicon (i.e., phonological neighborhood). Given the strong cognate effect observed in bilingual speech and the common assumption that during production the two languages are co-activated at the phonological level (Colomé, 2001; Colomé & Miozzo, 2010; Costa et al., 2000), one may suppose that phonologically similar words in the unspoken language (other than the translation itself) may also influence the production of the intended word in bilinguals (just as phonological neighbors influence monolingual speech production; e.g., Dell & Gordon, 2003; Sadat, Martin, Costa, & Alario, 2014).

Finally, we will examine whether the participant-related variable of general conflict resolution ability is related to the bilingual cost in speech production as claimed in previous studies (e.g., Bialystok et al., 2008). We will do so by measuring participants' performance in a classical Simon task (e.g., Bialystok et al., 2004). In this kind of task, the spatial compatibility between a presented stimulus and the side of the manual response are manipulated. Participants are asked to respond to a series of congruent, incongruent, and neutral trials, and the difference in response times between incongruent and congruent trials is thought to reflect general conflict resolution abilities (i.e., Simon effect; Simon & Wolf, 1963). This participant-related measure will be introduced as covariate into our analysis to potentially explain differences across speakers. In doing so, we will be able to explore whether there is a relationship between non-linguistic conflict resolution abilities and the potential cross-language conflict experienced by bilinguals.

1.5. This study

We tested 60 Spanish-Catalan bilinguals (30 L1 Spanish and 30 L2 Spanish speakers) in a large-scale picture-naming experiment and compared their performance to a group of 30 Spanish monolingual speakers (previously reported in Sadat et al., 2014). As in previous studies that investigated the bilingual cost (Gollan, Montoya, & Bonanni, 2005; Gollan, Montoya, Fennema-Notestine, et al., 2005; Gollan et al., 2008; Gollan et al., 2011; Ivanova & Costa, 2008; Sadat et al., 2012), our rationale was to compare the performance of the groups of bilingual speakers to a group of monolingual speakers.

The bilingual population under study here was exposed to their two languages from early on and has high levels of proficiency in both of them (equivalent to monolinguals). Thus, the monolingual and the two bilingual groups differ only in the amount of time they spend using the respective language of comparison (Spanish in this study). This is important since the weaker links hypothesis capitalizes on frequency of language use as the crucial predictor to explaining the bilingual cost in verbal performance. By selecting a group of bilingual speakers that was equally proficient in Spanish as their monolingual counterparts, we could directly address the predictions of the weaker links hypothesis.

Since this study used the monolingual group described in Sadat et al. (2014) as a control group, we used identical materials, design, and procedure as in that study. We included the most relevant item-related predictor variables to evaluate how each of these variables influences mono- and bilingual speech production. Given that we mainly focused on variables that govern the production facility of specific words (i.e., item-related variables), we tested a large set of 533 words to be produced twice, in two different runs. This allowed us to have a large number of observations per participant ($N = 1,066$) and speaker group ($N = 31,980$) and to go beyond dichotomous measures, accounting for effects over a continuous range of values. In addition, we characterized item-related variables, such as lexical frequency and translation similarity, as continuous measures to account for naming performance as a more detailed function of lexical characteristics. Appropriately assessing translation similarity as a continuous measure has been something of a challenge in bilingualism research. Most previous studies investigated the influence of this variable as a simple dichotomous measure (i.e., cognates versus non-cognates; but see Schepens, Dijkstra, & Grootjen, 2012; Wieling, Margaretha, & Nerbonne, 2012, for proposing Levenshtein edit distance as a continuous measure for phonological similarity), as a continuous measure based on orthographic similarity (Schwartz, Kroll, & Diaz, 2007; Van Orden, 1987),² or as a subjective measure by asking participants for translation similarity ratings (De Groot & Nas, 1991; Dijkstra et al., 1999; Friel & Kennison, 2001). In this study, we use a continuous measure that captures translation similarity in an even more fine-grained and phonetically more appropriate manner than previous studies (see below for a description of the ALINE measure used in this study). Moreover, we also included several participant-related predictors, such as individual measures of executive control (as suggested by the executive control account; Bialystok et al., 2008), vocabulary size (e.g., Bialystok et al., 2008; Luo, Luk, & Bialystok, 2010), and socioeconomic status (e.g., Morton & Harper, 2007), as control variables across mono- and bilingual groups to account for possible performance differences due to these variables.

The naming data of mono- and bilinguals was analyzed by means of linear mixed regression modeling performed at the single-trial level (Baayen, Davidson, & Bates, 2008). This method allows a fine-grained partitioning of variance down to the level of individual participants and items. Therefore, our analysis explored the effects of frequency and translation similarity at the participant and item level, which goes beyond previous descriptions of the bilingual cost. The analysis at the individual level will also help clarifying the main contributors to possible differences in the size of lexical frequency and translation similarity effects across participants.

2. Method

2.1. Participants

Sixty Spanish-Catalan bilinguals were recruited. Participants were undergraduate students at the University Pompeu Fabra, Barcelona, Spain. They all used Catalan and

Spanish at a native speaker level; that is, they were highly proficient at speaking, understanding, reading, and writing both of the languages (see Appendix S1 for a description of the bilingual community in Catalonia). Importantly, although being highly proficient, they were all unbalanced speakers with a preferred first learned and dominant language: Spanish was the dominant language for half of them (30 Spanish-dominant bilinguals speaking in L1; henceforth bilinguals in L1; 19 women) and the non-dominant language for the other half (30 Catalan-dominant bilinguals speaking in L2; henceforth bilinguals in L2; 19 women). All bilinguals had acquired their L1 (either Spanish or Catalan) from birth before being exposed to their L2 (i.e., they spoke in L1 with each parent) and reported being more comfortable and spending more time with friends and family in their L1 than L2 (see Table 1 for language history and proficiency ratings). None of the participants reported being able to have a simple conversation in any other language (mean percentage of overall current foreign language use for bilinguals in L1: 3%, $SD = 4$, and bilinguals in L2: 3%, $SD = 6$). All participants had normal or corrected-to-normal vision and were matched on age. They received 20 Euros for participating in the experiment.

The bilingual participants were compared to a group of 30 native Spanish monolingual speakers described in Sadat et al. (2014). Monolinguals were all students from the University of Murcia in Spain, growing up in Spanish-speaking families and using only Spanish for daily communication.

2.2. Materials

The stimuli were 533 black-and-white line drawings of common objects (identical to those used in Sadat et al., 2014; see Appendix S5 for a full list of the stimulus set). Spanish picture names met the following criteria: (a) they consisted of a single word; (b) they were present in the Spanish database BuscaPalabras (Davis & Perea, 2005); (c) they had

Table 1
Means (M) and standard deviations (SD) for participants' self-report ratings

	30 Spanish Monolinguals		30 Spanish-Catalan Bilinguals		30 Catalan-Spanish Bilinguals	
	M	SD	M	SD	M	SD
Age	22	2	21	2	22	3
Percent daily use of Spanish	99	4	63	18	30	19
Age exposed to Spanish	0	0	0	0	1	1
Age exposed to Catalan	–	–	1	2	0	1
Spanish proficiency	4.8	0.5	4.9	0.3	4.9	0.4
Catalan proficiency	–	–	4.6	0.6	5.0	0.2

Note. The measure “Percent daily use of Spanish” was obtained by asking participants to estimate their daily language use of Spanish, Catalan, and any other language with the constraint that their sum equal 100 percent. “Age exposed to Spanish/Catalan” refers to the mean age at which participants were continuously exposed to these languages. Proficiency ratings are on a 1–5 scale, where 1 indicates “very little knowledge of the language” and 5 indicates “native proficiency.” Proficiency values represent the average of the participants' responses in four domains (speech comprehension, speech production, reading, and writing).

no other meanings with higher frequency usage to be confused with (e.g., “tienda,” meaning “tent” or “shop” in English, or “sobre,” meaning “envelope” or “on”/“over”/“about”); (d) they had relatively high name agreement in Spanish (we only used pictures with name agreement values higher than 60% in Cuetos, Ellis, & Alvarez (1999), if available, or assessed through offline pretests with four Spanish monolinguals from among the university staff); (e) translations were not ambiguous in meaning³; and (f) there were no homophones within and across languages. They had black outlines and white surfaces and were presented 300 pixels wide \times 300 pixels high on a white rectangle with a monitor resolution of 800 \times 600 pixels.

For the multiple regression analysis, the following item-related predictors important for speech production were collected from Sadat et al. (2014):

- Word form frequency with values for written lexical frequency (range [0.07, 2.80] log occurrences per million, $M = 0.98$, $SD = 0.54$). A logarithmic transformation was applied to avoid the undue influence of extreme values in the regression.
- Correct target name agreement (range [10, 100] %, $M = 85$, $SD = 18$).
- Subjective estimates of age of acquisition (AoA) from adult ratings (range [2.3, 10.3] years; $M = 4.6$, $SD = 1.4$).
- Phonological neighborhood density in Spanish (PhND; range [0, 37] number of neighbor words, $M = 5$, $SD = 7$). This refers to the number of words in Spanish that can be formed from a given word by substituting, adding, or deleting one phoneme (Luce, 1986).
- Word length measured in phonemes (range [2, 11] number of phonemes, $M = 6$, $SD = 2$).
- First syllable frequency (range [0.08, 4.61] log occurrences per million, $M = 3.33$, $SD = 0.91$). A logarithmic transformation was applied to avoid the undue influence of extreme values in the regression.

In addition, the following item-related predictors specific to bilingual speech production were collected:

- Translation similarity measured by Levenshtein editing distance (range [0.6, 100] %, $M = 26$, $SD = 13$). This calculates how many phonemes of a word have to be changed to transform it into its translation and captures the amount of editing difference between two words (Levenshtein, 1966). This measure was standardized and expressed in percentages.
- Translation similarity measured by ALINE (Kondrak, 2000; range [0, 86] % of similarity, $M = 54$, $SD = 19$). This aligns the phonetic sequences to be compared and assigns similarity values to the common phonemes across translations. To find the best match of strings, it uses the phonetic similarity of surface forms (e.g., “alcachofa” [Spanish for artichoke] needs to be aligned two steps to the right to best match “carxofa” [Catalan]). In addition, ALINE associates different weights to each phoneme pair according to its saliency. This salience constraint leads, for example, to higher weight assignments in the case of identical consonant sounds over identi-

cal vowel sounds (see details in Kondrak, 2000). This measure was standardized and expressed in percentages.⁴

- Phonological neighbors in Catalan (range [1, 126] number of words, $M = 15$, $SD = 17$). This is defined as the number of Catalan lemmas that can be formed for each of the Spanish stimuli words by the substitution, addition, or deletion of a single phoneme at any position within the Spanish word. The number of Catalan lemma neighbors was estimated on the basis of a transcribed corpus of 137,028 Catalan words (Rafel i Fontanals, 1996). Recall that all participants were speaking Spanish throughout the experiment; hence, this variable captures the number of phonological neighbors in Catalan, the unspoken language.

In addition to the item-related predictors, several participant-related predictors were collected (these predictors were also available for the monolingual participants of Sadat et al., 2014):

- Executive control measures and button-press speed. Participants performed an adaptation of the Simon task (Bialystok et al., 2004; Simon & Wolf, 1963). In this task, participants were asked to manually respond as quickly and accurately as possible to the color of a visually presented cue on the screen. The color cue appeared either on the same side as the required response button (congruent trials), in the center of the screen (neutral trials), or on the opposite side as the required response button (incongruent trials). Interference effects were calculated by subtracting average performance in incongruent trials from congruent trials (monolinguals: $M = -46$ ms, $SD = 21$; bilinguals in L1: $M = -42$ ms, $SD = 21$; bilinguals in L2: $M = -36$ ms, $SD = 20$). Button-press speed was assessed by averaging the reaction times on the neutral trials (monolinguals: $M = 423$ ms, $SD = 39$; bilinguals in L1: $M = 407$ ms, $SD = 46$; bilinguals in L2: $M = 416$ ms, $SD = 47$).
- Socioeconomic status (Morton & Harper, 2007). Participants completed a questionnaire on their socioeconomic status with 11 questions (see Appendix S2; scale ranging from a minimum of 14 to a maximum of 35 points; monolinguals: $M = 22$, $SD = 4$; bilinguals in L1: $M = 22$, $SD = 5$; bilinguals in L2: $M = 26$, $SD = 4$).
- Vocabulary size (Bialystok et al., 2008; Luo et al., 2010). Participants completed a Spanish vocabulary size test (WAIS-III vocabulary subtest with 33 definitions; Wechsler, 1997). A native Spanish speaker evaluated the participants' answers according to the test instructions with zero to two points per definition (monolinguals: $M = 43$, $SD = 4$; bilinguals in L1: $M = 45$, $SD = 3$; bilinguals in L2: $M = 43$, $SD = 4$).

2.3. Design and procedure

Participants were randomly assigned to one of six experimental lists and were tested in a sound proof room. Stimulus presentation and the software voice-key were controlled via DMDX (Forster & Forster, 2003). The sensitivity of the voice key was adjusted for each participant. Participants were seated in front of the computer monitor at a

comfortable viewing distance. Each trial started with a fixation cross displayed at the center of the computer screen for 500 ms. After a 300 ms blank screen, the picture of the object to name was displayed. The picture remained on the screen until either the voice key detected the response or a 2,500 ms deadline was reached without any overt response detected. The next trial began 700 ms after the recording period finished.

The experiment consisted of a short training session followed by two sessions that were separated by a break of 15 min. In the training session, participants were asked to name eight practice pictures similar to the materials used in the experiment. They were instructed to name the pictures as quickly and as accurately as possible using single nouns. After that, in the first session, they had to name the whole set of 533 object pictures divided into eight blocks. The responses were monitored by the experimenter. If participants gave another name for the picture than the intended one, they were corrected by the experimenter at the end of the first session. In the second session, the same 533 pictures were presented in the same way as in the first session, but in a different order.⁵ Participants' responses were automatically recorded by the computer as digitized sound files, and errors were noted online by the experimenter. Each session lasted about 45 min. In total, the experiment including breaks lasted about 2 h.

2.4. Data analyses

All 63,960 vocal responses and speech onset markers (533 pictures \times 2 presentations \times 60 participants) were visually checked offline with the software CheckVocal (Protopapas, 2007) and corrected if necessary. None of the speech onsets provided by the experimental software voice key were transferred. Responses other than the intended target response were classified as errors and excluded from onset latency analysis.

The high number of item-based predictors that we considered a priori (see list above) increased the risk of model over-fitting due to multi-collinearity. Several preliminary steps of variable exploration were taken to attenuate this risk (see Appendix S3). The ALINE measure was retained as the only measure of translation similarity.

Onset latencies and accuracy rates were analyzed by mixed regression models at the single trial level (Baayen et al., 2008). In addition to fixed predictors considered in simple linear regressions, linear mixed-effects models account for random variation induced by specific words or speakers. All statistical analyses were run with the statistical software R (R Development Core Team, 2012) and linear mixed-effects models were computed with the package lme4 in R (Bates, Maechler, & Bolker, 2012). The Box-Cox test (using the function boxcox in the package MASS in R, Venables & Ripley, 2002) indicated that the reciprocal transformation of the latencies was the most appropriate transformation for the data to reduce skewness and approximate a normal distribution. We used $-10,000/RT$ as an order preserving transformation to facilitate the interpretation of our results. We followed Baayen et al.'s (2008) procedure of model criticism in which trials whose standardized residual value is above 2.5 were removed and the model was recomputed.

Our goal was to test for the presence of a bilingual cost in the current dataset (i.e., a significant effect of speaker group), and then test its robustness against the putative underlying variables introduced above (most notably, lexical frequency and translation similarity). If the cost is caused by either of those variables, it should not persist when these variables are included in the model. Note that for this rationale to work, frequency and translation similarity should be estimated as cost effects, just as bilingualism effects are estimated as costs in comparison to monolinguals (i.e., monolinguals are the reference level of comparison). In this way, their estimated effects are positive and can be directly pitted against positive speaker group effects. For this reason, the estimates of the frequency and translation similarity effects were linearly transformed to be referenced to the highest level of the predictor (high frequency and high translation similarity, respectively). This means that we will test whether the bilingual cost is still present under the circumstances of including words of highest frequency and words of highest translation similarity. This choice of referencing lexical frequency and translation similarity involves a linear operation and does not affect the statistical outcome of the model.

Data exploration was driven by the hypotheses described in the Introduction, and achieved by constructing several models of increasing complexity. During model construction, variables significantly contributing to a model's fit were retained and the others were excluded. For any of the significant fixed effects, we checked that the inclusion of individual random slopes for the respective variable would not alter the fixed effect. Any fixed effect that did not survive the inclusion of individual random slopes was discarded from the model (see Barr, Levy, Scheepers, & Tily, 2013). All decisions during model construction were based on model comparison by means of log-likelihood tests, as suggested by Baayen (2008). Control and participant-related predictors were entered in the models before the variables of focus (i.e., lexical frequency and translation similarity). This procedure ensured that any effect of the theoretically central predictors was significant over and above the variation explained by other secondary predictors (see Appendix S4 for details on model construction).

3. Results

3.1. Analysis of response latencies

After removing errors and non-target responses (i.e., responses that did not match the intended target name of the picture; 14%), 55,249 responses remained for analyses. As outlined in the Introduction, we added the 28,046 error-free and offline checked responses and from the monolingual speakers tested in Sadat et al. (2014) to be used as a control group, for a total of 83,295 trials.

In a linear model that included control and participant-related variables, the average naming latency was 910 ms ($SD = 118$) for monolinguals, 947 ms ($SD = 92$) for bilinguals in L1, and 964 ms for bilinguals in L2 ($SD = 93$). The difference between monolinguals and bilinguals in L1 ($p = .016$) and between monolinguals and bilinguals in L2 was

significant ($p = .007$). This establishes that both L1 and L2 groups showed the bilingual cost against monolinguals that is at stake in this article. In contrast, the difference between bilinguals in L1 and L2 was not significant ($p = .354$).

In the model where lexical frequency was included, this predictor had a significant effect and yet the difference between monolinguals and bilinguals in L1 and L2 remained significant. This shows that a bilingual cost resists the partialing out of the differential contribution of the frequency effect across groups. There was no significant interaction between speaker group and lexical frequency after including individual random slopes for lexical frequency.

In the next model, we added translation similarity to the model. This predictor had a significant effect, and importantly, now bilinguals in L1 and L2 were not significantly different from monolinguals (see Fig. 1). There was a significant increase in model fit with a reduction of 146 AIC when comparing this last model to the previous one ($\chi^2(6) = 157.81, p < .001$). Table 2 presents the statistical values for the fixed and random effects of the final linear mixed-effects model on naming latencies. As expected, translation similarity interacted significantly with speaker group. There was no translation similarity effect in monolinguals (since they do not have translations), but L1 and L2 speakers both showed significant effects. Bilingual latencies were faster for words with more translation similarity. This effect was significantly stronger in bilinguals speaking in L2 than L1 ($\chi^2(1) = 6.36, p = .011$). In sum, the inclusion of translation similarity as a predictor of naming latencies resulted in a disappearance of the previously observed bilingual cost.

Note that the other variables included in this and previous models for control purposes showed the expected effects (cf. Table 2). Trial order and session were significant, showing that responses to pictures became slower with increasing trial order within one

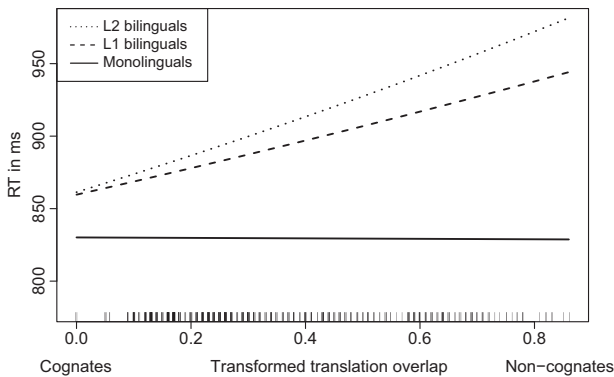


Fig. 1. Estimates of the partial effects of translation similarity and speaker group in the final model. The effect of translation similarity was absent in monolinguals (as expected) and was indistinguishable between the two bilingual groups. The speaker group difference was not significant for the highest translation similarity values (i.e., cognates) which were used as reference point for the model estimates. The difference between monolinguals and bilinguals in L1 and L2 was significant when estimated for low and centered translation similarity values. The distribution of the translation similarity values is depicted along the x-axis.

Table 2

Variance, standard deviation (*SD*), and variance–covariance structure of the random effects together with beta coefficients, standard errors (*SE*), *t*-values, and variance inflation factor (*VIF*) for the fixed effects in the final linear mixed effect model on naming latencies

Random Effects	Variance	<i>SD</i>	Correlations		
Item					
Intercept	1.48	1.28			
Bilinguals in L1	0.22	0.47	−0.297		
Bilinguals in L2	0.36	0.60	−0.359		0.863
Participant					
Intercept	1.93	1.39			
Freq _{max}	0.12	0.34	−0.439		
TrSim _{max}	0.26	0.51	−0.506		0.175
Fixed Effects	raw β	<i>SE</i> β	<i>t</i> -value	<i>VIF</i>	
Intercept	−16.13	1.21	−13.33		
Session	−0.10	0.02	−62.18		1.0
Trial order	<0.01	<0.01	23.43		1.0
Bilinguals in L1	0.30	0.33	0.93		1.7
Bilinguals in L2	0.29	0.33	0.87		1.7
BPSpeed	0.01	<0.01	3.37		1.0
Nagr	−0.06	<0.01	−17.67		1.5
AoA	0.38	0.05	7.90		1.8
PhND	0.02	<0.01	2.15		1.1
Freq _{max}	0.25	0.11	2.16		1.3
TrSim _{max}	−0.02	0.30	−0.08		1.3
Bilinguals in L1 × Nagr	0.01	<0.01	5.79		2.3
Bilinguals in L2 × Nagr	−0.02	<0.01	7.58		2.4
Bilinguals in L1 × AoA	−0.13	0.02	−5.22		2.4
Bilinguals in L2 × AoA	−0.12	0.03	−4.24		2.3
Bilinguals in L1 × TrSim _{max}	1.23	0.20	6.23		2.1
Bilinguals in L2 × TrSim _{max}	1.68	0.22	7.78		2.1

Notes. AoA, age of acquisition; BPSpeed, button-press speed; Freq_{max}, lexical frequency referenced at highest frequency; Nagr, correct target name agreement; PhND, phonological neighborhood density; SE, standard error; TrSim_{max}, phonological similarity across translations referenced at highest similarity.

session and that responses in the second session were faster than in the first. The measure of button-press speed was significant, showing that participants who were fast responders in a button-press task were also faster in naming pictures. As expected, there were significant effects of PhND, AoA, and correct target name agreement in all three speaker groups: Naming latencies increased with higher numbers of phonological neighbors, earlier learned words were named faster than later learned ones, and words with high percentages of correct target name agreement were named faster than words with lower percentages. The latter two item-related predictors interacted significantly with speaker group. Both bilingual groups showed a smaller effect than monolinguals for correct target name agreement and AoA. The effects of executive control, socioeconomic status, vocab-

ulary size, word length measured in phonemes, and first syllable frequency were not significant.

As a final step of the analysis, we focused on individual variations in the effects of lexical frequency and translation similarity, and their relation to individual speed of response. To do so, we compared the previously described full model with correlation parameters specified between individual intercepts and slopes to a model without those parameters (i.e., no-random correlation model). This allows us to investigate the relationship between magnitude of the variable effects (slope estimates of lexical frequency and translation similarity, respectively) and response speed (intercept estimate) for each individual participant (without ascribing the participants to separate speaker groups), and to test whether there is a significant correlation between these two estimates. Contrary to what was done in the previous models, lexical frequency and translation similarity were now entered as centered predictors to ensure the correct estimation of the correlation parameters between individual intercepts and slopes.⁶ The comparison of models with and without correlation parameters for lexical frequency showed no significant difference, hence failing to reveal any relationship between individual response speed and the individual lexical frequency effect size ($\chi^2(1) = 0.01, p = .924$). On the contrary, the same comparison involving translation similarity showed a significant relationship between individual response speed and the individual translation similarity effect size ($\chi^2(2) = 6.19, p = .045$).⁷ This latter result indicates that translation similarity effects vary with individual response speed, which is consistent with the observation of a bilingual cost in naming (i.e., faster naming latencies for monolinguals who did not show any translation similarity effect, and slower naming latencies for bilinguals who showed translation similarity effects).

3.2. Analysis of response accuracy

Responses containing speech errors (monolinguals: 1,681 trials; bilinguals in L1: 1,484 trials; bilinguals in L2: 2,048 trials) were contrasted with error-free responses (monolinguals: 28,046 trials; bilinguals in L1: 28,024 trials; bilinguals in L2: 27,225 trials) to accurately predict the probability of an error-free response. The analysis of response accuracy using generalized linear mixed-effects models paralleled the analysis on naming latencies in model construction and by including the same predictors.

In the model where lexical frequency was included, there was no difference between the odds of errors of monolinguals and bilinguals in L1 and L2. Lexical frequency had a significant effect for all three speaker groups, showing higher odds of error for lower than higher frequency words. There was a significant interaction between this variable and speaker group, showing that the odds of error were higher for bilinguals in L1 and L2 than for monolinguals with decreasing frequency of the words. This difference between the two bilingual groups was not significant.

When adding translation similarity in the next model, importantly, now the odds of errors for bilinguals in L1 and L2 were significantly lower than for monolinguals. There was a significant increase in model fit with a reduction of 199 AIC when comparing this

last model to the previous one ($\chi^2(6) = 211.02, p < .001$). Table 3 presents the statistical values for the fixed and random effects of the final generalized linear mixed-effects model on response accuracy. The odds of error were significantly less for bilinguals in L1 and L2 than monolinguals (both $ps < .001$). The difference between bilinguals in L1 and L2 was not significant ($p = .417$). As expected, translation similarity did not have an effect on the odds of error in monolinguals, but bilinguals in L1 and L2 were more likely to make errors with decreasing translation similarity of the words. This latter difference was significant, showing that the odds of error were higher for bilinguals in L2 than bilinguals in L1 with decreasing translation similarity.

Table 3

Variance, standard deviation (*SD*), and variance–covariance structure of the random effects together with beta coefficients, standard errors (*SE*), *z*-values, *p*-values, and variance inflation factor (*VIF*) for the fixed effects in the final linear mixed effect model on accuracy rates

Random Effects	Variance	<i>SD</i>	Correlations		
Item					
Intercept	1.37	1.17			
Bilinguals in L1	0.69	0.83	0.745		
Bilinguals in L2	1.36	1.17	0.315		0.745
Participant					
Intercept	4.81	2.19			
Freq _{max}	0.38	0.61	−0.694		
TrSim _{max}	1.15	1.07	−0.303		−0.191
Fixed Effects	raw β	<i>SE</i> β	<i>z</i> -value	<i>p</i> -value	<i>VIF</i>
Intercept	0.313	1.82	0.17	0.863	
Session	−2.28	0.06	−37.97	<0.001	1.0
Trial order	<0.01	<0.01	4.00	<0.001	1.0
Bilinguals in L1	−3.55	0.82	−4.32	<0.001	5.7
Bilinguals in L2	−2.88	0.83	−3.48	<0.001	5.5
Nagr	−0.09	<0.01	−19.95	<0.001	1.3
AoA	0.68	0.08	8.98	<0.001	1.3
Freq _{max}	1.16	0.24	4.74	<0.001	1.3
TrSim _{max}	0.44	0.49	0.90	0.369	1.1
Bilinguals in L1 × Nagr	<0.01	<0.01	0.04	0.971	1.7
Bilinguals in L2 × Nagr	0.03	0.01	4.75	<0.001	1.6
Bilinguals in L1 × AoA	−0.17	0.08	−2.13	0.033	1.7
Bilinguals in L2 × AoA	−0.03	0.09	−0.34	0.734	1.7
Bilinguals in L1 × Freq _{max}	0.99	0.30	3.31	<0.001	5.4
Bilinguals in L2 × Freq _{max}	0.78	0.31	2.51	0.012	5.3
Bilinguals in L1 × TrSim _{max}	3.68	0.53	7.01	<0.001	2.0
Bilinguals in L2 × TrSim _{max}	5.68	0.58	9.82	<0.001	2.1

Notes. AoA, age of acquisition; Freq_{max}, lexical frequency referenced at highest frequency; Nagr, correct target name agreement; SE, standard error; TrSim_{max}, phonological similarity across translations referenced at highest similarity.

Regarding the additional control variables, for all three speaker groups the odds of error were smaller for words that had higher correct target name agreement and were learned earlier. Bilinguals in L2 showed a smaller effect for correct target name agreement than monolinguals and bilinguals in L1, and bilinguals in L1 showed a smaller effect for AoA than monolinguals. Session and trial order were significant, showing that the odds of error were less in the second than in the first session and that they increased with increasing trial order. There was a significant effect of vocabulary size, showing that the odds of error were lower for participants with higher vocabulary scores. Measures of button-press speed, executive control, socioeconomic status, PhND, word length, and first-syllable frequency were not significant.

4. Discussion

In this study, we compared performance of mono- and bilinguals in a large-scale picture naming study. We estimated how participant- and item-related variables predict bilingual naming performance, and we explored the extent to which these variables modulated performance differences between mono- and bilingual speakers. Motivated by previous bilingualism research, our analysis focused on the contributions of lexical frequency and translation similarity, at the speaker group level and at the individual speaker level. These two variables have been shown to have a highly influential role on naming performance, yet their relationship with the bilingual cost was not firmly established so far.

First of all, the general observation of faster naming latencies for monolinguals than bilinguals was replicated. These results are in line with various picture-naming studies showing a naming cost for bilingual speakers of L2 (e.g., Gollan et al., 2008; Ivanova & Costa, 2008; Kohnert et al., 1998) and even L1 (Ivanova & Costa, 2008; Sadat et al., 2012). This result provided the basis for further investigations of the key variables underlying the bilingual cost.

The results of this study showed that the bilingual cost remained large and significant along the continuum of the lexical frequency variable, including the highest frequency words, both in L1 and in L2. This means that there still was a significant bilingual cost after the contribution of lexical frequency had been partialled out. Regarding the lexical frequency effect, it was indistinguishable between monolinguals and bilinguals speaking in L1 and L2. The weaker links hypothesis predicts that overall speed and magnitude of frequency effect are yoked together and should be similarly influenced by language use (i.e., speaker group). According to this view, naming latencies as well as the magnitude of the lexical frequency effect should both increase with reduced practice (from monolinguals to bilinguals speaking in L1 to bilinguals speaking in L2; Gollan et al., 2008). Thus, the present results of similar sized lexical frequency effects in mono- and bilinguals do not follow the prediction of the weaker links account (e.g., Cop, Keuleers, Drieghe, & Duyck, 2015; Duyck, Vanderelst, Desmet, & Hartsuiker, 2008; Gollan et al., 2008; Ivanova & Costa, 2008). However, the predictions of the weaker links account seem unclear regarding the interpretation of lexical frequency effects. The seminal paper by Oldfield

and Wingfield (1964) established the logarithmic relationship between lexical frequency and naming times in picture naming. In other words $RT \sim \log(\text{freq})$ and the frequency effect can be estimated as $\text{Freq}_{\text{eff}} = RT_{\text{HF}} - RT_{\text{LF}} \sim \log(\text{HF}) - \log(\text{LF})$. If we assume that bilingual speakers use their words a fraction α ($\alpha < 1$) of the time, the frequency for high- and low-frequency words will be approximated in bilinguals by $\alpha * \text{HF}$ and $\alpha * \text{LF}$, respectively. Then, exactly the same logarithmic relationship holds for the frequency effect in mono- and bilinguals; this is because for bilinguals $\text{Freq}_{\text{eff}} \sim \log(\alpha * \text{HF}) - \log(\alpha * \text{LF}) = \log(\alpha) + \log(\text{HF}) - (\log(\alpha) + \log(\text{LF})) = \log(\text{HF}) - \log(\text{LF})$, just as for monolinguals (see also Cop et al., 2015). Our data are consistent with this observation in the case of mono- and bilingual speakers. We did not observe a bilingual cost that is modulated by lexical frequency, but a cost that remains unchanged across the range of lexical frequency (i.e., similar frequency effects across speaker groups).

A possible candidate to explain a confounded frequency count in bilinguals would be to consider the cognate status of a word. One could argue that cognates will have functionally more frequent word forms compared to any other language-unique word because cognates occur in both languages. Thus, under the (admittedly strong) assumption that phonologically highly similar translations have an identical lexical representations across languages, they would always have a higher usage than phonologically dissimilar translations. However, this claim is not fully warranted and still debated (cf. Costa et al., 2005), since it remains unclear to what extent cognates would even share the same phonological representation (e.g., Caramazza, Bi, Costa, & Miozzo, 2004; for evidence of different representations for homophones within the same language). Alternatively, it is possible that for some reason specific to the bilingual language architecture, bilinguals would use cognate words overall more often than monolinguals do (Sadat, Pureza, & Alario, unpublished data). Whatever the origin for a confounded bilingual frequency count could be, in this context it is important to mention that in the present data there was no significant interaction between lexical frequency and translation similarity (e.g., see also Costa et al., 2000; Mulder, Dijkstra, Schreuder, & Baayen, 2014). Therefore, the effects of phonological similarity and frequency are assumed to be additive and considered separately in this discussion.

An alternative explanation for the bilingual cost that we considered here claims that delays in bilingual language processing emerge at very late time points of speech processing such as phonological and articulatory processing stages (Hanulová et al., 2011; Indefrey, 2006). In this study, the effects of phonological similarity across translations were tentatively related to this post-lexical account, and this variable has been shown to be highly influential for the presence of a bilingual cost. The bilingual cost, identified as a significant effect of speaker group, was present in all the models tested here, *except* when we included a predictor that coded for phonological similarity across translations. Note that, following a linear transformation of the predictor, translation similarity was referenced on the highest similarity value (i.e., cognates; once again this allows a direct statistical test of the amount of bilingual cost that survives the inclusion of the translation similarity variable). In this case, bilinguals speaking in L1 and L2 were not significantly different from monolinguals. In other words, this pattern of results shows that perfor-

mance did not differ across groups for phonological highly similar translations, and that the bilingual cost significantly increased as translation similarity decreased (see Fig. 1). This finding points to the *absence* of translation similarity as one of the main sources of the bilingual cost. One implication of the lack of naming latency differences between monolingual and bilingual speakers could be that for words with increased translation similarity, “bilingualism” is not a critical variable to account for word production latencies. This observation supports the hypothesis of the post-lexical account that posits that costs should emerge at rather late stages of language processing.

Interestingly, our results on response accuracy further corroborate the important role of translation similarity on bilingual speech performance. There was no significant difference in response accuracy between mono- and bilinguals in all the models tested here, *except* when we included a predictor that coded for phonological similarity across translations. Only when entering the predictor that coded for high translation similarity (i.e., cognates), bilingual speakers showed significantly lower odds of errors than monolinguals. Taken together with our results on onset latencies, this means that when bilinguals name words with high translation similarity (i.e., cognates), they do not only show no cost in performance speed, but also make fewer errors than their monolingual counterparts. This observation highlights the important role of translation similarity for the overall speech performance of bilinguals, when considering both naming latencies and accuracy. To our knowledge, there is no model that has been intended to explain bilingual accuracy data. However, one could imagine a possible extension of the two-step interactive-activation model of Dell and Gordon (2003) with an additional set of phonemes for the second language. If the present findings could be transferred to such a model, one could say that inter-language similarity is highest in the case of cognates (“friend” words), promoting speech accuracy. However, in the case of non-cognates, inter-language similarity is low (“foes”) and thus speech production accuracy would be highly reduced and rather similar to monolinguals.

Previous literature on the relation between the bilingual cost and cognate status of the words is sparse. In a post hoc analysis, Ivanova and Costa (2008) assessed the effects of cognate status on the bilingual cost and observed that the bilingual cost was similar for cognates and non-cognates (in their study, cognates: 32 ms; non-cognates: 35 ms). However, as stated by Ivanova and Costa, their results should be interpreted with caution, since they were assessed post hoc on a small and unbalanced set of stimuli. Christoffels, Firk, and Schiller (2007) reported a reduced, but still significant, cost for cognates when comparing bilingual speech production in L1 and L2. These results contrast with our present findings, in which there was no cost between L1 and L2 bilingual productions. A role of cognate status in bilingual naming performance has also been reported in tasks requiring different word-finding strategies than simple naming. For example, in a verbal fluency study reported in Sandoval, Gollan, Ferreira, and Salmon (2010), bilinguals produced fewer exemplars of a given semantic category than monolinguals. Interestingly, this type of bilingual cost was entirely driven by the fact that bilinguals produced fewer non-cognates than monolinguals while they produced equivalent numbers of cognates (see Fig. 5 in Sandoval et al. [2010]). Furthermore, some studies assessing tip-of-the-tongue

states reported no bilingual cost for cognate words or proper names (Gollan & Acenas, 2004; Gollan, Montoya, & Bonanni, 2005). Together, these and our findings converge on the idea that high translation similarity (or identity as in the case of many proper names) “equalizes” word production performance of mono- and bilingual speakers, whereas the absence of this similarity induces a cost in performance.⁸

Finally, we also explored the influence of participant-related variables on bilingual naming performance. The only significant participant-related variable was button-press speed, showing that participants who were faster in responding to a button-press task were also faster in naming pictures. Contrary to some studies claiming influencing effects of executive control ability (measured as conflict effect in a Simon task), vocabulary size, and socioeconomic status (e.g., Bialystok et al., 2008; Luo et al., 2010; Morton & Harper, 2007), no clear significant effects of these variables were observed. We do not wish to draw any conclusion from these absences of significance, especially because the focus of this study was to explore the effect of item-related predictors by using a large set of stimuli and, somewhat consequently, a moderate number of participants. It may be that more participants are needed to test for the influence of such participant-related properties. We also note the rather coarse estimates of participant-related measures used in this study. Finer assessment techniques would be beneficial to better determine the most important participant-related variables of language processing in bilinguals (see, e.g., Diependaele, Lemhöfer, & Brysbaert, 2013, for detailed measures of vocabulary size).

In essence, our findings indicate that part of the bilingual cost is tied to lexical frequency and, perhaps more important, that this cost is closely linked to a phonological property of the to-be-produced words, namely the degree of translation similarity. However, given that the effect of translation similarity has been attributed to several origins (Costa et al., 2005), our results do not show, in and of themselves, that the corresponding part of bilingual cost emerges at the level of phonological processing. In what follows, we will discuss how translation similarity is thought to affect speech production in bilinguals, and we describe the possible processing origins of the bilingual cost in this context.

4.1. On the origin of the effect of phonological similarity across translations

The prevalent explanation in the literature on cognate effects attributes them to processes occurring during lexical retrieval (Costa et al., 2000, 2005; but see Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010; Sánchez-Casas & García-Albea, 2005; Van Hell & De Groot, 1998, for explanations at the morphological and conceptual level). Due to interactivity in the speech production system (Dell, 1986), high phonological similarity would facilitate lexical retrieval of the word associated with the shared phonemes when compared to words that do not overlap in phonemes. In this context, our results showing the disappearance of the bilingual cost in the case of high translation similarity would mean that bilinguals do not benefit from translation similarity, but rather suffer from non-overlapping translations relative to monolinguals. This may suggest two sources of the bilingual cost.

On one hand, it is possible that phonological dissimilarity triggers interference either at the lexical or phonological level, and thus slows down speech in bilinguals compared to

monolinguals. This would mean that competition between language systems in bilinguals would depend on specific properties of the word to be produced, for example, being highest for phonologically non-overlapping translations and almost nonexistent in the case of highly overlapping ones. For now and with the current data, we cannot unequivocally discard this possibility. One challenge to this account follows from the observation that in monolingual speech production phonological similarity among words has been shown to slow down lexical retrieval (e.g., Gordon & Kurczek, 2013; Sadat et al., 2014; see also Chan & Vitevitch, 2010, for slowing due to words with a rich neighborhood network; but see Vitevitch, 2002). Based on the assumption that such slowing is due to competitive processes and applying this logic to bilinguals, this would entail that phonologically highly similar translations should be detrimental to bilingual speech production. Since the performance of mono- and bilinguals is similar in the case of highly overlapping translations, a consequence of the above would be that the lexical representations of translations do not compete for lexical selection (e.g., Costa & Caramazza, 1999; Costa, Miozzo, & Caramazza, 1999).

On other hand, one should consider the amount of inter-language interactions at the phonological and phonetic level in bilinguals and the extent to which the representations of two languages of a bilingual are shared. Recent studies using acoustical analysis revealed that cognate speech production is influenced by the two languages compared to non-cognate production in the case of Spanish-English and Catalan-Spanish bilinguals (Amengual, 2012; Brown & Amengual, 2015; Mora & Nadeu, 2012; Nip & Blumenfeld, 2015). That is, bilinguals show merged phonetic and phonological categories in the case of cognates which would remove any competition effects. However, since this overlap is not warranted in the case of non-cognates, it is more likely then that competition effects arise between different inter-language categories. Thus, the slowing of non-cognates results from the internal structure of overlap and spread within the mental lexicon. Previous studies have shown that lexical retrieval is influenced by the clustering or spread of the lexical representations within the mental lexicon (e.g., Chan & Vitevitch, 2010; Yates, Friend, & Ploetz, 2008). This refers to the general idea that more similar words (both semantically and phonologically) are represented in a more clustered manner than less similar ones. Due to the non-overlapping phonological and articulatory features, the structural representations of non-cognates would be less clustered than those of cognates. Given this additional dimension of inter-language similarity in bilinguals relative to monolinguals, translation similarity may play an important role in determining how lexical retrieval differs between these two groups of speakers. In the case of non-cognates, lexical representations would be less clustered and thus slower to retrieve than representations in monolinguals. Only in the case of cognates, lexical representations would resemble those of their monolingual counterparts and retrieval of the lexical item would be accomplished without an additional cost in bilinguals. Computational modeling would provide important insights into the mechanisms underlying phonological similarity effects in bilinguals. Since such a model will be based on architectures and mechanisms of monolingual speech production, future steps should elaborate on how phonological neighborhood and translation similarity effects could be captured in bilingual speech production.⁹

Previous research has not clarified whether reported processing benefits for cognates could result exclusively from the converging activation at the phonological level, and independent of conceptual similarity. Interestingly, our results on bilingual latencies showed no effect of purely phonological inter-language similarity as assessed by the number of phonological neighbors in the unspoken language (see also De Groot, Borgwaldt, Bos, & van den Eijnden, 2002; Lemhöfer et al., 2008, for no effect of orthographic inter-language neighbors). When bilinguals produced a word, there was no effect of the number of phonologically similar words of the unspoken language. Thus, inter-language influences of phonological similarity seem to be restricted to translations (but see Costa, Roelstraete, & Hartsuiker, 2006). The present results suggest that translation similarity effects do not solely derive from pure phonological influences across languages. Instead, explanations for cognate effects should consider additional assumptions relying on shared conceptual or lexical representations. One plausible explanation is that in order for phonological similarity effects to surface, there should be enough semantic activation present in the language production network. With respect to a possible bilingual extension of the speech production model by Dell and Gordon (2003), one could suggest that inter-language similarity is only influential in the case of semantic overlap, and that there is a modulation of L2 phonology activation dependent on semantics.

However, a recent study by Gollan and Goldrick (2012) suggested that the bilingual cost is present even in situations where no word retrieval is involved, and where only phonetic processing is required. This idea contrasts with our observation that for phonological similarity effects to arise there must be a semantic overlap. Gollan and Goldrick (2012) asked English monolinguals, early highly proficient Mandarin-English bilinguals, and early highly proficient Spanish-English bilinguals to repeat aloud tongue twisters consisting of non-words and words (tongue twisters are combinations of phonetically similar segments). Their results showed that overall both bilingual groups produced more tongue twister errors than monolinguals. Importantly, bilinguals also showed more errors in the case of non-word tongue twisters. It could be that in a special task setting like tongue twister production, phonological effects are more prone to arise than in single word production. This result suggests that there may be an additional and independent sublexical locus for the bilingual cost in speech production.

Before concluding, we note that the observation that the bilingual cost depends on the amount of translation similarity invites some predictions regarding different forms of bilingualism. Bilinguals speaking phonologically more similar languages are expected to experience less of a cost than bilinguals speaking two phonologically more distant languages, compared to monolinguals. This is because there will be overall more phonologically similar words in more similar language pairs than in more distant ones (see also Costa et al., 2012). A comparison of studies that have previously assessed the bilingual cost in different populations of bilinguals seems to endorse this pattern. As can be seen in Table 1 of Hanulová et al. (2011), the cost between mono- and bilinguals or L1 and L2 is at a minimum for Spanish-Catalan bilinguals, whereas more distant language combinations (e.g., Spanish-English) show larger costs. One caveat here is that the studies also investigated different populations of bilinguals. Most studies tested sequential bilin-

guals (i.e., they learned one language after the other with a large time delay between the two; e.g., Christoffels et al., 2007) or switched-dominant bilinguals (i.e., their first learned and dominant language became the non-dominant language over time; Gollan, Montoya, & Bonanni, 2005; Gollan, Montoya, Fennema-Notestine, et al., 2005; Gollan et al., 2008), whereas this study assessed performance in dominant, but early and relatively simultaneous bilinguals. It is still an open question as to how the present findings of the bilingual cost would apply to sequential bilinguals, and how much of our findings could be generalized to explain the L2 delay in speech production (i.e., slower response latencies in L2 than L1; Runnqvist et al., 2011). Note finally that the results of the study by Gollan and Goldrick (2012) in which tongue twister production was compared across Spanish-English, Mandarin-English, and English-only groups, indicate that there were differences in the error patterns at sublexical processing between the two bilingual groups. Mandarin-English bilinguals showed a more consistent cost compared to monolinguals over all tongue twister conditions than did Spanish-English bilinguals. This finding fits our proposal based on this study that the more phonologically dissimilar the languages of a bilingual, the larger the bilingual cost should be.

5. Conclusion

This study reveals that one of the key factors explaining performance differences in mono- and bilingual speakers is the amount of phonological similarity across translations. We argue that the relative absence of phonological similarity across translations promotes a bilingual cost, whereas similarity across translations helps bilinguals reaching monolingual levels of performance. Thus, our findings establish a direct relation between these two important phenomena of bilingual speech, previously described as the cognate effect and the bilingual cost observed in verbal tasks. This provides a new view of the involvement of phonological similarity in explaining the bilingual cost, and it suggests an important role of sublexical features in bilingual lexical retrieval performance.

Acknowledgments

This research was supported by a grant from the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013 Grant agreement no. 263575), three grants from the Spanish Government (PSI2008-01191, PSI2011-23033, Consolider Ingenio 2010 CSD2007-00012), and the Catalan Government (Consolidado SGR 2009-1521). This work, carried out within the Labex BLRI (ANR-11-LABX-0036), has benefited from support from the French Government, managed by the French National Agency for Research (ANR), under the project title Investments of the Future A*MIDEX (ANR-11-IDEX-0001-02). We thank the Brain and Language Research Institute and the Fédération de Recherche 3C (both at Aix-Marseille Université) for institutional support. Jasmin Sadat was supported by a pre-doctoral fellowship from the Spanish

Government (FPU-2008). Clara Martin was supported by the Spanish Government (Grant Juan de la Cierva) and is now supported by the Basque Foundation for Science (IKER-BASQUE) and the BCBL Institution. James Magnuson was supported by a U.S. National Institutes of Health grant to Haskins Laboratories (P01 HD001994, Jay Rueckl, PI). We thank Laure Minier and Elena Guerra for checking a portion of the vocal responses, and Paula Marne for help with the Spanish vocabulary ratings. We would also like to thank Harald Baayen, Natasha Tokowicz, and two anonymous reviewers for their comments on an earlier version of this manuscript.

Notes

1. It is not unlikely that the amount of competition between language systems or the effect of language use may depend on certain word properties (see previous section). We will come back to this issue in the Discussion, when interpreting the effects of phonological similarity across translations.
2. For this study, we did not consider any orthographic variables, since the role of orthographic properties has been shown to be negligible in picture naming performance (see review article by Alario, Perre, Castel, & Ziegler, 2007).
3. Although we tried to avoid words with ambiguous translations, 21 words with infrequent translation alternatives remained in the stimulus set. We asked participants after the experiment to translate those words from Spanish to Catalan to ensure that we used the dominant target translation. The results of this test confirmed the choice of the target translations. For the analysis, we also verified that our results would not change when excluding these words.
4. One important limitation of this study is that we do not provide information on the extent to which positional effects of translation similarity are important. Similarly as for phonological neighborhood density effects in monolinguals (e.g., Bien, Baayen, & Levelt, 2011), we suspect that similarity effects across translations would differ depending on the position of the similarity (e.g., beginning vs. end). For example, it may be the case that a translation with high similarity would only affect naming when the similarity occurred at the beginning as opposed to the end of a word. Thus, positional aspects of similarity (among other things) could be important when assessing the influence of phonological similarity across translations. Our translation similarity measure (ALINE; Kondrak, 2000) provided positional alignment between translations. However, further detailed investigations are needed on how translation similarity effects may vary according to position of similarity.
5. The procedure of familiarizing participants with the task and the materials in a first block follows common practice in the picture naming literature (e.g., Alario et al., 2004).
6. Due to the non-significant speaker group effect in the final model, we retained only the significant interactions with speaker group for the models testing the relation-

ship between individual intercepts and slopes. Note that the results remained unchanged when including speaker group as a fixed effect.

7. Note that when model comparisons were performed on the subset of bilingual speakers only, there was no significant correlation between individual response speed and the translation similarity effect, suggesting that this effect does not vary with response speed in bilingual speakers.
8. An alternative and less parsimonious account cannot be readily excluded, in which (a) the bilingual cost has a different and yet undescribed origin, and (b) phonological similarity compensates for this cost, to the point that it equalizes precisely mono- and bilingual performance for highly similar translations.
9. A third option would be to explain the bilingual cost in terms of monitoring processes. In this scenario, one would have to assume a mechanism that monitors all activated lexical items. Phonological similarity of an intended word to its phonological similar translation could increase the likelihood for this word to slip by the monitor, whereas dissimilar words could trigger additional monitoring processes and delay articulation. However, since the underlying mechanisms of monitoring processes and their application to the two languages of a bilingual are still debated, we refrain from further elaboration.

References

- Abutalebi, J., & Green, D. W. (2008). Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and Cognitive Processes*, 23(4), 557–582.
- Alario, F.-X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, 36, 140–155.
- Alario, F.-X., Perre, L., Castel, C., & Ziegler, J. C. (2007). The role of orthography in speech production revisited. *Cognition*, 102(3), 464–475.
- Alario, F.-X., Ziegler, J. C., Massol, S., & De Cara, B. (2012). Probing the link between cognitive control and lexical selection in monolingual speakers. *L'Annee Psychologique/Topics in Cognitive Psychology*, 112, 545–559.
- Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition*, 15(3), 517–530.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28.
- Balota, D. A., & Ferraro, F. R. (1993). A dissociation of frequency and regularity effects in pronunciation performance across young adults, older adults, and individuals with senile dementia of the Alzheimer type. *Journal of Memory and Language*, 32, 573–592.
- Balota, D. A., & Ferraro, F. R. (1996). Lexical, sublexical, and implicit memory processes in healthy young and healthy older adults and in individuals with dementia of the Alzheimer type. *Neuropsychology*, 10(1), 82.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.

- Barry, C., Hirsh, K., Johnston, R. A., & Williams, C. (2001). Age of acquisition, word frequency, and the locus of repetition priming of picture naming. *Journal of Memory and Language*, *44*, 350–375.
- Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.
- Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111.
- Belsley, D. A. (1984). Demeaning conditioning diagnostics through centering. *American Statistician*, *38*, 73–82.
- Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest*, *10*(3), 89–129.
- Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, *19*, 290–303.
- Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics*, *21*(6), 522–538.
- Bien, H., Baayen, R. H., & Levelt, W. J. M. (2011). Frequency effects in the production of Dutch deverbal adjectives and inflected verbs. *Language and Cognitive Processes*, *26*(4–6), 683–715.
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.
- Brown, E. L., & Amengual, M. (2015). Fine-grained and probabilistic cross-linguistic influence in the pronunciation of cognates: Evidence from corpus-based spontaneous conversation and experimentally elicited data. *Studies in Hispanic and Lusophone Linguistics*, *8*(1), 59–83.
- Caramazza, A., Bi, Y., Costa, A., & Miozzo, M. (2004). What determines the speed of lexical access: Homophone or specific-word frequency? A Reply to Jescheniak et al. (2003). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *30*, 278–282.
- Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin*, *98*(1), 67.
- Chan, K. Y., & Vitevitch, M. S. (2010). Network structure influences speech production. *Cognitive Science*, *34*(4), 685–697.
- Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research*, *1147*, 192–208.
- Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language*, *45*(4), 721–736.
- Colomé, À., & Miozzo, M. (2010). Which words are activated during bilingual word production? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 96.
- Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, *22*(5), 1216–1234.
- Costa, A., Calabria, M., Marne, P., Hernández, M., Juncadella, M., Gascón-Bayarri, J., Lleó, A., Ortiz-Gil, J., Ugas, L., Blesa, R., & Reñé, R. (2012). On the parallel deterioration of lexico-semantic processes in the bilinguals' two languages: Evidence from Alzheimer's disease. *Neuropsychologia*, *50*, 740–753.
- Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish-English and English-Spanish bilinguals. *Bilingualism*, *2*, 231–244.
- Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283–1296.
- Costa, A., Hernandez, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*, 59–86.
- Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, *41*, 365–397.
- Costa, A., Roelstraete, B., & Hartsuiker, R. (2006). The lexical bias effect in bilingual speech production: Evidence for feedback between lexical and sublexical levels across languages. *Psychonomic Bulletin & Review*, *13*, 612–617.

- Costa, A., Santesteban, M., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language*, 94(1), 94–103.
- Cuetos, F., Ellis, A. W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods*, 31(4), 650–658.
- Dalal, D. K., & Zickar, M. J. (2012). Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organizational Research Methods*, 15, 339–362.
- Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, 37(4), 665–671.
- De Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, 26(1), 99–107.
- De Groot, A., Borgwaldt, S., Bos, M., & van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language*, 47(1), 91–124.
- De Groot, A. M., & Nas, G. L. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, 30(1), 90–123.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283–321.
- Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon: Friends or foes? In N. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 9–37). Germany: Mouton de Gruyter.
- Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first and second language word recognition: A lexical entrenchment account. *The Quarterly Journal of Experimental Psychology*, 66(5), 843–863.
- Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, 41(4), 496–518.
- Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. R. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, 62, 284–301.
- Duyck, W., Vandereelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency-effect in second-language visual word recognition. *Psychonomic Bulletin & Review*, 15, 850–855.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods Instruments and Computers*, 35(1), 116–124.
- Friel, B. M., & Kennison, S. M. (2001). Identifying German-English cognates, false cognates, and non-cognates: Methodological issues and descriptive norms. *Bilingualism: Language and Cognition*, 4(3), 249–274.
- Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3), 474–496.
- Gollan, T. H., & Acenas, L. R. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish-English and Tagalog-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1), 246–269.
- Gollan, T. H., Fennema-Notestine, C., Montoya, R. I., & Jernigan, T. L. (2007). The Bilingual Effect on Boston Naming Test performance. *The Journal of the International Neuropsychological Society*, 13, 197–208.
- Gollan, T. H., & Goldrick, M. (2012). Does bilingualism twist your tongue? *Cognition*, 125(3), 491–497.
- Gollan, T. H., Montoya, R., & Bonanni, M. (2005). Proper names get stuck on bilingual and monolingual speakers' tip of the tongue equally often. *Neuropsychology*, 19(3), 278–287.
- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always a means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58(3), 787–814.
- Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, 33(7), 1220–1234.

- Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, *16*(4), 562–576.
- Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew-English bilinguals. *Bilingualism: Language and Cognition*, *4*(1), 63–83.
- Gollan, T. H., Slattery, T. J., Goldenberg, D., van Assche, E., Duyck, W., & 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.
- Gordon, J. K., & Kurczek, J. (2013). The aging neighborhood: Phonological density in naming. *Language and Cognitive Processes*, *29*(3), 326–344.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81.
- Guion, S. G., Flege, J. E., Liu, S. H., & Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, *21*(02), 205–228.
- Hanulová, J., Davidson, D. J., & 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*(7), 902–934.
- Harrell, F. E. Jr., with contributions from many other users (2010). Hmisc: Harrell Miscellaneous. R package version 3.8-3.
- Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin Review*, *18*(4), 625–658.
- Hothorn, T., Buehlmann, P., Dudoit, S., Molinaro, A., & Van Der Laan, M. (2006). Survival ensembles. *Biostatistics*, *7*(3), 355–373.
- Indefrey, P. (2006). A meta-analysis of hemodynamic studies on first and second language processing: Which suggested differences can we trust and what do they mean? *Language Learning*, *56*, 279–304.
- Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*(2), 277–288.
- Kohnert, K. J., Hernandez, A. E., & Bates, E. (1998). Bilingual performance on the Boston naming test: Preliminary norms in Spanish and English. *Brain and Language*, *65*(3), 422–440.
- Kondrak, G. (2000). A new algorithm for the alignment of phonetic sequences. Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference, 288–295. Seattle, Washington; Apr 29 to May 4, 2000.
- Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 12–31.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*(8), 707–710.
- Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon*. Ph.D. dissertation, Indiana University.
- Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition*, *114*(1), 29–41.
- Mackay, I. R. A., & Flege, J. E. (2004). Effects of the age of second language learning on the duration of first and second language sentences: The role of suppression. *Applied Psycholinguistics*, *25*(03), 373–396.
- Mora, J. C., & Nadeu, M. (2012). L2 effects on the perception and production of a native vowel contrast in early bilinguals. *International Journal of Bilingualism*, *16*(4), 484–500.
- Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, *10*(6), 719–726.
- Mulder, K., Dijkstra, T., Schreuder, R., & Baayen, H. R. (2014). Effects of primary and secondary morphological family size in monolingual and bilingual word processing. *Journal of Memory and Language*, *72*, 59–84.

- Nip, I. S., & Blumenfeld, H. K. (2015). Proficiency and linguistic complexity influence speech motor control and performance in Spanish language learners. *Journal of Speech, Language, and Hearing Research*, 58, 653–668.
- Oldfield, R. C., & Wingfield, A. (1964). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17(4), 273–281.
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66(2), 232–258.
- Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, 39(4), 859–862.
- R Development Core Team (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.
- Rafel i Fontanals, J. (1996). *Diccionari de freqüències: Llengua no literària*. Barcelona, Spain: Institut d'Estudis Catalans.
- Roberts, P. M., Garcia, L. J., Desrochers, A., & Hernandez, D. (2002). English performance of proficient bilingual adults on the Boston naming test. *Aphasiology*, 16(4–6), 635–645.
- Runnqvist, E., Strijkers, K., Sadat, J., & Costa, A. (2011). On the temporal and functional origin of L2 disadvantages in speech production: A critical review. *Frontiers in Psychology*, 2, 379.
- Sadat, J., Martin, C. D., Alario, F.-X., & Costa, A. (2012). Characterizing the bilingual disadvantage in noun phrase production. *Journal of Psycholinguistic Research*, 41(3), 159–179.
- Sadat, J., Martin, C. D., Costa, A., & Alario, F.-X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, 68, 33–58.
- Sadat, J., Pureza, R., & Alario, F.-X. (under review). Traces of an early learnt second language in speech production of switched-dominance bilinguals.
- Sánchez-Casas, R., & García-Albea, J. E. (2005). The representation of cognate and noncognate words in bilingual memory. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 226–250). New York: Oxford University Press.
- Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition*, 13, 231–252.
- Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distributions of cognates in Europe as based on Levenshtein distance. *Bilingualism: Language and Cognition*, 15(1), 157–166.
- Schwartz, A. I., Kroll, J. F., & Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes*, 22(1), 106–129.
- Simon, J. R., & Wolf, J. D. (1963). Choice reaction time as a function of angular stimulus-response correspondence and age. *Ergonomics*, 6(1), 99–105.
- Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, 28(4), 516–536.
- Spieler, D. H., & Balota, D. A. (2000). Factors influencing word naming in younger and older adults. *Psychology and Aging*, 15(2), 225.
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9(307).
- Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(25).
- Székely, A., D'Amico, S., Devescovi, A., Federmeier, K., Herron, D., Iyer, G., Jacobsen, T., & Bates, E. (2003). Timed picture naming: Extended norms and validation against previous studies. *Behavior Research Methods, Instruments, & Computers*, 35(4), 621–633.
- Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(2), 135–178.
- Oldfield, R. C., & Wingfield, A. (1964). The time it takes to name an object. *Nature*, 202, 1031–1032.
- Van Hell, J. G., & De Groot, A. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, 1(03), 193–211.

- Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, 15(3), 181–198.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.
- Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 735–747.
- Wechsler, D. (1997). *Wechsler adult intelligence scale* (3rd ed.). New York: The Psychological Corporation.
- Wieling, M., Margaretha, E., & Nerbonne, J. (2012). Inducing a measure of phonetic similarity from pronunciation variation. *Journal of Phonetics*, 40, 307–314.
- Yates, M., Friend, J., & Ploetz, D. M. (2008). Phonological neighbors influence word naming through the least supported phoneme. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6), 1599–1608.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Description of the bilingual community in Catalonia

Appendix S2. Excerpt of the socioeconomic questionnaire

Appendix S3. Preliminary steps of variable exploration

Appendix S4. Model construction during data exploration

Appendix S5. Full list of the stimulus set