

Frequency Drives Lexical Access in Reading but Not in Speaking: The Frequency-Lag Hypothesis

Tamar H. Gollan, Timothy J. Slattery, and
Diane Goldenberg
University of California, San Diego

Eva Van Assche and Wouter Duyck
Ghent University

Keith Rayner
University of California, San Diego

To contrast mechanisms of lexical access in production versus comprehension we compared the effects of word frequency (high, low), context (none, low constraint, high constraint), and level of English proficiency (monolingual, Spanish–English bilingual, Dutch–English bilingual) on picture naming, lexical decision, and eye fixation times. Semantic constraint effects were larger in production than in reading. Frequency effects were larger in production than in reading without constraining context but larger in reading than in production with constraining context. Bilingual disadvantages were modulated by frequency in production but not in eye fixation times, were not smaller in low-constraint contexts, and were reduced by high-constraint contexts only in production and only at the lowest level of English proficiency. These results challenge existing accounts of bilingual disadvantages and reveal fundamentally different processes during lexical access across modalities, entailing a primarily semantically driven search in production but a frequency-driven search in comprehension. The apparently more interactive process in production than comprehension could simply reflect a greater number of frequency-sensitive processing stages in production.

Keywords: frequency effect, semantic constraint, reading, production, bilingual

Language users have a strong intuition that words refer to the same thing whether they are being perceived or produced. This is particularly true for a given individual—each time a person says *apple*, hears someone say *apple*, or reads the word *apple*, it is likely that many of the same basic concepts come to mind (e.g., a sweet, crunchy fruit). Of course, an important piece of what makes communication through language possible is that words generally refer to more or less the same thing, regardless of who is producing or perceiving them. So imagine a conversation in which one person asks, “Do we have any apples?” and the other responds, “There’s lots of apples in the fridge.” The apples that the first person asked

about and the apples that the other mentioned in answering brought the same basic ideas to both people’s minds. Although conceptual overlap between people (and uses) is not always 100%, which leads to occasional miscommunications and sometimes disappointment (e.g., if one finds a green apple when hoping for a red one), most of the time there is sufficient overlap for communication to succeed. These observations highlight the common goal in perceived and produced word use: to communicate meaning regardless of modality (speaking, listening, writing, or reading).

There are some striking parallels between language production and comprehension. However, historically, researchers of language processing have divided themselves into those who study production and those who study comprehension. This division is not necessarily problematic in that similar (or different) conclusions may be reached independently. However, a complete understanding of language use will eventually require at least some direct comparisons of comprehension and production. For example, according to one view there are intimate connections between comprehension and production as they occur in dialogue, perhaps the most basic and natural form of language use (Pickering & Garrod, 2004), which necessarily involves rapid transition between production and comprehension. To the extent that common mechanisms are revealed in comprehension and production, this would facilitate such exchanges. On the other hand, differences that arise would need to be accommodated within any theoretical view of alignment between comprehension and production. Similarly, one intriguing and oft-cited theory proposes intimate connections between comprehension and production: In the motor theory of speech perception (Liberman, Cooper, Shankweiler, & Studdert-

Tamar H. Gollan, Department of Psychiatry, University of California, San Diego; Timothy J. Slattery, Diane Goldenberg, and Keith Rayner, Department of Psychology, University of California, San Diego; Eva Van Assche and Wouter Duyck, Department of Experimental Psychology, Ghent University, Ghent, Belgium.

The data in Experiment 1 were first reported as part of Diane Goldenberg’s undergraduate honor’s thesis. This research was supported by Grants HD050287, HD26765, and HD051030 from the National Institute of Child Health and Human Development, the National Institutes of Health, and by a postdoctoral fellowship from the Research Foundation–Flanders. The authors wish to thank Rachel Thayer for drawing the pictures used in Experiment 1 and Ken Forster, Merrill Garrett, and Nina Silverberg for years of discussion that provided the foundation for many of the ideas presented here.

Correspondence concerning this article should be addressed to Tamar H. Gollan, Department of Psychiatry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0948. E-mail: tgollan@ucsd.edu

Kennedy, 1967; Liberman & Mattingly, 1985), speech is perceived via top-down synthesis through the production system. On the other hand, some critics of the motor theory have instead argued the reverse: In the sensory theory of speech production, sensory representations of speech guide production (Lotto, Hickok, & Holt, 2009; Venezia & Hickock, 2009). Extreme versions of these proposals would posit that production and comprehension share the same representations and processes and would need to be modified to accommodate any differences observed between modalities.

Similarities Between Comprehension and Production

One obvious similarity between comprehension and production is that both processes exhibit robust frequency effects, such that words that are used frequently are processed more quickly than words that are not used as often (see e.g., Forster & Chambers, 1973; Oldfield & Wingfield, 1965; Rayner & Duffy, 1986). In both comprehension and production the relationship between counts of frequency and access time is logarithmic, such that small increases in frequency of use at the low end of the use scale have large effects on lexical access time but negligible effects beyond a certain high-frequency count. That is, frequency effects seem to be subject to a ceiling (see also the section entitled Constraint, Frequency, and Bilingual Effects) after which increased uses have little further effect on lexical accessibility (see e.g., Griffin & Bock, 1998; Howes & Solomon, 1951). A possibly related finding is that in both comprehension and production, repetition sometimes affects low-frequency words relatively more than high-frequency words (see e.g., Griffin & Bock, 1998; Oldfield & Wingfield, 1965; Raney & Rayner, 1995; Scarborough, Cortese, & Scarborough, 1977; though note that frequency effects are usually not eliminated by repetition: Forster & Davis, 1984; Jescheniak & Levelt, 1994; Navarrete, Basagni, Alario, & Costa, 2006; but see Griffin & Bock, 1998; Raney & Rayner, 1995). Frequency effects are often viewed as the signature of lexical access in both comprehension and production (see e.g., Almeida, Knobel, Finkbeiner, & Caramazza, 2007; Caramazza, 1997; Dell, 1990; Forster & Chambers, 1973; Graves, Grabowski, Mehta, & Gordon, 2007; Levelt, Roelofs, & Meyer, 1999; Rayner, 1998, 2009; Rubenstein, Garfield, & Millikan, 1970), although the locus of the frequency effect has been rigorously debated in both comprehension (see e.g., Murray & Forster, 2004) and production (see e.g., Jescheniak & Levelt, 1994; see Knobel, Finkbeiner, & Caramazza, 2008, who suggested that frequency likely affects multiple processing stages in production, and Kittredge, Dell, Verkuilen, & Schwartz, 2008, for evidence to this effect).

Another factor that affects comprehension and production similarly is context. Words that occur in high-constraint or predictable contexts (e.g., *Snow White took a bite of a poisonous apple*) are recognized/read and produced more quickly (Ehrlich & Rayner, 1981; Fischler & Bloom, 1979; Griffin & Bock, 1998; Rayner & Well, 1996; Stanovich & West, 1979; Tulving & Gold, 1963) than when they occur in a more neutral context (e.g., *The photographer took a picture of a large apple*). Early research on the effects of context on the processing of ambiguous words (those with more than one meaning) in comprehension revealed that both meanings are accessed initially, with effects of context arising only at a later processing stage (Swinney, 1979). In these studies, participants

listened to constraining sentences such as *The man was not surprised when he found several spiders, roaches, and other bugs in the corner of his room*. As they listened they were probed at a certain point in the sentence (e.g., at *bugs*), and then later in the sentence, for recognition of words related to both meanings of *bugs* (i.e., insects and spy-related devices). Initially, participants were faster to respond to words related to both meanings of ambiguous words (relative to unrelated words), whereas later on, facilitation was observed for only context-relevant meanings of the ambiguous word. This implies that participants had initial access to all word meanings but also that context led participants to suppress the irrelevant meaning by the end of the sentence (although the appropriate meaning can be selected quickly and, for ambiguous words with one highly dominant meaning, access may not be exhaustive; Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1986; see also Gernsbacher & Faust, 1991). Language production studies also revealed some evidence for exhaustive access to both meanings of ambiguous words. For example, speakers produce low-frequency words more quickly if there is a higher frequency homophone neighbor (e.g., a picture of a *bee* is named more quickly than a picture of an object with a frequency-matched name that does not have a high-frequency counterpart, such as *be*; see e.g., Jescheniak & Levelt, 1994; but see Caramazza, Costa, Miozzo, & Bi, 2001). Similarly, speakers benefit from the distractor *dance* when trying to name a picture of a toy *ball* because *dance* activates the Cinderella type of *ball*, which facilitates access to the sounds needed to produce the name *ball* (Cutting & Ferreira, 1999).

Differences Between Comprehension and Production

Despite the similarities between comprehension and production, there are both theoretical and empirical reasons to expect differences between them, starting with the distinct nature of the access problem in the two modalities. In comprehension, individual words must be distinguished from form-similar neighbors to identify the correct meaning, and this must often be done with an incomplete/noisy signal. Contrastingly, in production highly specific representations of meaning activate lexical candidates that must be selected from among meaning-related candidates. Thus, in production, lexical access is guided by meaning (and syntactic class) and ends with the identification of a lexical form, whereas in comprehension the process is reversed, going from lexical forms to semantics. Consequently, an ideal organizational scheme for the mental lexicon in comprehension would not work very efficiently in production, and vice versa (Fay & Cutler, 1977).

One subfield in the literature on language processing that has adhered somewhat less to the traditional division of investigators into those who study comprehension and those who study production is cognitive neuropsychology. Within this domain, there is some powerful evidence that separate systems serve comprehension and production (Rapp, 2001). The approach taken here is one of detailed investigation of the pattern of preserved versus impaired abilities in single-case studies of individuals with cognitive impairment. With this approach, the functional architecture of the normal cognitive system is revealed through in-depth analysis of cognitive deficits and attempts to identify the locus (or loci) of impairment in well-articulated language-processing models. To identify the nature of a deficit, it is necessary to examine perfor-

mance in every modality of input and output and at multiple levels of processing and representation (e.g., sublexical, lexical, and semantic). The theoretical assumptions inherent to this approach are that the cognitive architecture is the same across individuals, that both the functional and anatomical (brain) architecture are modular to at least some degree, and that brain damage does not introduce new compensatory mechanisms that did not exist prior to brain damage (Coltheart, 2001). In this literature, double dissociations between modalities have been identified. These were apparent in the classical distinction between Broca's and Wernicke's aphasia but also in more recently identified and specific deficits, for instance with patients who experience problems when writing but not reading and vice versa, or difficulties understanding spoken but not written words despite intact auditory acuity (for a review, see Hillis, 2001). The dissociations observed in this literature imply separate systems for comprehension and production, as well as separate lexicons (and access) for each modality (speaking, writing, reading, and listening; Caramazza, 1997). This type of architecture may not be very economical, but it is not the most conceivable extreme; the line is drawn at the level of meaning where it is assumed that the different modalities ultimately contact a shared semantic system (i.e., the notion of modality-specific semantics is not supported; Shelton & Caramazza, 2001).

Other evidence for distinct processes of lexical access in comprehension versus production concerns the fact that form-related neighbors (i.e., neighborhood density) consistently facilitate access in production but sometimes delay access in comprehension (Dell & Gordon, 2003; Slattery, 2009; for review see Andrews, 1997). Conversely, semantically related words facilitate comprehension (for review see Neely, 1991) but function like competitors and delay lexical access in production (see e.g., Lupker, 1979; Schriefers, Meyer, & Levelt, 1990; but see Mahon, Costa, Peterson, Vargas, & Caramazza, 2007). Such differences have led some to conclude that distinct representations of lexical form serve comprehension and production (see e.g., Roelofs, 2003; Zwitserlood, 2003), with one set serving as the end point for production processes and the other as the point of entry for comprehension. In this view, comprehension and production overlap only at the level of meaning. Although some models do posit a second amodal layer of lexical representations, this level of representation would subservise lexical access itself only in production, not in comprehension (Levelt et al., 1999).

Interactions Between Frequency and Constraint

Interestingly, both in comprehension (Forster & Chambers, 1973; Murray & Forster, 2004; Rayner, 1998; Reichle, Pollatsek, Fisher, & Rayner, 1998) and production (Jescheniak & Levelt, 1994; Levelt et al., 1999) it has been assumed that frequency effects arise primarily at the point at which lexical forms are accessed. Thus, even if separate form lexicons might serve comprehension and production, both are considered to be sensitive to frequency. Moreover, it has been debated in both comprehension and production research whether processing stages during lexical access are interactive or discrete. And, in both fields, a number of investigators have considered the joint effects of frequency and constraint as a means to decide between models (with interactions supporting interactivity and a lack of interactions supporting seriality). The logic behind this was that frequency and semantic

constraint effects should not interact if (a) lexical access proceeds in discrete stages and (b) frequency and semantic constraint effects arise in different processing stages.

The placement of frequency effects at the locus of retrieving lexical forms necessarily separates frequency from constraint effects in both comprehension and production. This is because forms will be accessed relatively late in production, after resolution of competition between semantically related candidates, but early in comprehension, which necessarily begins with the analysis of lexical form and meaning as the end point of the process. On the other hand, frequency effects might not be limited to the point of accessing lexical forms (e.g., if retrieval of lexical semantics were also frequency-sensitive), in which case interactions could arise without interactivity. Thus, the locus or loci of the frequency effect itself in the two modalities could be another source of apparent differences between them.

Some of the studies that investigated constraint and frequency effects reported interactions between them such that low-frequency targets benefited more from constraining context than did high-frequency targets (see e.g., Griffin & Bock, 1998; Stanovich & West, 1979; Van Petten & Kutas, 1990). However, others studies found no interaction, particularly in the literatures on written word recognition and reading, in which constraining context affected low- and high-frequency words equally (see e.g., Forster, 1981; Rayner, Ashby, Pollatsek, & Reichle, 2004). The identification of a Frequency \times Constraint interaction in production, but not in comprehension, raises the possibility that mechanisms of lexical access interact with top-down factors more in production than in comprehension or that access to lexical semantics (in which constraint effects could operate) is frequency-sensitive for production but not comprehension (this could entail either separate lexical-semantic representations for comprehension and production or just differently weighted connections to lexical-semantic representations; frequency-sensitive going from concepts in but not frequency-sensitive going from lexical form representations in). However, to date only one study has investigated this interaction in production (Griffin & Bock, 1998), and no study has compared production and comprehension directly. Thus, the apparent difference between production and comprehension could instead reflect cross-study differences between materials, tasks, and participants or some combination of these.

Constraint, Frequency, and Bilingual Effects

Constraint and frequency have also figured prominently in the literature on bilingual language processing (another subfield of psycholinguistics that exhibits a less obvious division of researchers into those who study comprehension versus production). A widely accepted assumption in models of bilingual language processing is that bilinguals cannot shut off one language completely to effectively function like monolinguals (see e.g., Dijkstra & van Heuven, 2002; Duyck, Van Assche, Drieghe, & Hartsuiker, 2007; Hermans, Bongaerts, de Bot, & Schreuder, 1998; Van Assche, Duyck, Hartsuiker, & Diependaele, 2009; for reviews see Costa, 2005; Kroll, Bobb, Misra, & Guo, 2008; La Heij, 2005; Sebastián-Gallés & Kroll, 2003). The presence of dual-language activation could place bilinguals at a disadvantage when compared with monolinguals, particularly in the domain of language production, for which a single language must be chosen for response, so that

translation equivalents could function as competitors. Some evidence to support this notion is that bilinguals exhibit small but significant disadvantages relative to monolinguals on a number of production tasks, including reduced category fluency (see e.g., Gollan, Montoya, & Werner, 2002; Rosselli et al., 2000), higher rates of tip-of-the-tongue state retrieval failures (see e.g., Gollan & Silverberg, 2001), slower picture-naming times (see e.g., Gollan, Montoya, Fennema-Notestine, & Morris, 2005; Ivanova & Costa, 2008), and slower sentence production times (MacKay & Flege, 2004). Similar bilingual disadvantages have been reported in language comprehension tasks such as lexical decision (see e.g., Ransdell & Fischler, 1987). Some have suggested that semantically constraining context can reduce the extent to which both languages are active (Schwartz & Kroll, 2006), at least during later stages of word comprehension (Libben & Titone, 2009; but see Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, in press, for cross-lingual interactions with a high constraint context during all stages of word comprehension). The proposal that bilinguals effectively function like monolinguals in a semantically constraining context leads to the prediction that bilingual disadvantages should be reduced under such conditions. Taking this same logic a step further, the bilingual disadvantage might even be reduced in a neutral context relative to no context (e.g., picture-naming of objects in isolation) if a monolingual context of any kind (e.g., a unilingual sentence) can serve to reduce dual-language activation. These predictions are based on the assumption that dual-language activation leads bilinguals to respond more slowly than do monolinguals.

An additional cause of bilingual disadvantages might reflect a simpler and emergent property of frequency effects and patterns of bilingual language use. By virtue of using each language only part of the time, bilinguals have used word forms particular to each language relatively less frequently than have monolinguals (i.e., bilingualism entails a frequency lag; Gollan, Montoya, Cera, & Sandoval, 2008; Gollan et al., 2005; Gollan et al., 2002; Gollan & Silverberg, 2001; Sandoval, Gollan, Ferreira, & Salmon, 2010; for similar ideas see Ivanova & Costa, 2008; Lehtonen, & Laine, 2003; Mägiste, 1979; Nicoladis, Palmer, & Marentette, 2007; Pearson, Fernandez, Lewedeg, & Kimbrough, 1997; Ransdell & Fischler, 1987). Supporting this account, 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. Similarly, slowing related to language dominance is particularly large for retrieval of low-frequency words in both comprehension (Duyck, Vanderelst, Desmet, & Hartsuiker, 2008) and production (Gollan et al., 2008; but see Ivanova & Costa, 2008). In this view, lexical representations evolve toward ceiling-level, stable end states, and the monolingual lexicon is closer to these ceiling levels of lexical accessibility than is the bilingual lexicon (for further explanation, see Gollan et al., 2008; Murray & Forster, 2004). As language use and proficiency increases, low-frequency words begin to catch up with high-frequency words in levels of activation, and frequency effects should become smaller.

Thus, (a) monolinguals exhibit smaller frequency effects than do bilinguals tested in their dominant language and (b) bilinguals exhibit smaller frequency effects when tested in their dominant language than in their nondominant language. On this frequency-lag (a.k.a. the “weaker links”; Gollan et al., 2008, p. 787) hypoth-

esis, bilingual disadvantages should become smaller only when frequency effects also become smaller. For example, the bilingual disadvantage might be reduced with semantic constraint only for production (in which frequency effects are reduced with semantic constraint) and not for comprehension (if frequency effects remain unchanged in semantically constraining context). On the other hand, if bilingual disadvantages are reduced by manipulations that have no influence on the frequency effect, then this would imply that something other than a frequency lag causes the bilingual disadvantage.

To summarize, there are reasons to expect both similarities and differences between the processes of lexical access in comprehension and production. Further considering these similarities and differences seems likely to advance both specific fields, and there are some informative discussions along these lines in the literature (Fay & Cutler, 1977; Garrett, 2001; Hillis, 2001; Jurafsky, 2003; Schiller & Meyer, 2003; Treiman, Clifton, Meyer, & Wurm, 2003). However, the almost complete separation between production and recognition research has precluded a direct empirical comparison of comprehension and production in psycholinguistic research (for a noteworthy exception, see a recent fMRI study revealing shared brain regions for processing meaning across modalities; Awad, Warren, Scott, Turkheimer, & Wise, 2007). In the present study, we therefore considered the joint effects of frequency, contextual constraint, and bilingualism on production (picture naming) and comprehension (reading and lexical decision), with the goal of distinguishing processes that are modality-specific from those that are general to language processing. Because there have been far fewer attempts to investigate Frequency \times Constraint interactions in language production than in comprehension, we begin with production using a picture-naming task (Experiment 1). Subsequently, we use the same materials (only substituting words for pictures) to investigate the effects of frequency, constraint, and language proficiency in visual word recognition, operationalized through lexical decision and eye-tracking during reading (Experiment 2). We then end with direct comparisons of the observed effects across modalities (comparison of Experiments 1 and 2).

In both Experiments 1 and 2, our primary goal was to test for a Frequency \times Constraint interaction and to determine whether the pattern of results observed generalizes to processing at all levels of proficiency. From the perspective of the frequency-lag hypothesis, the manipulation of English proficiency level effectively functions like a tool that allows a between-subjects manipulation of frequency. That is, if bilinguals use lexical forms less frequently than do monolinguals, then this provides an additional opportunity to test for interactions with constraint in both comprehension and production. We predicted that, on the basis of the frequency-lag hypothesis, bilingual effects might resemble frequency effects (e.g., interacting with constraint only in production and not in comprehension), especially at the lowest level of English proficiency, at which the frequency of use difference between bilinguals and monolinguals is the greatest. A secondary goal was to explore the nature of the bilingual disadvantage, investigating whether it occurs in both comprehension and production and whether it is reduced by semantically constraining or neutral context.

Experiment 1: Picture Naming

Method

Participants. Fifty monolingual English speakers, 50 Spanish–English bilinguals, and 45 Dutch–English bilinguals participated. Monolinguals and Spanish–English bilinguals were undergraduates at the University of California, San Diego, and received course credit for their participation; Dutch–English bilinguals were undergraduates at Ghent University in Belgium and either received course credit or were paid for their participation.¹ These three groups of participants were chosen for their expected three-tiered level of relative English proficiency, with the highest English proficiency among monolinguals, followed by English-dominant Spanish–English bilinguals, and finally the Dutch-dominant Dutch–English bilingual group, which was expected to have the lowest level of English proficiency. Spanish–English bilinguals were expected to have a higher level of English proficiency than were Dutch–English bilinguals because of their relatively earlier age of acquisition (AoA) of English and their years of immersion in an English-speaking environment. Conversely Spanish–English bilinguals were expected to have a lower level of English proficiency than were monolinguals because of their relatively later AoA of English and their relatively reduced use of English relative to monolinguals (Gollan et al., 2008). Participants completed a language history questionnaire in which they provided subjective ratings of language proficiency, and as objective measures of English proficiency participants completed a 40-item multiple choice vocabulary test (Shipley, 1946). Table 1 displays a summary of the participant characteristics.

Materials. Table 2 contains a summary of the materials' characteristics. A list of 45 pairs of high- and low-frequency names of objects was constructed using a standard frequency corpus (Davies, 2007). High- and low-frequency name pairs differed maximally in frequency count while being matched for number of letters and initial phoneme (e.g., for every *apple* an *arrow*). The pictures were black-and-white line drawings produced by an artist who was instructed to make the drawings as simple as she could to elicit the target names. Pictures with names that could not be easily depicted with a simple line drawing were replaced. The high- and low-frequency targets were not matched for codability, but we tried as much as possible to select targets with just one name. An estimate of codability can be obtained by counting the number of alternative correct names that speakers produced (e.g., *skates* and *ice skates*, *statue* and *Statue of Liberty*, *basket* and *picnic basket*). Virtually all items had between zero and two alternative names. About two thirds of the items in both frequency groups had no alternative names. The average number of alternative names was slightly higher for low- than for high-frequency names, but this difference was not significant ($ps = .33, .13, \text{ and } .09$ for monolinguals, Spanish–English bilinguals, and Dutch–English bilinguals, respectively). A slight difference between high- and low-frequency targets emerged when considering the proportion of cases in which speakers actually used an alternative name (instead of the experimentally intended target name). For high-frequency targets this occurred just over 1% of the time, but for low-frequency targets it was about 4% of the time (small but significant differences for all speaker groups; $ps = .03$ for monolinguals and Spanish–English bilinguals, $p = .05$ for Dutch–English bilin-

guals). Ten of the target names (five high- and five low-frequency) were cognates in Spanish and English (e.g., *paper* is *papel* in Spanish), and 41 (22 high- and 19 low-frequency) were cognates in Dutch and English (e.g., *apple* is *appel* in Dutch).² For each target a high- and low-constraint sentence context was created. High-constraint sentences were designed to strongly elicit the target word (e.g., *Snow White ate a poisoned _____*), whereas low-constraint sentences were designed to be congruent with the target completion while also allowing many other possible completions (e.g., *The artist painted the bowl and the _____*). To determine the extent to which sentences created high- versus low-constraint contexts, 20 undergraduates at the University of California, San Diego, who did not participate in the experiment were asked to provide one-word completions for each sentence (target completions for high-constraint sentences were produced over 85% of the time, and for low-constraint sentences less than or equal to 3% of the time; see Table 2). Sentences that were not high or low in constraint were replaced and renormed with 20 additional undergraduates who also did not participate in the experiment. High-constraint sentences were slightly longer than low-constraint sentences, but length did not differ across the frequency contrast (both $ps \geq .24$). High- and low-frequency pairs were divided into three lists of 15 pairs (thus each list contained 30 targets, of which 15 were high frequency, 15 low frequency).

Procedure. Each participant completed all three lists but in only one of each of the three conditions (no context, low-constraint context, high-constraint context). Participants saw each picture only once. Assignment of material list to experimental condition was counterbalanced such that each of the three lists was presented in each of the three conditions a roughly equal number of times between participants. The no-context condition was presented separately from the context condition, and the two context conditions were presented together with low- and high-constraint sentences. The order of testing with versus without context was counterbalanced across participants, and materials were presented in a different random order for each participant.

The materials were presented using PsyScope 1.2.5 (Cohen, MacWhinney, Flatt, & Provost, 1993) on a Macintosh computer with a 17-in. monitor for the monolinguals and Spanish–English bilinguals and using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) for the Dutch–English bilinguals. Care was taken to ensure that stimulus characteristics (e.g., color, resolution, and picture sizes) were identical across labs. Participants wore headset

¹ Note that Belgium consists of a Dutch-speaking community and a smaller French-speaking community. The present participants belong to the first group and are late Dutch–English bilinguals, because they have been exposed to English through popular media (music, television, movies, Internet, etc.) and received formal English-language instruction in school starting from age 14.

² The greater number of cognates with Dutch reflects the fact that these languages are more similar (both Germanic) and would reduce to some extent the degree to which Dutch–English bilinguals would be disadvantaged given that cognates are easier for bilinguals to produce (see e.g., Costa, Caramazza, & Sebastián-Gallés, 2000). Although we did not manipulate Dutch–English cognate status in our materials, we considered whether some of the effects of greatest interest might be modulated by cognate status (e.g., the Frequency \times Constraint interaction in Experiments 1 and 2) and found no evidence for any influence of cognate status.

Table 1
Participant Characteristics in Experiments 1 (Picture Naming) and 2 (Reading)

Characteristic	Picture naming						Reading					
	Monolinguals (n = 50)		Spanish–English bilinguals (n = 50)		Dutch–English bilinguals (n = 45)		Monolinguals (n = 36)		Spanish–English bilinguals (n = 36)		Dutch–English bilinguals (n = 27)	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Age in years	20.3	1.6	20.0	3.5	19.3	1.5	20.4	1.5	20.3	1.6	19.4	1.5
Age of acquiring English	0.1	0.3	3.7	2.9	10.8	2.5	0.2	0.6	3.4	3.2	10.9	2.8
English self-rated speaking ^a	7.0	^b	6.5	0.8	5.1	0.8	7.0	^b	6.6	0.6	4.9	0.9
English self-rated reading ^a	7.0	0.1	6.5	0.8	5.5	0.7	6.9	0.2	6.7	0.5	5.6	0.7
Percentage daily use of English	98.4	4.2	83.1	14.0	6.3	7.1	99.9	0.3	81.7	16.6	6.0	6.5
Percentage daily reading in English	97.0	4.4	82.0	12.9	15.1	13.0	98.3	2.7	78.5	13.4	13.9	8.4
Percentage correct on English vocabulary test (Shipley, 1946)	77.4	6.5	72.7	7.4	61.7	7.7	78.0	10.6	72.4	6.5	57.4	8.0

^a Proficiency level based on self-ratings on a scale ranging from 1 (little to no knowledge) to 7 (like a native speaker). ^b All monolinguals rated themselves at ceiling on this question.

microphones that recorded naming response times, and responses were audiotaped for later verification of accuracy. Participants were instructed to name pictures as quickly and accurately as possible. An experimenter was present throughout the experiment to record responses and the accuracy of voice key triggers.

In the no-context condition, trials began with a central fixation point (“+”) that was presented for 500 ms and followed by a picture in a 6-cm × 6-cm area in the center of the computer screen. The picture disappeared once the voice key was triggered. A blank screen was presented for 1.5 s between trials. Participants initiated subsequent trials by pressing the space bar.

In the context conditions, participants were instructed to silently read the sentences, which were presented one word at a time using self-paced visual presentation. Words appeared in the center of the screen, and participants pressed the space bar to move from one word to the next so that each word replaced the previous word in the center of the screen. At the end of each sentence, a picture of an object that was meant to complete the sentence appeared, and participants were instructed to name the pictures as quickly and accurately as possible. Pictures remained on the screen until the voice key was triggered. The next trial began 1.5 s later. To encourage participants to pay attention to the sentences, the word *REMEMBER* appeared on the screen after the voice key was triggered on six of the 60 trials. Three high-constraint and three low-constraint sentences were preselected for this purpose. At the prompt, participants were asked to recall the sentence aloud, and the experimenter noted the accuracy. These responses were coded as exact or close (which were counted as correct) or as incorrect; the rate of correct responses was about 80% or higher in both Experiments 1 and 2 for all participant groups.

Results

Reaction times (RTs) under 250 ms and over 5,000 ms were excluded from analyses. In addition, mean RTs were calculated for each participant collapsed across condition, and RTs more than 2.5 standard deviations above each participant’s mean were excluded from analyses. These trimming procedures excluded 2.2% of the

monolinguals’, 2.6% of the Spanish–English bilinguals’, and 2.1% of the Dutch–English bilinguals’ correct and valid RTs. In anticipation of comparisons of the data to those reported in Experiment 2, the no-context mean correct response RTs were analyzed separately in a 2 × 3 analysis of variance (ANOVA) with frequency (high, low) as a repeated measures factor in the participants analyses and group (monolingual, Spanish–English, Dutch–English) as a repeated factor in the items analyses. The in-context data were submitted to a 2 × 2 × 3 ANOVA with frequency (high, low) and constraint (high, low) as repeated measures factors and group (monolingual, Spanish–English, Dutch–English) as a between-subjects factor in the participants analyses (F_1). In the items analyses (F_2), constraint and group were repeated measures factors, and frequency was between items. There were five items (*braid*, *faucet*, *skunk*, *stool*, and *truck*) with missing data points (naming errors) in the no-context and low-constraint-context conditions for Dutch–English bilinguals and one item with missing data for Spanish–English bilinguals in the low-constraint condition (which is consistent with the expected relative English proficiency of these two groups). Thus, the degrees of freedom vary slightly from the 2 × 2 × 3 to the 2 × 3 ANOVAs reported in the next sections depending on the number of items available for each analysis. The results are shown in Figure 1 (see also the summary of results from both Experiments 1 and 2 shown in Table 4).

Picture naming out of context. Speakers named pictures with high-frequency names faster than pictures with low-frequency names, $F_1(1, 142) = 74.65, MSE = 16,917, \eta_p^2 = .35, p < .01; F_2(1, 83) = 11.41, MSE = 135,065, \eta_p^2 = .12, p < .01$, and monolinguals named pictures more quickly than did bilinguals, $F_1(2, 142) = 33.42, MSE = 63,213, \eta_p^2 = .32, p < .01; F_2(1, 83) = 105.92, MSE = 263,879, \eta_p^2 = .56, p < .01$. As previously reported for out-of-context picture naming (Gollan et al., 2008; Ivanova & Costa, 2008), bilinguals exhibited significantly larger frequency effects than did monolinguals, $F_1(2, 142) = 6.99, MSE = 16,917, \eta_p^2 = .09, p < .01; F_2(1, 83) = 5.71, MSE = 46,183, \eta_p^2 = .06, p = .02$. Additionally, the finding of greater frequency effects in bilinguals than monolinguals was

F1

Table 2
Name Characteristics for Items in High- and Low-Frequency Pairs in Experiments 1 and 2, and t Tests for the Difference Between High- and Low-Frequency Targets

Frequency	Time magazine corpus ^a	CELEX total ^a	CELEX written ^a	CELEX spoken ^a	Kucera & Francis ^a	Length in syllables	Length in phonemes	Length in letters	Orthographic N	Phonological N	High-constraint norming ^b	Low-constraint norming ^b
High												
M	79.8	85.4	88.2	48.9	82.0	1.4	4.2	5.4	3.7	10.8	87.5	3.0
SD	78.3	75.6	76.0	90.0	76.8	0.5	1.1	0.8	3.8	7.2	14.0	6.0
Max	420.6	403.7	391.4	560.8	350.0	3.0	9.0	9.0	19.0	29.0		
Min	14.9	12.7	13.7	0.8	6.0	1.0	3.0	4.0	0.0	0.0		
Low												
M	9.3	9.8	10.4	3.3	9.4	1.4	4.4	5.4	3.3	9.7	85.5	2.0
SD	5.1	7.6	8.0	4.0	8.7	0.5	1.0	0.8	3.2	7.4	13.0	4.5
Max	21.2	40.1	42.2	17.7	34.0	3.0	8.0	9.0	13.0	37.0		
Min	1.3	0.0	0.0	0.0	0.0	1.0	3.0	5.0	0.0	0.0		
r ^c	6.03	6.67	6.83	3.40	6.31	0.20	0.92	0.14	0.45	0.72	0.67	0.90
p	.00	.00	.00	.00	.00	.84	.36	.89	.65	.47	.51	.37

Note. Time Magazine Corpus is from Davies (2007); CELEX is from Baayen, Piepenbrock, & Glikers (1995); and Kucera & Francis refers to Kucera & Francis (1967). N = neighborhood density; Max = maximum; Min = minimum.

^a Values represent frequency per million. ^b Percentages of people (n = 20) who entered the target word as the sentence completion in the constraint-norming study. For both high- and low-frequency words the maximum value for high-constraint sentences was 20 and minimum value was 10, and for low-constraint sentences the maximum value was 0, and the minimum value was 0. ^c dfs = 88.

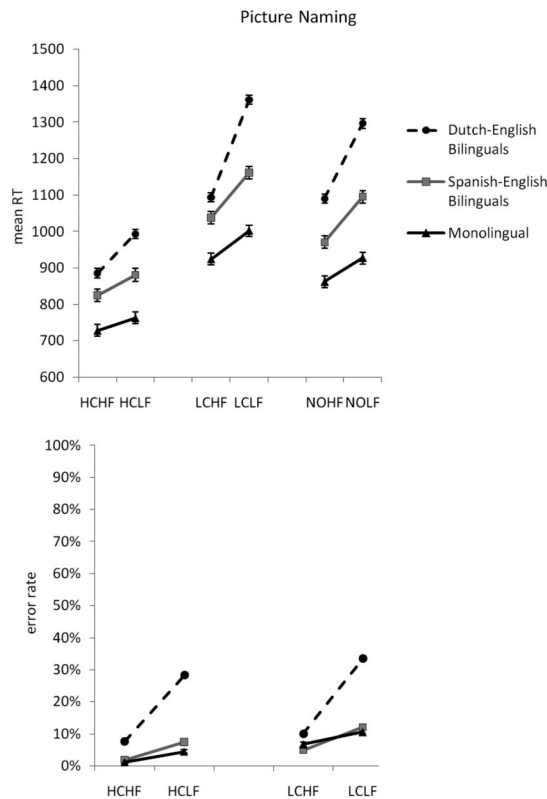


Figure 1. Picture-naming times in milliseconds (top panel) and error rates (bottom panel) in each of the six conditions for each participant type in Experiment 1. Means are taken from participants analyses, and error bars are standard errors for each condition. RT = reaction time; HC = high semantic constraint; LC = low semantic constraint; HF = high frequency; LF = low frequency; NO = no context.

robust for both bilingual groups. Spanish-English bilinguals exhibited a greater frequency effect than did monolinguals, $F_1(1, 98) = 6.06$, $MSE = 7,301$, $\eta_p^2 = .06$, $p = .02$; $F_2(1, 88) = 4.47$, $MSE = 10,282$, $\eta_p^2 = .05$, $p = .04$, and the frequency effect was even larger in Dutch-English bilinguals, who exhibited a marginally greater frequency effect than that seen in Spanish-English bilinguals, $F_1(1, 93) = 3.41$, $MSE = 23,071$, $\eta_p^2 = .04$, $p = .07$; $F_2(1, 85) = 2.65$, $MSE = 41,459$, $\eta_p^2 = .03$, $p = .11$.

Error analyses for picture naming out of context exhibited a pattern of results similar to that for the RT data, with robust main effects of frequency and group (all $ps < .01$) and robust Frequency \times Group interactions (both $ps < .01$; excluding Dutch-English bilinguals, this interaction was just marginally significant; $p = .11$).

Picture naming in context. Speakers named pictures with high-frequency names faster than pictures with low-frequency names, $F_1(1, 142) = 138.17$, $MSE = 12,902$, $\eta_p^2 = .49$, $p < .01$; $F_2(1, 85) = 12.62$, $MSE = 145,862$, $\eta_p^2 = .13$, $p < .01$, and more quickly in high- than low-constraint contexts, $F_1(1, 142) = 478.14$, $MSE = 18,997$, $\eta_p^2 = .77$, $p < .01$; $F_2(1, 85) = 143.05$, $MSE = 69,847$, $\eta_p^2 = .63$, $p < .01$. In addition, monolinguals named pictures faster than did bilinguals, $F_1(2, 142) = 20.80$, $MSE = 120,294$, $\eta_p^2 = .23$, $p < .01$; $F_2(1, 85) = 105.47$, $MSE = 55,368$, $\eta_p^2 = .55$, $p < .01$.

As previously reported for monolingual picture naming (Griffin & Bock, 1998), frequency effects were smaller in high- than low-constraint contexts, $F_1(1, 142) = 41.87$, $MSE = 7,050$, $\eta_p^2 = .23$, $p < .01$; $F_2(1, 85) = 3.91$, $MSE = 69,847$, $\eta_p^2 = .04$, $p = .05$. In addition, this interaction was not driven by one participant group; monolinguals ($p = .02$) and both bilingual groups (both $ps < .01$) exhibited a significant Frequency \times Constraint interaction (some of these planned comparisons were just marginally significant by F_2 ; see Table 4). As found in out-of-context naming, slowing related to bilingualism was significantly greater in production of low- than high-frequency picture names; that is, bilinguals exhibited significantly larger frequency effects than did monolinguals, $F_1(2, 142) = 16.92$, $MSE = 12,902$, $\eta_p^2 = .19$, $p < .01$; $F_2(1, 85) = 7.03$, $MSE = 55,368$, $\eta_p^2 = .08$, $p = .01$, mainly due to slower responses for pictures with low-frequency names. This interaction was primarily driven by low-constraint contexts, in which Spanish–English bilinguals exhibited a greater frequency effect than did monolinguals, though in context this was only significant by participants, $F_1(1, 98) = 4.67$, $MSE = 5,998$, $\eta_p^2 = .04$, $p = .04$; $F_2 < 1$, and Dutch–English bilinguals exhibited an even greater frequency effect—significantly greater than that seen in Spanish–English bilinguals, $F_1(1, 93) = 17.44$, $MSE = 13,981$, $\eta_p^2 = .16$, $p < .01$; $F_2(1, 85) = 4.43$, $MSE = 74,408$, $\eta_p^2 = .05$, $p = .05$. In high-constraint contexts, Spanish–English bilinguals did not exhibit greater frequency effects than did monolinguals (both $ps \geq .25$), and Dutch–English bilinguals did not exhibit a greater frequency effect than did Spanish–English bilinguals (both $ps \geq .12$). Thus, both frequency effects and the bilingual disadvantage were modulated by context, as would be expected on a frequency-lag account.

With bilingualism considered as a kind of frequency manipulation, there is some indication that bilinguals benefited more from constraint than did monolinguals, $F_1(2, 142) = 3.16$, $MSE = 18,997$, $\eta_p^2 = .04$, $p = .045$; $F_2(1, 85) = 4.86$, $MSE = 45,670$, $\eta_p^2 = .05$, $p = .03$, and this interaction was driven by Dutch–English bilinguals' responses to low-frequency words; the Constraint \times Participant Group interaction was no longer significant when the Dutch–English bilinguals were excluded from the analysis (both $ps \geq .20$). There was also some evidence for a three-way Frequency \times Constraint \times Participant Type interaction—significant only by participants, $F_1(2, 142) = 6.23$, $MSE = 7,050$, $\eta_p^2 = .08$, $p < .01$; $F_2(1, 85) = 1.31$, $MSE = 45,670$, $\eta_p^2 = .02$, $p = .26$ —such that bilinguals at the lowest level of English proficiency benefited more from semantic constraint than did monolinguals. These results are generally consistent with the frequency-lag account (in that both frequency effects and the bilingual disadvantage were attenuated by constraint context).

The error data exhibited a pattern of results similar to that for the RT data. There were robust main effects of frequency, constraint, and group (all $ps < .01$), as well as robust Frequency \times Group interactions (both $ps < .01$) that remained significant after excluding the Dutch–English bilinguals (both $ps \leq .03$). None of the other two-way interactions, nor the three-way interaction, was significant in the error data.

To constrain the interference account of bilingual disadvantages, we asked whether context alone (low semantic constraint) was sufficient to reduce the bilingual disadvantage. To this end, we compared out-of-context (bare picture naming) to low-constraint-context naming responses in two analyses (Spanish–English bilin-

guals vs. monolinguals, and Dutch–English bilinguals vs. monolinguals). As reported earlier, there were main effects of frequency and group (all $ps < .01$). In addition, there was a main effect of context such that response times were slower following sentence context relative to out-of-context naming (all $ps \leq .01$), possibly reflecting ongoing contextual sentence analysis during the process of picture naming. However, there was no indication that low-constraint context affected Spanish–English bilinguals to a greater extent than monolinguals and also no differential effect of context on Dutch–English bilinguals compared with monolinguals (all $F_s < 1$). Thus, it appears that semantic constraint (not just sentence context in the target language) is needed to reduce the disadvantage associated with processing a second and nondominant language.

Discussion

To summarize the results of Experiment 1, picture-naming times exhibited robust main effects of frequency, constraint, and bilingual status. Also, these effects interacted, such that speakers accessed low-frequency names more slowly than high-frequency names, particularly in out-of-context or low-constraint contexts (replicating Griffin & Bock, 1998). An additional finding in Experiment 1, which could be taken as further support for the notion of a Frequency \times Constraint interaction (assuming second-language use entails a frequency lag), was that Dutch–English bilinguals benefited more than did monolinguals from high-constraint semantic context. Consistent with the notion of a frequency lag for bilinguals in production, slowing related to bilingualism (Spanish–English bilinguals vs. monolinguals), and second-language production (Dutch–English bilinguals vs. monolinguals) was greater for producing low- than high-frequency names, again particularly in low-constraint and out-of-context picture naming (replicating and extending Gollan et al., 2008). Moreover, the presence of low-constraint semantic context alone was not sufficient to reduce the bilingual disadvantage (i.e., low relative to no context). Thus, context influenced the bilingual disadvantage only where it also reduced frequency effects, a result that is consistent with the frequency-lag account.

The finding that semantically constraining context reduced both frequency effects and slowing related to second-language production supports the proposal of an overlapping locus for the effects of (a) frequency, (b) bilingual disadvantages, and (c) language dominance. Two previous picture-naming studies compared frequency effects between monolinguals and bilinguals in their dominant language, as well as within bilinguals between their dominant and nondominant languages. In both cases, bilinguals exhibited a larger frequency effect than did monolinguals when tested in their dominant language (Gollan et al., 2008; Ivanova & Costa, 2008). However, in one study (Gollan et al., 2008), as in Experiment 1, frequency effects were also larger in the nondominant than in the dominant language, whereas in the other study (Ivanova & Costa, 2008) frequency effects were equally large in the bilinguals' two languages. Thus, the two studies produced similar findings with respect to the bilingual/monolingual contrast but discrepant results with respect to Language Dominance \times Frequency interactions. A number of differences could explain the slightly different pattern of results across studies, but the current data rule out some of these possible explanations.

Gollan et al. (2008) had a within-subjects language-dominance contrast (comparing dominant English responses to nondominant Spanish responses within the same group of participants), whereas Ivanova and Costa (2008) had a between-subjects language-dominance contrast (they held language constant and tested two different groups of speakers who were either dominant in Spanish or dominant in Catalan and nondominant in Spanish). The current design was the same as in Ivanova and Costa and used a between-subjects manipulation of language dominance, holding the language of testing (i.e., English) constant. However, we still obtained the same pattern of results as did Gollan et al. (a larger frequency effect in the nondominant than in the dominant language). Thus, it would seem that the more consistently obtained pattern is a larger frequency effect for the nondominant language. This result is also predicted on theoretical grounds: If reduced frequency of use leads to larger frequency effects across the monolingual/bilingual contrast, then reduced use of a nondominant language should also lead to larger frequency effects across the language-dominance contrast.

The difference between studies may reflect the fact that Spanish and Catalan are similar languages and equally dominant in the environment where Ivanova and Costa (2008) conducted their study; indeed, the difference in naming RTs between dominant Spanish and nondominant Spanish was much smaller in Ivanova and Costa (about 50 ms for both high- and low-frequency targets; see repetition 1 in Table 3, pp. 794) when compared with the differences between dominant English and nondominant Spanish in Gollan et al. (about 250 ms for high-frequency targets and 500 ms for low-frequency targets; see Figure 1 of Gollan et al., 2008, pp. 795) and when compared with the differences between dominant English and nondominant English in Experiment 1 of the current study (about 120 ms for high-frequency targets and 201 ms for low-frequency targets in out-of-context naming; see Figure 1). Thus, the increase in the magnitude of the frequency effect (a difference score) may become detectable only at a certain point, where language-dominance differences are sufficiently robust.

Returning to the larger implications of the data reported here, Experiment 1 replicated the previously reported Frequency \times Contextual Constraint interaction for spoken-language production and also demonstrated that this interaction arises at three different levels of English proficiency. In addition, we replicated the Frequency \times Bilingualism interaction for production and showed that the bilingual disadvantage becomes smaller only in conditions where the frequency effect also becomes smaller and that low-constraint target-language sentence context alone is not sufficient to reduce the disadvantage (only semantically constraining context). In Experiment 2, we tested whether frequency, constraint, and bilingualism exhibit similar effects in written-word recognition and reading using the same materials as in Experiment 1 and participants from the same language groups.

Experiment 2: Reading and Lexical Decision

As reviewed earlier, several studies of reading comprehension have identified significant effects of frequency and constraint but without any interaction between them (see Rayner et al., 2004, for a summary). Thus, given the findings of Griffin and Bock (1998) and the replication thereof in Experiment 1, there is an apparent difference between effects identified across modalities. In Exper-

iment 2, we aimed to establish whether this apparent contrast could be replicated in a single study or whether between-study variation in methodology might have spuriously introduced the contrast. Because the bulk of studies investigating the joint effects of frequency and contextual constraint in language comprehension involved reading, we contrasted spoken production with reading. Note that in comparing comprehension and production it is ultimately important to also compare auditory word recognition with production. However, because of possible involvement of production processes in auditory recognition, reading likely provides a stronger starting point for detecting any possible differences between comprehension and production. For example, some have proposed that every spoken utterance is first launched through the auditory comprehension system prior to overt production to check for errors (the *monitor* in language production; Levelt, 1989). If so, it would be more difficult to detect differences between comprehension and production, even if they did exist. Thus, reading may provide an estimate of comprehension processes in a manner as divorced as possible from production (though note that even here similarities could arise because of inner speech during reading).

Method

Participants. Thirty-six monolingual English speakers, 36 Spanish–English bilinguals, and 27 Dutch–English bilinguals who did not participate in Experiment 1 were recruited for participation in the comprehension tasks (reading and lexical decision). As with the production studies, the monolinguals and Spanish–English bilinguals were undergraduates at the University of California, San Diego, and received course credit for their participation; the Dutch–English bilinguals were undergraduates at Ghent University in Belgium and either received course credit or were paid for their participation. Table 1 displays a summary of the participant characteristics.

Materials. The materials for the comprehension studies were the same as those from the production studies but with a few important changes. First and most obviously, the pictures were replaced by their respective word names in all conditions. Second, for the conditions in which context was provided (sentence reading), additional text was added to the sentence following the target words. This was done because in eye-tracking studies of reading it is common to avoid having target words occur in sentence initial or final positions. Fixations on words in these positions can be influenced by nonlanguage aspects of the experiment such as orienting to the appearance of the stimuli and the decision to terminate reading on a trial. Additionally, including posttarget text avoids sentence wrap-up effects from influencing the eye-tracking measures on the target words (Rayner, Kambe, & Duffy, 2000). Finally, for each target word we created an appropriate and pronounceable nonword for use in the lexical decision task. Nonwords were legal and pronounceable strings created by altering two letters of each of the target words (e.g., *apple* became *airle*, and *arrow* became *anrew*).

Procedure. As in Experiment 1, each participant completed all three lists in one of each of the three conditions (no-context lexical decision, low-constraint context, high-constraint context). Assignment of material list to experimental condition was counterbalanced between participants such that each of the three ma-

terials lists was presented in each of the three conditions an equal number of times. Therefore, each participant read 60 targets embedded in sentences (15 sentences in each of the frequency by constraint conditions) and performed lexical decisions for 30 words and 30 nonwords in isolation. Across all the participants, each word appeared equally often in each of the context conditions (high constraint, low constraint, and without context), and no participant saw a target in more than one condition. As with the production study, the no-context condition was presented separately from the context condition, and the two context conditions were presented together with low- and high-constraint sentences intermixed in a different random order for each participant. The order of testing with versus without context was counterbalanced across participants.

Eye movements were recorded via an SR Research Eyelink 1000 eye tracker, which recorded the position of the reader's eye every millisecond and has a spatial resolution of less than 0.04° . Participants were seated 55 cm (21.6 in.) away from a 19-inch ViewSonic VX922 LCD monitor. Head movements were minimized with chin and head rests. Although viewing was binocular, eye movements were recorded from the right eye. The same equipment was used for data collection in both locations (San Diego and Ghent).

At the start of the sentence-reading experiment, participants completed a calibration procedure by looking at a sequence of three fixation points randomly presented horizontally across the middle of the computer screen. A validation procedure then repeated this process to check the error of the calibration. If this error was greater than 0.4° of visual angle, the entire procedure was repeated. At the start of each with-context trial, a black square (50 pixels wide and 50 pixels tall) appeared on the left side of the computer screen. This square coincided with the left side of the first letter in the upcoming sentence. The sentence replaced this square on the screen once a stable fixation was detected within its area. Sentences were presented as black letters on a white background in 14-point Courier New font with 3.2 letters equaling 1° of visual angle. Participants were instructed to read silently for comprehension and to press a button on a keypad when they finished reading. To encourage participants to pay attention to the sentences, the word *REMEMBER* appeared on the screen following this button press on six of the 60 trials. At this prompt, participants were asked to recall the sentence aloud, and the experimenter noted the accuracy. Three high-constraint and three low-constraint sentences were preselected for this purpose. The first eight trials were practice trials to get the participants accustomed to the procedure.

The lexical decision task was carried out in the same lab space as was the sentence-reading study and utilized the same chin and head rests on the same computer as when the eye-tracking data were collected during sentence reading. Prior to each trial, participants fixated a marker in the center of the computer monitor. Following a 1-s delay, a letter string was presented in the center of the monitor, and participants were told to press the left button on the keypad if the string was a word and to press the right button if it was a nonword. Participants were instructed to respond as quickly as possible while minimizing errors. The order of the items was randomized separately for each participant.

Results

As in Experiment 1, we first present the results from reading out of context (lexical decision), followed by the reading in context (eye-tracking) results. As shown in Table 3, we examined the following four standard eye-movement measures (Rayner, 1998, 2009): (a) *first fixation duration* (the duration of the first fixation on the target word),³ (b) *gaze duration* (the sum of all fixations on the target word before moving to another word), (c) *total reading time* (the sum of all fixations on the target word including regressions), and (d) *skipping probability* (the probability that the target word was skipped on first pass reading). In addition, we computed an alternative measure in which gaze duration was adjusted to account for skipping; in these analyses, gaze duration for a skipped target word was assigned a value of zero (Just & Carpenter, 1980; Murray, 2000; Rayner, Slattery, Drieghe, & Livesedg, in press). Briefly summarized, we obtained the same results in first fixation, gaze duration, and total reading times; these differed slightly from skip-adjusted gaze durations and considerably from skipping rates. Specifically, only skipping probabilities exhibited a robust Frequency \times Constraint interaction (albeit of a different nature from that seen in Experiment 1), and the interaction was either completely ($F_s < 1$) or largely absent from the other measures. For ease of exposition, the detailed results for first fixation durations and total reading times are presented in the Appendix.

Lexical decision out of context. The results of the lexical decision task are shown in Figure 2, including responses to nonwords (the means are shown for completeness, though nonwords were not included in any statistical comparisons; see also summary of results in Table 4). Participants produced correct responses to high-frequency words faster than to low-frequency words, $F_1(1, 98) = 90.09$, $MSE = 2,140.56$, $\eta_p^2 = .48$, $p < .001$; $F_2(1, 88) = 49.16$, $MSE = 6,314.58$, $\eta_p^2 = .36$, $p < .001$, and monolinguals responded more quickly than did bilinguals $F_1(2, 98) = 3.12$, $MSE = 19,925.92$, $\eta_p^2 = .06$, $p < .05$; $F_2(2, 176) = 20.34$, $MSE = 5,068.74$, $\eta_p^2 = .19$, $p < .001$. In contrast to the production data, in which slowing related to bilingualism was robust for both bilingual groups, pairwise comparisons indicated that the monolinguals were only marginally faster than Spanish–English bilinguals (both $ps < .06$) and significantly faster than the Dutch–English bilinguals (both $ps < .05$). In addition, the difference between Spanish–English and Dutch–English bilinguals was significant in the items analysis ($p < .05$) but did not approach significance in the participant analysis ($F < 1$). Of interest, there was a Frequency \times Group interaction, $F_1(2, 98) = 11.97$, $MSE = 2,140.56$, $\eta_p^2 = .20$, $p < .001$; $F_2(2, 176) = 10.96$, $MSE = 5,068.74$, $\eta_p^2 = .11$, $p = .001$. However, only Dutch–English bilinguals exhibited a larger frequency effect than did the other two groups (both $ps < .005$), whereas Spanish–English bilinguals and monolinguals exhibited the same size frequency effect (both $F_s < 1$). This contrasts with the production data, in which both Spanish–English and Dutch–English bilinguals exhibited larger frequency effects than did monolinguals.

³ We also analyzed single-fixation durations (those cases in which the target was fixated a single time during first pass reading). The basic pattern of results was the same for this measure as for first fixation and gaze duration.

Table 3
Participant Means (and Standard Errors) for Eye-Tracking Measures in Experiment 2

Measure and frequency	Monolinguals		Spanish–English bilinguals		Dutch–English bilinguals	
	Low constraint	High constraint	Low constraint	High constraint	Low constraint	High constraint
First fixation (ms)						
Low frequency	216 (5.1)	216 (5.3)	228 (5.1)	217 (5.2)	233 (5.9)	239 (6.1)
High frequency	210 (4.5)	205 (4.7)	212 (4.5)	205 (4.7)	221 (5.2)	220 (5.4)
Gaze duration (ms)						
Low frequency	235 (6.9)	231 (6.8)	261 (6.9)	240 (6.8)	280 (8.0)	273 (7.8)
High frequency	222 (6.7)	213 (5.8)	239 (6.7)	219 (5.8)	260 (7.8)	249 (6.7)
Total time (ms)						
Low frequency	288 (12.4)	284 (12.8)	328 (12.4)	299 (12.8)	391 (14.3)	372 (14.8)
High frequency	268 (11.1)	256 (10.2)	292 (11.1)	274 (10.2)	346 (12.8)	326 (11.8)
Skipping rate (%)						
Low frequency	13.1 (1.9)	22.3 (2.2)	9.9 (1.9)	13.2 (2.2)	4.7 (2.1)	9.7 (2.5)
High frequency	18.3 (1.9)	21.1 (2.1)	14.7 (1.9)	11.7 (2.1)	8.0 (2.2)	11.5 (2.4)
Alternative gaze duration (ms)						
Low frequency	204 (9.0)	181 (8.8)	238 (9.0)	209 (8.8)	269 (10.4)	249 (10.2)
High frequency	182 (8.3)	168 (7.5)	206 (8.3)	193 (7.5)	240 (9.6)	222 (8.7)

Error analyses exhibited a pattern of results similar to that of the RT data, with robust main effects of frequency and group (all $ps < .001$) and robust Frequency \times Group interactions (both $ps < .001$; the main effect of participant type and the interaction with fre-

quency was not significant without the Dutch–English bilinguals; $F_s < 1$).

Thus, the out-of-context (lexical decision) data exhibit striking differences with respect to the production data reported in Experiment 1. In particular, slowing associated with bilingualism was not as robust (only the Dutch–English bilinguals responding to written words in their second language were markedly slower than the monolinguals). Similarly, whereas the Frequency \times Group interaction was present in production at all levels of English proficiency (comparing monolinguals with Spanish–English bilinguals, and monolinguals with Dutch–English bilinguals), in lexical decision the frequency effect was larger only in the participant group that was least proficient in English.

A similarity that did emerge was that when the targets were in a nondominant language (English for the Dutch–English bilinguals), the frequency effect was significantly larger both in picture naming and in lexical decisions than when the targets were in a dominant language (English for the Spanish–English bilinguals), confirming previous results for both production and comprehension (respectively, Gollan et al., 2008, and Duyck et al., 2008).

Reading in context (eye tracking). Gaze durations are our primary measure of target word processing because they are the sum of all fixations on the target words prior to leaving it in either direction (contingent on it not being initially skipped). Gaze durations are often interpreted as a measure of lexical access (Rayner, 1998, 2009). This measure is also the most appropriate for comparison with other tasks, such as production and lexical decision (which have only a single dependent measure of processing time), and has been shown to correlate best (of the various fixation time measures) with lexical decision and naming times (Schilling, Rayner, & Chumbley, 1998). To facilitate comparison with Experiment 1, the means for the gaze durations are shown in Figure 3 (these same means appear in Table 3). Because of the large F3 difference in baseline between eye-tracking measures and explicit responses, lexical decision times are shown separately (see Figure 3). Gaze durations were shorter on high- than low-frequency words, $F_1(1, 96) = 77.39, MSE = 498.64, \eta_p^2 = .45, p < .001$; $F_2(1, 88) = 25.03, MSE = 2,082.80, \eta_p^2 = .22, p < .001$, and

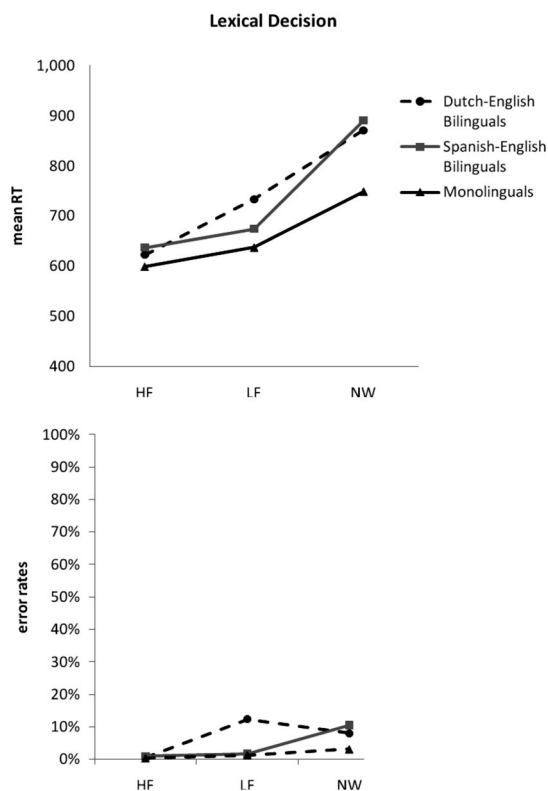


Figure 2. Lexical decision times in milliseconds (top panel) and error rates (bottom panel) for high- and low-frequency targets in the out-of-context condition for each participant type in Experiment 2. Means are taken from participants analyses, and error bars are standard errors for each condition. HF = high frequency; LF = low frequency; NW = nonword.

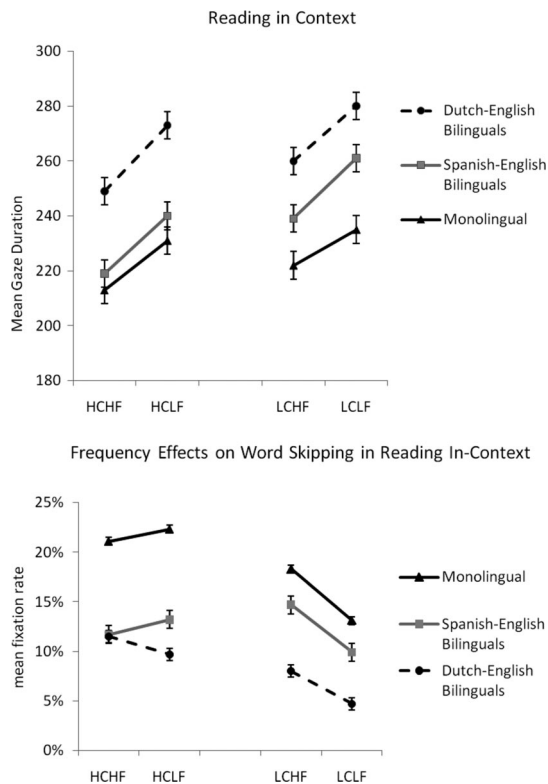


Figure 3. Gaze duration times in milliseconds (top panel) and word-skipping rates (bottom panel) in each of the four in-context conditions for each participant type in Experiment 2. Means are taken from participants analyses, and error bars are standard errors for each condition. HC = high semantic constraint; LC = low semantic constraint; HF = high frequency; LF = low frequency.

shorter on high- than on low-constraint target words, $F_1(1, 96) = 24.32$, $MSE = 587.83$, $\eta_p^2 = .20$, $p < .001$; $F_2(1, 88) = 20.20$, $MSE = 1,329.62$, $\eta_p^2 = .19$, $p < .001$. There was also a main effect of group indicating that gaze durations were shorter for monolinguals than for bilinguals, $F_1(2, 96) = 10.80$, $MSE = 4,679.98$, $\eta_p^2 = .18$, $p < .001$; $F_2(2, 176) = 88.42$, $MSE = 943.96$, $\eta_p^2 = .50$, $p < .001$. Planned comparisons indicate that gaze durations for monolinguals were significantly shorter than for Spanish-English bilinguals (both $ps < .05$), and both of these groups of readers had shorter gaze durations than did the Dutch-English bilinguals (all $ps < .01$); thus, slowing associated with bilingualism was more robust (applying to both bilingual groups) in context with the gaze duration measure than it was out of context with the lexical decision task (the Dutch-English group was significantly slower, but the Spanish-English group was only marginally slower, than monolinguals).

In stark contrast to the production data (in Experiment 1), there was no indication that frequency interacted with constraint in gaze duration times (both $F_s < 1$), the bilingual disadvantage was not larger for low-frequency words than for high-frequency words (both $F_s < 1$), and there was no three-way Frequency \times Constraint \times Group interaction ($F_s < 1$). There was a Constraint \times Group interaction, $F_1(2, 96) = 3.51$, $MSE = 587.83$, $\eta_p^2 = .07$, $p < .05$; $F_2(2, 176) = 4.53$, $MSE = 1,046.40$, $\eta_p^2 = .05$, $p < .05$;

however, it was entirely different in nature from that observed in the production data and not the interaction that would be predicted on the basis of the three-tiered manipulation of English proficiency. Specifically, Spanish-English bilinguals benefited more than did monolinguals ($ps < .005$) and marginally more than did Dutch-English bilinguals ($ps < .08$) from semantically constraining context (but see the analysis of skipping rates later), but constraint effects between the two groups that differed most in English proficiency (monolinguals and Dutch-English bilinguals) did not differ (both $F_s < 1$). As can be seen later, this difference seemed to be attributable to a tradeoff between gaze durations and skipping rates and between group differences in these two measures.

Thus, contrary to the results of the production study RTs, there was no indication of a Frequency \times Constraint interaction in any of the standard fixation duration measures (see also Appendix), and Frequency \times Bilingualism interactions were also rather different across the two experiments. The contrast between results observed in production (Experiment 1) versus comprehension (Experiment 2) was not unexpected, given previously published results (see e.g., Griffin & Bock, 1998, versus Rayner et al., 2004). However, the observation of this contrast in the same study, using the same materials and the same participant populations, provides confidence that the differences observed across studies reflect an actual difference between modalities. Also of note with respect to the predictions of the interference account of bilingual disadvantages, in production there was some evidence for an attenuation of the bilingual disadvantage in context, whereas in reading if anything the bilingual disadvantage was more robust in context (though in Experiment 2 this difference could reflect task differences between lexical decision and gaze duration rather than a context effect). In addition, an unexpected finding from the perspective of the frequency-lag hypothesis was the presence of a disadvantage without an interaction with frequency in gaze durations.

Another important dependent measure in eye movement studies of reading is skipping rate. Skipping rates of target words are often interpreted as a measure of processing difficulty, with skipping rates increasing for targets that are easier to process (i.e., shorter words, higher frequency words, more predictable words). In the current study (see Panel B of Figure 3 and Table 3), the skipping rate was 14.22% for high-frequency target words but only 12.16% for low-frequency target words, a difference that was significant in the participants' analysis but only marginal over items, $F_1(1, 96) = 4.81$, $MSE = 85.96$, $\eta_p^2 = .05$, $p < .05$; $F_2(1, 88) = 2.80$, $MSE = 226.10$, $\eta_p^2 = .03$, $p = .098$. More importantly, contextual constraint (or predictability) has a strong influence on skipping rates (Rayner, 1998, 2009). Not surprisingly then, high-constraint targets were skipped 14.91% of the time, whereas low-constraint targets were only skipped 11.46% of the time, $F_1(1, 96) = 10.01$, $MSE = 115.67$, $\eta_p^2 = .09$, $p < .001$; $F_2(1, 88) = 7.56$, $MSE = 190.57$, $\eta_p^2 = .08$, $p < .01$. There was also a main effect of group on skipping rate, $F_1(2, 96) = 10.75$, $MSE = 316.47$, $\eta_p^2 = .18$, $p < .001$; $F_2(2, 176) = 47.14$, $MSE = 101.16$, $\eta_p^2 = .35$, $p < .001$, because monolinguals were more likely to skip targets (18.70%) than were Spanish-English bilinguals (12.39%), who were more likely to skip targets than were Dutch-English bilinguals (8.47%; all $ps < .05$).

Of interest, skipping rates paralleled the production data in that frequency effects were larger when semantic constraint was low than when it was high, $F_1(1, 96) = 7.75$, $MSE = 69.45$, $\eta_p^2 = .08$, $p < .01$; $F_2(1, 88) = 4.07$, $MSE = 190.57$, $\eta_p^2 = .04$, $p < .05$. In addition, there was some indication of a Group \times Constraint interaction, $F_1(2, 96) = 2.76$, $MSE = 115.67$, $\eta_p^2 = .05$, $p = .068$; $F_2(2, 176) = 3.30$, $MSE = 123.85$, $\eta_p^2 = .04$, $p < .05$, but this was in the opposite direction of the effect seen in fixation duration measures, with Spanish–English bilinguals exhibiting a smaller constraint effect than did monolinguals (both $ps < .05$) and trending in the same direction when compared with Dutch–English bilinguals (both $ps < .12$). However, there was no indication of a Group \times Frequency interaction (both $F_s < 1$). The three-way interaction did not approach significance ($F_s < 1$).

The different pattern of results for fixation duration and skipping rate measures complicates the interpretation of the eye movement data. However, two other eye movement studies that also manipulated contextual constraint and frequency found an interaction between these variables in skipping rates despite not finding such an interaction in the overall fixation durations (Hand, Mielliet, O’Donnell, & Sereno, 2010; Rayner et al., 2004). Both of these earlier studies ignored the interaction in skipping when interpreting fixation duration effects because there is evidence to suggest that the decision of when to move the eyes and the decision of where to move them are independent of each other (Rayner, 1998, 2009). However, due to the goals of the current studies, resolving such complications in the data patterns of these measures is critical. One technique that has been used to reconcile such differences (see Just & Carpenter, 1980; Murray, 2000; Rayner et al., in press) is to average in a zero to gaze durations for every skip. The logic inherent to this procedure is that skipped words are processed in the parafovea to a large enough extent that they don’t require a direct fixation for lexical access and that this will be more likely to happen when overall lexical processing is easier. After averaging zeros in, gaze duration times remained shorter on high- than low-frequency target words, $F_1(1, 96) = 49.28$, $MSE = 1,060.60$, $\eta_p^2 = .34$, $p < .001$; $F_2(1, 88) = 16.45$, $MSE = 4,329.66$, $\eta_p^2 = .16$, $p < .001$, and shorter on high- than on low-constraint targets, $F_1(1, 96) = 27.14$, $MSE = 1,338.28$, $\eta_p^2 = .22$, $p < .001$; $F_2(1, 88) = 23.45$, $MSE = 2,178.12$, $\eta_p^2 = .21$, $p < .001$, and were affected by group, $F_1(2, 96) = 15.93$, $MSE = 7,266.21$, $\eta_p^2 = .25$, $p < .001$; $F_2(2, 176) = 125.84$, $MSE = 1,365.18$, $\eta_p^2 = .59$, $p < .001$; gaze durations with zeros averaged in were shorter for monolinguals than for the bilinguals (all $ps < .005$) and were shorter for Spanish–English bilinguals than for Dutch–English bilinguals (both $ps < .01$).

The procedure of averaging zeros into gaze durations also introduced a hint of a Constraint \times Frequency interaction, but this was only marginally significant in the participants analysis and did not approach significance by items, $F_1(1, 96) = 3.43$, $MSE = 587.14$, $\eta_p^2 = .03$, $p = .067$; $F_2(1, 88) = 1.33$, $MSE = 2,178.12$, $\eta_p^2 = .02$, $p = .253$. Additionally, the Constraint \times Group interaction that had been significant in both the duration and skipping measures (though in opposite directions) did not approach significance in this alternative measure (both $F_s < 1$). None of the other interactions approached significance (all $F_s < 1$).

Thus, after taking measures to consolidate the differences between skipping rates and fixation times, a clearer picture emerges.

Whereas we observed significant Frequency \times Constraint interactions in language production, these effects are largely absent in reading. Similarly, whereas we found significant Frequency \times Group interactions in language production, these effects were completely absent for reading in context.

Comparing Comprehension and Production

The contrast between the results reported in Experiment 1 and those reported in Experiment 2 imply a distinct effect of frequency on lexical access in production versus comprehension. In comprehension, frequency seems to operate independently of context, primarily serving to initiate lexical access via form-driven frequency-sensitive search process. In production, frequency effects seem to be modulated by context, influencing access but only as a part of a meaning-initiated search that can be affected by frequency but is not frequency-driven. Direct comparisons of comprehension and production are difficult to achieve because of the different baselines inherent to the different tasks (fixation times in reading are relatively short, whereas picture-naming and lexical decision times are much longer). To facilitate more direct comparisons across modalities, we conducted two sets of analyses, including (a) a baseline adjustment and (b) vincentile plots.

To adjust for baseline differences in our measures, we calculated proportionally adjusted frequency effects for each participant in each of the three conditions in Experiments 1 and 2 as follows:

$$(\text{low frequency} - \text{high frequency}) / [(\text{low frequency} + \text{high frequency}) / 2]$$

For this purpose, we used picture-naming times from Experiment 1 and gaze duration (without zeros averaged in) and lexical decision times in Experiment 2. These proportional frequency values were then submitted to a $2 \times 3 \times 3$ ANOVA with modality (production, comprehension) and participant group (monolinguals, Spanish–English bilinguals, Dutch–English bilinguals) as nonrepeated factors and contextual constraint (out of context, low constraint, and high constraint) as repeated factors to determine the relative size of the frequency effect in comprehension and production and whether this is modulated by context. The results of this analysis are shown in Figure 4 with separate panels for each participant group. F4

Briefly summarized, the relative magnitude of the frequency effect in production (Experiment 1) versus comprehension (Experiment 2) varied with contextual constraint. Monolinguals and Spanish–English bilinguals exhibited a similar pattern of results, which was slightly different from that seen in Dutch–English bilinguals. However, in all three participant groups, semantic constraint reduced the size of the frequency effect in picture naming. Overall, proportionally adjusted frequency effects across experiments revealed significant main effects of language such that monolinguals exhibited smaller frequency effects than did bilinguals, $F(2, 238) = 15.76$, $MSE = 0.02$, $\eta_p^2 = .12$, $p < .01$, a main effect of modality such that production resulted in larger frequency effects than did comprehension, $F_1(1, 238) = 6.22$, $MSE = 0.02$, $\eta_p^2 = .03$, $p = .01$, and a main effect of contextual constraint such that frequency effects were smallest in high-constraint semantic context, $F_1(1, 238) = 4.41$, $MSE = 0.02$, $\eta_p^2 = .02$, $p = .04$.

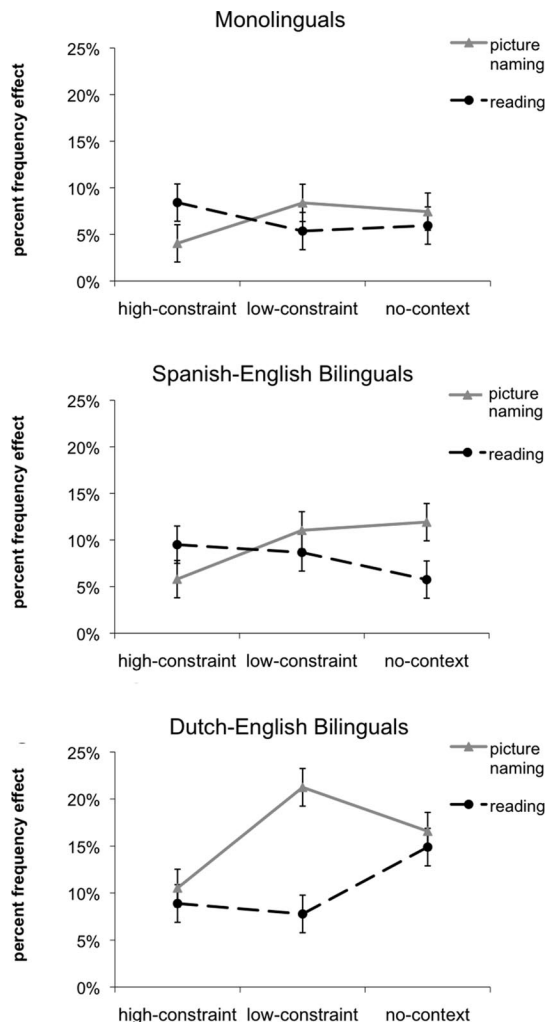


Figure 4. The size of the frequency effect in each condition in picture naming (Experiment 1) and reading (Experiment 2) as a proportion of baseline— $(LF - HF)/(LF + HF)/2$ —for monolinguals (top panel), Spanish–English bilinguals (middle panel), and Dutch–English bilinguals (bottom panel). The size of the frequency effect was calculated using picture-naming times for all conditions in Experiment 1, gaze duration times for in-context reading, and lexical decision times for out-of-context reading in Experiment 2.

These main effects were qualified by a significant Contextual Constraint \times Modality (experiment) interaction such that frequency effects were larger in production than in comprehension in no- or low-constraint contexts but smaller or similarly sized in production than in comprehension when in high-constraint contexts. None of the other interactions approached significance (all p s $>$.16), except for the Language \times Modality interaction ($p = .07$), which was driven exclusively by Dutch–English bilinguals. In comparing monolinguals with Dutch–English bilinguals (excluding the Spanish–English group), there were main effects of language ($p < .01$), modality ($p = .03$), and contextual constraint ($p = .05$) and a significant Language \times Modality interaction ($p = .03$). However, in comparing monolinguals with Spanish–English bilinguals (excluding Dutch–English bilinguals), the main effect of

language remained significant ($p = .03$) and the Constraint \times Modality interaction was quite robust ($p < .01$), but none of the other main effects or interactions approached significance.

Planned comparisons focusing on the context conditions (low vs. high constraint) in each participant group revealed a significant crossover Modality \times Contextual Constraint interaction in monolinguals ($p = .04$) and a marginally significant crossover interaction in the same direction in Spanish–English bilinguals ($p = .08$), such that frequency effects were larger in production than in comprehension in low-constraint contexts but smaller in production than in comprehension in high-constraint contexts. Dutch–English bilinguals exhibited a similar interaction ($p = .01$), but the size of the frequency effects in production did not dip below that in comprehension in high-constraint contexts (no crossover). In comparing the different modalities for the out-of-context conditions, Spanish–English bilinguals exhibited significantly greater frequency effects in production than in comprehension ($p = .01$), but for monolinguals the frequency effect did not vary in magnitude across modalities (see Panel A of Figure 4), whereas Dutch–English bilinguals exhibited a large frequency effect in comprehension—as large as that in production (see Panel C of Figure 4).

To further facilitate the comparison between comprehension and production, illustrating the effects of frequency and constraint across the full range of processing times in each task, we constructed vincentile plots (Vincent, 1912). Vincentile plots (Ratcliff, 1979; Staub, White, Drieghe, Hollway, & Rayner, 2010) have recently been recommended (see e.g., Balota, Yap, Cortese, & Watson, 2008; Yap & Balota, 2007) for deriving better informed theoretical implications based on between-condition latency differences that can arise from differences in mode RT, in the tails of the RT distributions, or in both. Vincentile plots are constructed by first sorting the responses for each participant by condition in order (fastest to slowest) and then assigning times for each participant into bins ranging from fastest to slowest. For current purposes we assigned each participants’ data points into five bins, with a maximum of three data points per participant per bin per condition (some bins had fewer than three data points for each participant because of missing data/outlier removal). In Experiment 1 the data from all 145 participants, and in Experiment 2 the data from all 99 participants, were averaged into each bin. Subsequently, an average time for each bin in each condition was calculated, and when graphed these means illustrate how the effects of interest vary across the distribution. The means for each bin as a function of frequency and constraint are shown in Figure 5, with the production data in Panel A and the reading data (unadjusted gaze duration times) in Panel B (for simplicity of presentation, out-of-context data are not shown).

The most immediately obvious difference between production and comprehension is in the effect of context. Though constraining context speeds the processing in both comprehension and production, the effect appears to be much larger in production: Here, the low- and high-constraint naming times do not overlap at all, whereas in reading the separation is less clear. More specifically, for picture-naming times, semantic constraint seems to be more powerful than frequency effects (e.g., high-constraint low-frequency points are faster than the low-constraint high-frequency points), whereas in reading frequency was a stronger determinant of gaze duration times (low-frequency means tend to be slower than high-frequency means regardless of constraint). The vincen-

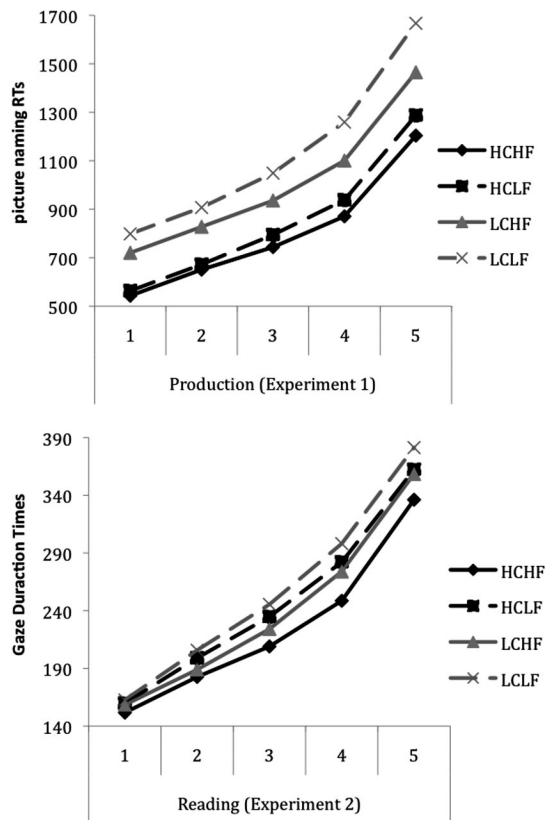


Figure 5. The effects of frequency and constraint on picture-naming times (top panel) and gaze duration times (not adjusted for skipping; bottom panel) in five vincentile bins, reflecting participants' fastest and slowest (and three levels in between) response times in each condition (see text for additional explanation). Note that the scales on the y-axes were set to maximize differences between the four points in each bin. HC = high semantic constraint; LC = low semantic constraint; HF = high frequency; LF = low frequency.

tile bins also illustrate the consistency of the reported effects across the distribution of times. In Panel A, the separation between high- and low-frequency targets is larger in low constraint than it is in high constraint at all five time bins (i.e., the frequency effect is smaller in semantically constraining context for production). In Panel B, the separation between high- and low-frequency means for gaze duration is about equal in low-constraint (triangles and Xs) and high-constraint conditions (diamonds and squares; i.e., the frequency effect is not modulated by context for reading). The figures also show that frequency effects in gaze duration times are small in the first time bin, as well as in production with low-constraint context in the first two bins.

A $2 \times 2 \times 5$ ANOVA with frequency, constraint, and vincentile bins as factors confirmed these conclusions. In production there were significant two-way interactions (all $ps < .01$) between frequency and vincentile points and between constraint and vincentile points (such that effects were larger in slower RT bins), as well as the already reported Frequency \times Constraint interaction, but there was no indication that the latter interaction varied with vincentile bin, $F_1(1, 139) = 1.33$, $MSE = 39,394$, $\eta_p^2 = .01$, $p = .25$. In comprehension, the two-way Frequency \times Vincentile

Points interactions and Constraint \times Vincentile Points interactions were significant (both $ps < .01$), but the Frequency \times Constraint interaction, as well as the three-way interaction, was not (both $F_s < 1$). Altogether, the proportional analyses and the vincentile plots confirm our conclusions with respect to the effects of frequency and constraint in comprehension versus production, while illustrating their consistency across the distribution of RTs.

General Discussion

The goal of this study was to identify the mechanisms of lexical access that are modality-specific and to distinguish them from those that are language-general, aiming at a better understanding of the differences and commonalities that underlie the cognitive processes associated with language production versus comprehension. A review of these different literatures indicated a contrast between studies of picture naming and studies of word recognition with respect to the joint effects of word frequency and semantically constraining context on lexical access. The data reported here provide strong evidence confirming this striking contrast between modalities.

Using the same materials in picture naming, lexical decision, and eye tracking, with three groups of participants at different levels of English proficiency, we observed robust main effects of frequency, semantic constraint, and participant group on both picture-naming and eye-tracking measures. However, whereas picture naming produced robust Frequency \times Constraint interactions, these were largely absent in reading measures. Similarly, only picture naming (and not reading) revealed an interaction in the expected direction between English proficiency level (a proxy of frequency) and constraint (such that slowing associated with a low proficiency level was attenuated in production but not in comprehension). Slowing associated with bilingualism appeared to be more robust in production than in comprehension (at least for lexical access out of context). Direct comparisons of picture-naming and reading measures (after adjusting for the baseline level in the respective response measures; see Figure 4) revealed that frequency effects were larger in production than in comprehension without semantic constraint but were smaller in production than in comprehension in high-constraint semantic contexts (for all participants with English as their dominant language, including monolinguals). In addition, constraint effects were larger in production than in comprehension (see Figure 5).

Table 4 summarizes the pattern of statistical significance for all major contrasts in Experiments 1 and 2. A summary of effect sizes is also provided for effects of crucial importance for this study (e.g., note that the constraint effects are highly robust statistically in both experiments, although the size of the effect appears to be much larger in picture naming than in reading).

The Role of Frequency in Lexical Access

The results reported here imply differences in the cognitive mechanisms underlying language comprehension versus production. As outlined earlier, frequency would logically be assigned a different role when the process of lexical access begins with meaning and ends with access to lexical forms (in production), when compared with a process that begins by accessing lexical forms and ends with processing of meaning (in comprehension).

Table 4
Summary of Statistical Significance and Effect Size for Contrasts of Interest

Effect	Experiment 1				Experiment 2			
	Significance		η_p^2		Significance		η_p^2	
	F_1	F_2	F_1	F_2	F_1	F_2	F_1	F_2
Out-of-context 2 × 3 ANOVA								
Main effects								
Frequency: high < low	**	**	.35	.12	**	**	.48	.36
Participant group: monolingual < Spanish-English < Dutch-English	**	**	.32	.56	*	**	.06	.19
Dutch-English vs. monolingual	**	**	.40	.56	*	*	.08	.26
Spanish-English vs. monolingual	**	**	.20	.53	†	†	.04	.33
Spanish-English vs. Dutch-English	**	**	.14	.32	$F < 1$	*	.01	.05
Interaction								
Frequency × Participant Group: bilingual disadvantage bigger for low- than high-frequency targets	**	*	.09	.06	**	**	.20	.11
Dutch-English vs. monolingual	**	*			**	**		
Spanish-English vs. monolingual	*	*			$F < 1$	$F < 1$		
Spanish-English vs. Dutch-English	†	†			**	**		
In-context 2 × 2 × 3 ANOVA								
Main effects								
Frequency: high < low	**	**	.49	.13	**	**	.45	.22
Constraint: high < low	**	**	.77	.63	**	**	.20	.19
Participant group: monolingual < Spanish-English < Dutch-English	**	**	.23	.55	**	**	.18	.50
Dutch-English vs. monolingual	**	**	.29	.56	**	**	.25	.67
Spanish-English vs. monolingual	**	**	.14	.65	*	*	.06	.25
Spanish-English vs. Dutch-English	**	**	.08	.26	**	**	.10	.39
2-way interactions								
Frequency × Constraint: frequency effect is bigger in low than high constraint	**	*	.23	.04	$F < 1$	$F < 1$	—	—
Monolingual	*	†			$F < 1$	$F < 1$		
Spanish-English	**	ns^a			$F < 1$	$F < 1$		
Dutch-English	**	†			$F < 1$	$F < 1$		
Frequency × Participant Group: bilingual disadvantage bigger for low- than high-frequency targets	**	**	.19	.08	$F < 1$	$F < 1$	—	—
Low constraint								
Dutch-English vs. monolingual	**	**			$F < 1$	$F < 1$		
Spanish-English vs. monolingual	*	$F < 1$			ns^a	$F < 1$		
Spanish-English vs. Dutch-English	**	*			$F < 1$	$F < 1$		
High constraint								
Dutch-English vs. monolingual	*	ns^a			$F < 1$	$F < 1$		
Spanish-English vs. monolingual	ns^a	ns^a			$F < 1$	$F < 1$		
Spanish-English vs. Dutch-English	ns^a	ns^a			$F < 1$	$F < 1$		
Constraint × Participant Group: bilingual disadvantage bigger in low than high constraint	*	*	.04	.05	*	*	.07	.05
Low frequency								
Dutch-English vs. monolingual	**	*			$F < 1$	$F < 1$		
Spanish-English vs. monolingual	ns^a	$F < 1$			*	*		
Spanish-English vs. Dutch-English	*	†			†	†		
High frequency								
Dutch-English vs. monolingual	$F < 1$	$F < 1$			$F < 1$	$F < 1$		
Spanish-English vs. monolingual	$F < 1$	$F < 1$			ns^a	†		
Spanish-English vs. Dutch-English	$F < 1$	$F < 1$			ns^a	ns^a		
3-way interactions: bilinguals benefit more from constraint than do monolinguals only for low-frequency targets	**	ns^a	.08	.02	$F < 1$	$F < 1$	—	—
Dutch-English vs. monolingual	**	ns^a						
Spanish-English vs. monolingual	ns^a	$F < 1$						
Spanish-English vs. Dutch-English	*	ns^a						

Note. Dashes indicate that data are not reported; effect sizes were added for only effects of interest. ANOVA = analysis of variance.
^a But $F > 1$.

† $p \leq .10$. * $p \leq .05$. ** $p \leq .01$.

The data we reported suggest an unequivocal role for frequency in lexical access during reading that is relatively (if not completely) independent from later processing effects such as the interpretation of meaning. In contrast, in production the role of frequency is attenuated (or even eliminated in some cases; Griffin & Bock, 1998) by semantic context. Thus, in comprehension, lexical access itself is a frequency-driven process, whereas lexical access during production is primarily a semantically driven process with the expression of frequency effects coming and going on the basis of the relative absence or presence of semantic constraint.

Two aspects of the data we reported would seem to challenge this claim. The first is the Frequency \times Constraint interaction in word-skipping rates for reading in context. In particular, readers were less likely to skip a target word if it was a low-frequency word that occurred in a low-constraint context, although frequency did not influence skipping rates otherwise (see Panel B of Figure 2). Note that the analogous interaction in picture naming was of a different form; specifically, in naming times, frequency effects were present in all conditions (including high-constraint contexts), but they were larger in low-constraint semantic contexts. Although the skipping data are interesting in their own right, the vast majority of words are not skipped, and it isn't as clear what to make of skipping data as it is with fixation time measures (because assumptions must be made with respect to how skipped words are processed). To incorporate the skipping data into our analyses we relied on a measure that combines skipping rates and gaze duration times (see Table 3). This approach makes the (likely overly conservative) assumption that skipped words are processed in no time. Despite this conservative approach, the analyses of skipping adjusted gaze duration times failed to produce robust Frequency \times Constraint interactions (only a marginal effect in the participants analysis and no effect by items). Thus, we might temper our claims here: Rather than say that there was absolutely no evidence for a Frequency \times Constraint interaction in reading measures, we found that the evidence for an interaction was both qualitatively different in reading than in production and considerably less robust in reading than in production.

A second question concerns the magnitude of the frequency effect in the two modalities. For all participant groups, frequency effects were as large (or in some cases even larger) in production as in comprehension, both without context and in low-constraint semantic context. For the least English-proficient group (Dutch-English bilinguals), this was true even in high-constraint semantic context. If frequency plays a less central role in lexical access for production, then why were frequency effects sometimes so large in production? In particular, how does one reconcile the claim that frequency plays a central role in lexical access for comprehension but not production with the fact that frequency effects weren't consistently smaller in production than in comprehension? Importantly, our comparisons of comprehension and production relied on baseline adjusted measures of the frequency effect (see Figure 4), and thus it is not the case that frequency effects were large in production simply because picture-naming times are substantially longer than fixation times in reading (see Figures 1 and 2). In addition, monolinguals also exhibited the modulation of frequency effects by modality and constraint (such that frequency effects tended to be larger in production than in reading without constraining context but smaller than in reading with constraining context);

hence, the modulation applied even at the highest proficiency levels.

Frequency Lag for Production Versus Comprehension

One interesting, though admittedly speculative, possibility is that production could lag behind comprehension in frequency of use. If reading has a higher cumulative frequency over a lifetime than does speaking (at least for some people, including the participants in the current study), this could explain why frequency effects can still be quite large in production even though access is not primarily frequency-driven in production. Although people learn to speak as children long before they learn to read, reading rapidly becomes a high-frequency activity early in life (in school at age 6 and after school in homework, e-mail, and reading for pleasure, etc.). Moreover, when they read it is often one-sided (they don't write at the same time), whereas when speaking they more often divide their time with auditory comprehension (except for folks who talk to themselves). In addition, many people often listen to a single speaker. Current estimates of how many words are spoken in a day by college students is about 16,000 (Mehl, Vazire, Ramirez-Esparza, Slatcher, & Pennebaker, 2007). Assuming college students read anywhere from 2 to 8 hr per day, at a rate of about 300 words per minute (Pelli et al., 2007; Taylor, 1965), this would translate into many more words read than spoken per day⁴ ($300 \times 60 \times 2 = 36,000$ and $300 \times 60 \times 8 = 144,000$). Fn4

If frequency effects become smaller and smaller with increased use, this should lead to smaller frequency effects in the more often used modality. Similarly, a lifetime of using two languages leads older bilinguals to have smaller frequency effects than do young bilinguals (albeit only in the nondominant language; Gollan et al., 2008). However, an implicit assumption in this line of reasoning is that frequency effects reflect a simple and cumulative count of the absolute number of exposures; this view has been questioned, at least for visual word recognition (Murray & Forster, 2004; see also Gollan et al., 2008). Moreover, it would be difficult to actually verify whether people indeed speak less than they read. And even if this lag could be verified, it might apply at only high levels of education, although there are reasons to believe that the production lag should apply more broadly. Specifically, picture naming is often difficult even for patients with very mild brain damage, whereas reading is often resistant to severe brain damage and is therefore often used in estimating predamage IQ levels (Lezak, Howieson, & Loring, 2004). The sensitivity of production but not reading to brain damage implies fundamental differences between them. A different but related reason why production might lag behind comprehension is that production may require a greater number of exposures to reach ceiling levels of proficiency than does reading. In this case, production would lag behind comprehension even if people were to speak as often as, or even more often than, they read.

The notion of a frequency lag for production relative to comprehension, or that a greater number of exposures is needed to approach ceiling levels of lexical accessibility for production than for comprehension, fits well with the observation that comprehen-

⁴ This way of estimating reading frequency was suggested to us by Mark Liberman.

sion precedes production when monolingual children learn their first language. Similarly, adults learning a new language can comprehend (and often also read) much better than they can speak the new language. That said, it is not transparent how one might test whether ceiling had been reached when comparing speaking and reading or how this might be implemented within processing models (perhaps with different selection thresholds for comprehension vs. production). What is needed here is to specify the stages of lexical access for comprehension versus production and where frequency effects are represented in each.

The Locus of the Frequency Effect

To determine more specifically why frequency, constraint, and their interaction (or lack thereof) might vary with modality, one finds it useful to consider which representations and processes might be shared or different between them. As already noted, it is generally assumed that comprehension and production access shared conceptual representations and that lexical representations are not shared at the level of form. It is also generally agreed that (even though they are not necessarily shared between modalities) lexical form representations are frequency-sensitive in both comprehension and production. But beyond these extreme end points (on which all models would agree), several different options can be considered. For production, some have argued that frequency effects are restricted exclusively to the retrieval of lexical forms (implemented as higher resting activation levels for high- than for low-frequency words within the lexical representation itself; Jescheniak & Levelt, 1994). With this rather severe restriction of the locus of frequency effects in production it is difficult to see why frequency effects would be greater in production than in comprehension without making additional assumptions (such as those just outlined in the section entitled Frequency Lag for Production Versus Comprehension).

Number of Frequency-Sensitive Processing Stages

Instead, it may be that frequency effects arise in relatively more processing stages in production than in comprehension. Consistent with this claim, Kittredge et al. (2008) recently challenged the proposal that only lexical form access is frequency-sensitive by demonstrating that semantic errors (which are often considered to arise prior to word form selection) are frequency-sensitive. That is, the probability of making a semantic related error (e.g., saying *orange* instead of *apple*) was lower for high-frequency (e.g., *apple*) than for low-frequency (e.g., *cherry*) target words, even after controlling for other variables that are correlated with frequency (e.g., imageability, density, length). Although phonological errors (e.g., *pillow* instead of *pineapple*) were more sensitive to frequency than were semantic errors, in both cases the frequency effect was highly robust (both $ps < .01$), and thus they proposed that both lexical semantics and lexical form access are frequency-sensitive. Similarly, Knobel et al. (2008) suggested that frequency effects could arise at three loci in the production architecture: in the connections from meaning to lexical form, in the lexical representations themselves, and in the connections from the lexical representations to the sounds needed to assemble spoken words.

On this view, production would involve at least one more frequency-sensitive processing stage than would comprehension.

If lexical–semantic (lemma) selection is the critical extra frequency-sensitive stage for production, then the greater effect of frequency on production than on comprehension could be explained only if lemmas are not shared between modalities. Levelt et al. (1999, p. 7) proposed two distinct layers of whole-word lexical representations in which lemmas represent semantic and syntactic information, are accessed prior to word-form representations (called lexemes), and are shared between modalities, stating the following:

Recognizing a word, whether spoken or written, involves accessing its syntactic potential, that is, the perceptual equivalent of the lemma, we assume activation of the corresponding lemma-level node. In fact . . . all production lemmas are perceptual lemmas; the perceptual and production networks coincide from the lemma level upwards.

Given the hypothesis of shared lemmas, it might instead be stipulated that the outgoing (production) connections from meaning to lemmas also are frequency-sensitive, whereas the ingoing (comprehension connections) from written and spoken forms to lemmas are not frequency-sensitive. Note also that the notion of a modality-neutral layer of lexical representations (i.e., lemmas) is not universally accepted (see e.g., Caramazza, 1997). Still, the same general problem arises in models without lemmas—some accommodation must be made to explain why frequency effects are sometimes larger in production than in comprehension despite greater sensitivity in production processes to other variables (e.g., semantic constraint).

The extent to which lemmas are or are not shared between comprehension and production also has implications for the extent to which production and comprehension can or cannot be based on each other (in whichever direction: motor theory of perception, Liberman & Mattingly, 1985; or sensory theory of speech, Venezia & Hickock, 2009) and how alignment occurs between modalities in dialogue (Pickering & Garrod, 2004). At one extreme, alignment might be pushed entirely outside the language system and up to the conceptual level; at the other extreme, any differences between modalities might be explained in terms of processing rather than representation (Ferreira, 2004). In developing such theories, it would be useful to spell out precisely how much sharing is necessary for alignment to succeed, both in terms of representation and processing involved.

It might be suggested that a different pattern of results will be found when comparing spoken production with listening comprehension (instead of reading), and if so this would open up the possibility that lemmas are shared only for listening and speaking and therefore also that frequency and constraint might interact in auditory comprehension. This would have to be tested in a separate study. However, if reading involves phonological activation (for review see Frost, 1998), the distinction between auditory and reading comprehension is weak. In addition, it could be argued that task and measure-specific factors in the current paradigms played a role in the differences we observed and that there are fundamental differences in what is being measured in comprehension and production. Naming involves elaborate planning and motor execution of a series of phonemes in a specific order, with precise pronunciation and, for nonnative speakers, also a change in accent. Although ultimately comprehension may certainly also involve specific and elaborate processing, this may not be measured ex-

perimentally (instead what is measured is processing to the point that a saccade or a decision can be made). Thus, because word frequency influences access to lexical forms for both comprehension and production, the absolute size of the effect may have more to do with the specific processes associated with form retrieval in each case and the number of exposures needed to reach ceiling levels of proficiency in each case. Although word frequency is not usually assumed to influence all of these processes, the possibility for a frequency effect at each of these stages exists.

Specifically, at least in principle, frequency could play a role at all processing steps needed to access and produce a picture name including concept selection, lexical–semantic (lemma) access, lexical form (lexeme) access, and even phonetic-articulatory planning (see section entitled Implications of Bilingual Effects). Frequency effects are usually assumed not to affect conceptual access (see e.g., Knobel et al., 2008; Oldfield & Wingfield, 1965), but some studies have shown effects along these lines (e.g., one study identified similarly sized frequency effects in lexical decision and an object decision task; Kroll & Potter, 1984). In contrast, gaze durations could primarily reflect processing steps up to the point of lexical access but not subsequent steps in the comprehension process, because conceptual access could easily continue after gaze shifts to the subsequent word. However, the presence of semantic constraint effects on gaze durations (Rayner & Well, 1996; and other effects of semantic manipulations on gaze duration; see e.g., Duffy et al., 1988) suggests that this view may be too simplistic.

Implications for Interactivity

Returning to one of the starting points for the present study, recall the apparent contrast between studies suggesting production appears inherently more interactive than comprehension. If frequency and semantic constraint affect the same processing stage (lexical–semantic retrieval), then interactions between them could be explained without positing interactivity between processing levels (i.e., without postulating that word–form retrieval can feed activation back up to influence lexical–semantic retrieval). Thus, production will appear to be more interactive than is comprehension, not because there is more interactivity in the system when going from meaning to words than the reverse but rather because relatively more processing stages are frequency-sensitive in production than in comprehension, and in production (but not in comprehension) frequency and meaning influence the same processing stage.

Within the proposal of more frequency-sensitive processing stages in production than in comprehension, it is quite impressive that frequency effects can sometimes be eliminated in production in high semantic constraint (Griffin & Bock, 1998). The relative weakness of frequency effects in language production research has also been described in this way in the context of naturally occurring speech errors: “Put briefly, the places in language generation for which lexical frequency effects should have a natural expression in speech error patterns show pallid or inconsistent effects” (Garrett, 2001, p. 234). All this is to say that although frequency effects may be large in production without semantic constraint, when semantic constraint is present, frequency effects in production are rather weak.

Implications of Bilingual Effects

Our three-tiered manipulation of English proficiency level (with monolinguals being more proficient than English-dominant Spanish–English bilinguals, who in turn were more proficient in English than Dutch-dominant Dutch–English bilinguals) provides a further means for evaluating differences between comprehension and production, while also having implications for models of bilingual language processing. Most broadly relevant, we suggested that where Participant Group \times Constraint interactions are found, this may constitute additional evidence for Frequency \times Constraint interactions assuming that bilingualism entails a frequency-lag and that there is limit on the extent to which activation can accumulate in lexical representations with frequency of use (Gollan et al., 2008). Here, we assume that the monolingual lexicon is closer to ceiling levels of activation and that monolinguals are thus accessing higher frequency lexical representations than are bilinguals. In this respect, we obtained further evidence for interactive effects of frequency and constraint in production but not in comprehension. Specifically, Dutch–English bilinguals (at the lowest level of English proficiency) benefited from semantic constraint more than did monolinguals, but only in Experiment 1 (picture naming) and not in Experiment 2 (eye tracking). Although there was some tendency in the fixation times for Spanish–English bilinguals to benefit more from context than did monolinguals, this did not seem to be related to the level of English proficiency, because Dutch–English bilinguals (less English-proficient than Spanish–English bilinguals) did not exhibit this effect. Instead, the interaction in Experiment 2 seemed to reflect a trade-off with skipping rates such that Spanish–English bilinguals’ skipping rates did not vary with context (whereas for the other two participant groups, skipping rates did vary with context). Altogether these results then again provide evidence for a Frequency of English Use \times Semantic Constraint interaction that is stronger in production than in comprehension.

The pattern of slowing related to bilingualism and second language use in Experiments 1 and 2 provides some additional evidence for separation between modalities (with distinct frequency counters in comprehension and production). In picture naming out of context, bilinguals were significantly slower than monolinguals, even when producing picture names in their dominant language (i.e., comparing Spanish–English bilinguals with monolinguals). In contrast, out-of-context reading in the lexical decision task seemed to be less sensitive to slowing related to bilingualism (slowing was statistically robust only at the lowest level of English proficiency, Dutch–English bilinguals). A possible explanation for the sometimes greater bilingual disadvantage in production than in comprehension is that bilinguals (but not monolinguals) may exhibit frequency effects in articulation. Articulation processes are usually assumed to be at ceiling for monolinguals but may not be for bilinguals, who have had less practice with producing nonnative sound combinations. The proposal that production lags behind comprehension in frequency of use (whether because speaking is a less frequent activity than is reading or because speaking requires more frequent use to achieve proficiency) also provides a ready explanation for this difference.

In this view, the frequency lag associated with bilingualism may be greater for production than for comprehension because all language users are closer to ceiling levels of lexical accessibility in

reading. Thus, Spanish–English bilinguals are closer to “catching up” with monolinguals in terms of frequency of use in reading than in speaking, and only Dutch–English bilinguals lag clearly behind monolinguals in both modalities, exhibiting significant slowing and greater frequency effects (albeit the increased frequency effect was obtained only in lexical decisions in Experiment 2). This interpretation predicts that in comprehension measures of more difficult targets than those employed here, bilinguals should exhibit a disadvantage even in the dominant language. Supporting this view, vocabulary tests (which include difficult items typically toward the end of the test) often reveal lower scores in bilinguals than in monolinguals (Bialystok, Luk, Peets, & Yang, 2010; Portocarrero, Burright, & Donovanick, 2007). Indeed, in both Experiments 1 and 2, Spanish–English bilinguals obtained slightly but significantly lower Shipley English vocabulary scores than did monolinguals (both $ps < .01$).

Another reason why slowing related to bilingualism might sometimes be more robust in production than in comprehension, and for the Participant Group \times Constraint interaction in Experiment 1, is dual-language activation. In production, a single response, and therefore also an output language, must by definition be selected, so that dual-language activation might therefore be quite disruptive (Hermans et al., 1998; but see Costa & Caramazza, 1999). In contrast, in recognition, readers may rely solely on bottom-up activation, without even having to make a language selection, as long as the input matches a stored lexical representation. To the extent that translation equivalents might become active in language comprehension, this should be relatively less disruptive, given that both activate the intended meaning. Previous studies have led to the hypothesis that semantically constraining context may act to reduce, or even eliminate entirely (Libben & Titone, 2009; Schwartz & Kroll, 2006; van Hell & de Groot, 2008), the effects of dual-language activation (but see Duyck et al., 2007; Van Assche et al., in press, and Experiment 2 of this study).

Previously, we speculated that context of any kind might have similar effects relative to no context. Anecdotally, all bilinguals have had the experience that it feels easier to retrieve words in a nondominant language when surrounded by context that supports activation of that language. However, we obtained no evidence that slowing related to bilingualism or second language use was attenuated by low-constraint context (in production where we could make this comparison; see the last paragraph in the Results section of Experiment 1). Even with high-constraint semantic context, only Dutch–English bilinguals benefited more from context than did monolinguals, but both bilingual groups responded more slowly than did monolinguals (in Experiment 1). In comprehension there were even some effects in the opposite direction. That is, Spanish–English bilinguals were only marginally slower than monolinguals in lexical decisions but had significantly longer gaze duration times (see Table 4 for a summary). This could be an uninteresting consequence of the different types of measures (lexical decision vs. gaze durations), or the contrast could suggest a kind of trade-off such that context reduces some aspects of the bilingual disadvantage but also introduces some disadvantage associated with the added burden of having to process syntax in a nondominant language.

Thus, our inclusion of participants at three different levels of English proficiency strengthens the conclusions drawn here by

demonstrating generalizability to a broad range of types of language users and by manipulating frequency in more than one way (comparing high- and low-frequency words and also comparing speakers who have used English exclusively/monolinguals vs. those who have used English relatively less frequently as a second language). At the same time, the results reported here provided additional evidence to constrain the frequency-lag account of bilingual disadvantages, demonstrating that it is more robust in production than in comprehension and challenging this account by showing that slowing associated with bilingualism can arise without any interaction with frequency (e.g., in gaze duration times in Experiment 2). Similarly, our results constrain the interference model of bilingual disadvantages by demonstrating that context does not reduce bilingual disadvantages much (only at the lowest proficiency level and only if it is a highly semantically constraining context). Importantly, the two accounts of the bilingual disadvantage (frequency lag and dual-language activation) are not mutually exclusive (see also Gollan et al., 2008), and it is likely that more than one mechanism will be needed to explain processing differences between bilinguals and monolinguals, including disadvantages and advantages (Bialystok, Craik, Green, & Gollan, 2009).

Conclusions

In the present study, we considered the joint effects of frequency, semantic constraint, and language proficiency on lexical access in language comprehension and production and provided empirical evidence pointing to fundamental differences in lexical search between comprehension and production, thus confirming theoretically motivated reasons for expecting these differences. More specifically, multiple aspects of our data suggested that (a) production is a process driven by semantic constraint, whereas (b) comprehension is frequency-driven. In addition, relative to comprehension, production involves one or more of the following options: less practice, more difficulty, and/or a greater number of frequency-sensitive processing stages. For this reason, (c) frequency effects can sometimes be larger in production than in comprehension (even in monolingual speakers and even though access in production is semantically driven), (d) frequency and constraint interact in production but not (or much less so) in comprehension, and (e) bilingual disadvantages and language dominance effects are sometimes larger for production than for comprehension. It remains to be seen whether production will always necessarily lag behind comprehension for lexical access out of semantic context or whether practice could attenuate this basic difference between modalities. However, the fundamental differences in cognitive processes underlying lexical selection in comprehension versus production likely cannot be attenuated by any amount of language use. In this respect, the data and interpretation presented here do two things: one, demonstrate flexibility in the language system to adjust its representational and search mechanisms in order to optimize the differing needs when deriving meaning from lexical form compared with finding forms to express intended meanings, and two, place constraints on the extent to which the mechanisms of language comprehension and production can be shared or aligned in naturally occurring language use.

References

- Almeida, J., Knobel, M., Finkbeiner, M., & Caramazza, A. (2007). The locus of the frequency effect in picture naming: When recognizing is not enough. *Psychonomic Bulletin & Review*, *14*, 1177–1182.
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, *4*, 439–461.
- Awad, M., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. (2007). A common system for the comprehension and production of narrative speech. *Journal of Neuroscience*, *27*, 11455–11464. doi:10.1523/JNEUROSCI.5257-06.2007
- Baayen, R. H., Piepenbrock, R., & Glikers, L. (1995). The CELEX lexical database [CD-ROM]. Philadelphia, PA: Philadelphia Linguistic Data Consortium, University of Pennsylvania.
- Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language*, *59*, 495–523. doi:10.1016/j.jml.2007.10.004
- Bialystok, E., Craik, F. I. M., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. *Psychological Science in the Public Interest*, *10*, 89–129.
- Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*, 525–531.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208. doi:10.1080/026432997381664
- Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. C. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning Memory & Cognition*, *27*, 1430–1450. doi:10.1037/0278-7393.27.6.1430
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*, 257–271.
- Coltheart, M. (2001). Assumptions and methods in cognitive neuropsychology. In B. Rapp (Ed.), *The handbook of cognitive neuropsychology: What deficits reveal about the human mind* (pp. 3–21). Ann Arbor, MI: Psychology Press.
- Costa, A. (2005). Lexical access in bilingual production. In J. F. Kroll and A. M. B. de Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic approaches* (pp. 308–325). New York, NY: Oxford University Press.
- Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish-English and English-Spanish bilinguals. *Bilingualism: Language and Cognition*, *2*, 231–244. doi:10.1017/S1366728999000334
- 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*, 1283–1296. doi:10.1037/0278-7393.26.5.1283
- Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 318–344. doi:10.1037/0278-7393.25.2.318
- Davies, M. (2007). *Time Magazine corpus: 100 million words, 1920s–2000s*. Retrieved September 1, 2008, from <http://corpus.byu.edu/time>
- Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *5*, 313–349. doi:10.1080/01690969008407066
- Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon friends or foes? In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 9–38). Berlin, Germany: Mouton de Gruyter.
- Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*, 175–197.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*, 429–446. doi:10.1016/0749-596X(88)90066-6
- Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 663–679. doi:10.1037/0278-7393.33.4.663
- Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency-effect in second-language visual word recognition. *Psychonomic Bulletin & Review*, *15*, 850–855. doi:10.3758/PBR.15.4.850
- Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word recognition and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*, 641–655. doi:10.1016/S0022-5371(81)90220-6
- Fay, D., & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*, *8*, 505–520.
- Ferreira, F. (2004). Production-comprehension asymmetries. *Behavioral and Brain Sciences*, *27*, 196. doi:10.1017/S0140525X04280050
- Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Language Verbal Behavior*, *18*, 1–20. doi:10.1016/S0022-5371(79)90534-6
- Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology*, *33A*, 465–495.
- Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*, 627–635. doi:10.1016/S0022-5371(73)80042-8
- Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 680–698. doi:10.1037/0278-7393.10.4.680
- Frost, R. (1998). Towards a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, *123*, 71–99. doi:10.1037/0033-2909.123.1.71
- Garrett, M. (2001). Now you see it, now you don't: Frequency effects in language production. In E. Doupoux (Ed.), *Language, brain and cognitive development* (pp. 227–240). Cambridge, MA: MIT Press.
- Gernsbacher, M. A., & Faust, M. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 245–262. doi:10.1037/0278-7393.17.2.245
- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*, 787–814. doi:10.1016/j.jml.2007.07.001
- Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*, 1220–1234.
- Gollan, T. H., Montoya, R. I., & Werner, G. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, *16*, 562–576. doi:10.1037/0894-4105.16.4.562
- Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew-English bilinguals. *Bilingualism: Language and Cognition*, *4*, 63–83. doi:10.1017/S136672890100013X
- Graves, W. W., Grabowski, T. J., Mehta, S., & Gordon, J. K. (2007). A neural signature of phonological access: Distinguishing the effects of word frequency from familiarity and length in overt picture naming. *Journal of Cognitive Neuroscience*, *19*, 617–631. doi:10.1162/jocn.2007.19.4.617
- Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the

- relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, 38, 313–338. doi:10.1006/jmla.1997.2547
- Hand, C. J., Mielliet, S., O'Donnell, P. J., & Sereno, S. C. (2010). The frequency-predictability interaction in reading: It depends where you're coming from. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1294–1313. doi:10.1037/a0020363
- Hermans, D., Bongaerts, T., de Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, 1, 213–229. doi:10.1017/S1366728998000364
- Hillis, A. (2001). The organization of the lexical system. In B. Rapp (Ed.), *The handbook of cognitive neuropsychology: What deficits reveal about the human mind* (pp. 185–210). Ann Arbor, MI: Psychology Press.
- Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word probability. *Journal of Experimental Psychology*, 41, 401–410. doi:10.1037/h0056020
- Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, 127, 277–288. doi:10.1016/j.actpsy.2007.06.003
- Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 824–843. doi:10.1037/0278-7393.20.4.824
- Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probability theory in linguistics* (pp. 39–96). Cambridge, MA: MIT Press.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354. doi:10.1037/0033-295X.87.4.329
- Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, 25, 463–492. doi:10.1080/02643290701674851
- Knobel, M., Finkbeiner, M., & Caramazza, A. (2008). The many places of frequency: Evidence for a novel locus of the lexical frequency effect in word production. *Cognitive Neuropsychology*, 25, 256–286. doi:10.1080/02643290701502425
- Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica*, 128, 416–430. doi:10.1016/j.actpsy.2008.02.001
- Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior*, 23, 39–66. doi:10.1016/S0022-5371(84)90499-7
- Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- La Heij, W. (2005). Selection processes in monolingual and bilingual lexical access. In J. F. Kroll and A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 289–307). New York, NY: Oxford University Press.
- Lehtonen, M., & Laine, M. (2003). How word frequency affects morphological processing in monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 6, 213–225. doi:10.1017/S1366728903001147
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38. doi:10.1017/S0140525X99001776
- Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological Assessment* (4th ed.). New York, NY: Oxford University Press.
- Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 381–390. doi:10.1037/a0014875
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74, 431–461. doi:10.1037/h0020279
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36. doi:10.1016/0010-0277(85)90021-6
- Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences*, 13, 110–114. doi:10.1016/j.tics.2008.11.008
- Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition*, 7, 485–495.
- MacKay, I., & Flege, J. (2004). Effects of the age of second-language (L2) learning on the duration of L1 and L2 sentences: The role of suppression. *Applied Psycholinguistics*, 25, 373–396. doi:10.1017/S0142716404001171
- Mägiste, E. (1979). The competing language systems of the multilingual: A developmental study of decoding and encoding processes. *Journal of Verbal Learning and Verbal Behavior*, 18, 79–89. doi:10.1016/S0022-5371(79)90584-X
- Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 503–535. doi:10.1037/0278-7393.33.3.503
- Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007, July 6). Are women really more talkative than men? *Science*, 317, 82. doi:10.1126/science.1139940
- Murray, W. S. (2000). Sentence processing: Issues and measures. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 649–664). Oxford, England: Elsevier. doi:10.1016/B978-008043642-5/50030-9
- Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, 111, 721–756. doi:10.1037/0033-295X.111.3.721
- Navarrete, E., Basagni, B., Alario, F. X., & Costa, A. (2006). Does word frequency affect lexical selection in speech production? *Quarterly Journal of Experimental Psychology*, 59, 1681–1690. doi:10.1080/17470210600750558
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theory. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.
- Nicoladis, E., Palmer, A., & Marentette, P. (2007). The role of type and token frequency in using past tense morphemes correctly. *Developmental Science*, 10, 237–254. doi:10.1111/j.1467-7687.2007.00582.x
- Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 17, 273–281. doi:10.1080/17470216508416445
- Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Kimbrough, D. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, 18, 41–58. doi:10.1017/S014271640009863
- Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, 7(2), 20, 1–36.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral Brain Sciences*, 27, 169–190. doi:10.1017/S0140525X04000056
- Portocarrero, J. S., Burrell, R. G., & Donovanick, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology*, 22, 415–422. doi:10.1016/j.acn.2007.01.015
- Raney, G. E., & Rayner, K. (1995). Word frequency effects and eye

- movements during two readings of a text. *Canadian Journal of Experimental Psychology*, 49, 151–173. doi:10.1037/1196-1961.49.2.151
- Ransdell, S. E., & Fischler, I. (1987). Memory in a monolingual mode: When are bilinguals at a disadvantage? *Journal of Memory and Language*, 26, 392–405. doi:10.1016/0749-596X(87)90098-2
- Rapp, B. (2001). *The handbook of cognitive neuropsychology: What deficits reveal about the human mind*. Ann Arbor, MI: Psychology Press.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 86, 446–461. doi:10.1037/0033-2909.86.3.446
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422. doi:10.1037/0033-2909.124.3.372
- Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62, 1457–1506. doi:10.1080/17470210902816461
- Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z reader model. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 720–732. doi:10.1037/0096-1523.30.4.720
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14, 191–201.
- Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 53, 1061–1080. doi:10.1080/02724980050156290
- Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (in press). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*.
- Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3, 504–509.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105, 125–157. doi:10.1037/0033-295X.105.1.125
- Roelofs, A. (2003). Modeling the relation between the production and recognition of spoken word forms. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 115–158). Berlin, Germany: Mouton de Gruyter.
- Rosselli, M., Ardila, A., Araujo, K., Weekes, V. A., Caracciolo, V., Padilla, M., & Ostrosky-Solis, F. (2000). Verbal fluency and repetition skills in healthy older Spanish-English bilinguals. *Applied Neuropsychology*, 7, 17–24. doi:10.1207/S15324826AN0701_3
- Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 9, 487–494. doi:10.1016/S0022-5371(70)80091-3
- 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. doi:10.1017/S1366728909990514
- Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 1–17. doi:10.1037/0096-1523.3.1.1
- Schiller, N. O., & Meyer, A. S. (2003). *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Berlin, Germany: Mouton de Gruyter.
- Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, 26, 1270–1281.
- Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, 29, 86–102. doi:10.1016/0749-596X(90)90011-N
- Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, 55, 197–212. doi:10.1016/j.jml.2006.03.004
- Sebastián-Gallés, N., & Kroll, J. F. (2003). Phonology in bilingual language processing: Acquisition, perception, and production. In N. Schiller & A. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 279–317). Berlin, Germany: Mouton de Gruyter.
- Shelton, J. R., & Caramazza, A. (2001). The organization of semantic memory. In B. Rapp (Ed.), *The handbook of cognitive neuropsychology: What deficits reveal about the human mind* (pp. 423–444). Ann Arbor, MI: Psychology Press.
- Shipley, W. C. (1946). *Institute of Living Scale*. Los Angeles, CA: Western Psychological Services.
- Slattery, T. J. (2009). Word misperception, the neighbor frequency effect, and the role of sentence context: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1969–1975. doi:10.1037/a0016894
- Stanovich, K. E., & West, R. F. (1979). Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory & Cognition*, 7, 77–85.
- Staub, A., White, S. J., Drieghe, D., Hollway, E. C., & Rayner, K. (2010). Distributional effects of word frequency on eye fixation durations. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1280–1293. doi:10.1037/a0016896
- Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645–659. doi:10.1016/S0022-5371(79)90355-4
- Taylor, S. E. (1965). Eye movements in reading: Facts and fallacies. *American Educational Research Journal*, 2, 187–202.
- Treiman, R., Clifton, C., Jr., Meyer, A. S., & Wurm, L. H. (2003). Language comprehension and production. In I. B. Weiner, A. F. Healy, R. W. Proctor, & J. A. Schinka (Eds.), *Handbook of Psychology: Vol. 4. Experimental psychology* (pp. 527–548). New York, NY: Wiley.
- Tulving, E., & Gold, C. (1963). Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *Journal of Experimental Psychology*, 66, 319–327. doi:10.1037/h0048802
- Van Assche, E., Duyck, W., Hartsuiker, R., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in sentence context. *Psychological Science*, 20, 923–927. doi:10.1111/j.1467-9280.2009.02389.x
- Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, 64, 88–107.
- van Hell, J. G., & de Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica*, 128, 431–451. doi:10.1016/j.actpsy.2008.03.010
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, 18, 380–393.
- Venezia, J. H., & Hickok, G. (2009). Mirror neurons, the motor system, and language: From the motor theory to embodied cognition and beyond. *Language and Linguistics Compass*, 3, 1403–1416. doi:10.1111/j.1749-818X.2009.00169.x
- Vincent, S. B. (1912). The function of the vibrissae in the behavior of the white rat. *Behavioral Monographs*, 1(No. 5), 1–81.

Yap, M. J., & Balota, D. A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 274–296. doi:10.1037/0278-7393.33.2.274

Zwitserslood, P. (2003). The internal structure of words: Consequences for listening and speaking. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 79–114). Berlin, Germany: Mouton de Gruyter.

Appendix

First Fixation Duration and Total Fixation Times

Fixations shorter than 80 ms that were within one character of a previous or subsequent fixation were combined with that fixation; all other fixations less than 80 ms were eliminated (2.5% of all fixations). Trials with track loss on or near the target word were also eliminated (4.6% of all trials). In addition, mean first fixation duration, gaze duration, and total time on the target were calculated for each participant, and trials in which these measures were more than 2.5 standard deviations above the participant's mean were excluded from analyses. This trimming procedure excluded 4.2% of the monolinguals', 4.2% of the Spanish–English bilinguals', and 4.4% of the Dutch–English bilinguals' remaining sentence-reading trials.

First fixations on high-frequency target words were shorter than those on low-frequency targets, $F_1(1, 96) = 46.67$, $MSE = 347.25$, $\eta_p^2 = .33$, $p < .001$; $F_2(1, 88) = 29.35$, $MSE = 822.53$, $\eta_p^2 = .25$, $p = .001$. However, the main effect of constraint was only marginally significant on first fixation duration, $F_1(1, 96) = 2.87$, $MSE = 334.08$, $\eta_p^2 = .03$, $p = .094$; $F_2(1, 88) = 2.58$, $MSE = 811.63$, $\eta_p^2 = .03$, $p = .112$. There was also a main effect of group, $F_1(2, 96) = 3.63$, $MSE = 2,510.09$, $\eta_p^2 = .07$, $p < .05$; $F_2(2, 176) = 28.51$, $MSE = 582.29$, $\eta_p^2 = .25$, $p < .001$. Planned comparisons indicate that first fixation durations for monolinguals and Spanish–English bilinguals were not significantly different in the participants analysis ($F < 1$), but monolinguals' first fixation durations were shorter than Spanish–English bilinguals' in the items analysis, $F_2(1, 89) = 4.24$, $p < .05$. Additionally, Dutch–English bilinguals had longer first fixations than did monolinguals ($p = .016$) and marginally longer first fixations than did Spanish–English bilinguals ($p = .057$).

There was no indication of a Frequency \times Constraint interaction in the first fixation data (the benefit of a semantically constraining context was not larger for low- than high-frequency targets; both $F_s < 1$); there was no Group \times Frequency interaction (both $F_s < 1$); slowing related to bilingualism was not greater for low- than high-frequency targets; and there was no three-way interaction between these variables, $F_1(2, 96) = 1.12$, $MSE = 273.33$, $\eta_p^2 = .02$, $p = .33$; $F_2(2, 176) = 1.35$, $MSE = 571.79$, $\eta_p^2 = .02$, $p = .263$. There was a significant Constraint \times Group interaction, $F_1(2, 96) = 3.82$, $MSE = 333.36$, $\eta_p^2 = .07$, $p < .05$; $F_2(2, 176) = 3.80$, $MSE = 571.79$, $\eta_p^2 = .04$, $p < .05$, not the interaction predicted on the basis of the three-tiered manipulation of English proficiency; Spanish–English bilinguals benefited more from semantically constraining contexts than did Dutch–English bilin-

guals ($ps < .05$) and marginally more than did monolinguals ($ps < .06$), but constraint effects between the two groups that differed most in English proficiency (monolinguals and Dutch–English bilinguals) did not differ (both $F_s < 1$). Summarizing the first fixation data, the most striking difference with respect to Experiment 1 was the lack of interactions in the reading data but the presence of interactions in the production data: between frequency and constraint and between frequency and participant group.

The pattern of results for total fixation times was similar to that found for first fixations and gaze duration. That is, total fixation time was shorter on high- than low-frequency targets, $F_1(1, 96) = 78.47$, $MSE = 1,397.63$, $\eta_p^2 = .45$, $p < .001$; $F_2(1, 88) = 23.58$, $MSE = 6,206.42$, $\eta_p^2 = .21$, $p < .001$, and shorter on high- than low-constraint targets, $F_1(1, 96) = 11.83$, $MSE = 2,002.53$, $\eta_p^2 = .11$, $p < .001$; $F_2(1, 88) = 11.21$, $MSE = 3,943.63$, $\eta_p^2 = .11$, $p < .001$, and there was again a main effect of group indicating that total fixation times were shortest for monolinguals, $F_1(2, 96) = 16.58$, $MSE = 14,584.22$, $\eta_p^2 = .26$, $p < .001$; $F_2(2, 176) = 102.37$, $MSE = 3,788.89$, $\eta_p^2 = .54$, $p = .001$. Planned comparisons indicate that total fixation times for monolinguals and Spanish–English bilinguals were shorter than for Dutch–English bilinguals ($ps < .001$); however, the difference between monolinguals and Spanish–English bilinguals was only marginal ($ps < .08$). An examination of the interaction effects in the total time measure showed that there was still no indication that frequency interacted with constraint (both $F_s < 1$). Additionally, for the first time in the reading data, there was some hint of a greater frequency effect in bilinguals. However, this Participant Group \times Frequency interaction was only marginally significant in the participants analysis, $F_1(2, 96) = 3.08$, $MSE = 1,397.63$, $\eta_p^2 = .06$, $p = .051$, and did not approach significance in the items analysis $F_2(2, 176) = 1.29$, $MSE = 3,788.89$, $\eta_p^2 = .01$, $p = .278$. Finally, the Group \times Constraint interaction that had been significant for first fixation and gaze duration was not significant for total fixation time, $F_1(2, 96) = 1.15$, $MSE = 2,002.53$, $\eta_p^2 = .02$, $p = .322$; $F_2(2, 176) = 1.89$, $MSE = 3,121.03$, $\eta_p^2 = .02$, $p = .154$, and the three-way interaction did not approach significance ($F_s < 1$).

Received March 26, 2010

Revision received October 27, 2010

Accepted October 27, 2010 ■

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

1