In the Mind's Eye: Eye-Tracking and Multi-modal Integration During Bilingual Spoken-Language Processing

Sarah Chabal [1,*]

Email schabal@u.northwestern.edu

Viorica Marian [1]

[1] Department of Communication Sciences and Disorders, Northwestern University, 2240 N Campus Dr., Evanston, IL, 60208 USA

Abstract

The human language system integrates information from multiple sources and modalities. Bilinguals, who experience increased processing demands due to competition between their two languages, may be especially likely to rely on cues from multiple modalities. The cross-modal integration involved in language processing has been frequently studied using eye-tracking, an approach that can accommodate the simultaneous presence of both auditory and visual inputs. Eye-tracking research has demonstrated that bilinguals activate both of their languages in parallel (e.g., Spivey and Marian 1999; Weber and Cutler 2004), leading to competition both within (e.g., the English word "marker" competes with the phonologically similar word "marble") and between (e.g., the English word "marker" competes with the Russian word "marka") their two languages (Marian and Spivey 2003a, b). Interestingly, this competition arises even when languages do not share phonology (e.g., bimodal bilinguals; Shook and Marian 2012) and in the absence of explicit linguistic input (Chabal and Marian ~~2013~~ 2015), demonstrating the highly interactive nature of bilingual language processing. This interactivity in the language system yields changes to bilingual cognitive function. For example, bilinguals' need to control phonological competition is associated with enhanced executive control (Blumenfeld and Marian 2011), and their experience suppressing information from the non-target language improves their ability to learn novel vocabulary (Bartolotti and Marian 2012). We conclude that multi-modal investigations of language processing such as those employing eye-tracking are not only ecologically valid, as they closely resemble real-world multi-modal situations, but also can demonstrate how language interacts with other cognitive and perceptual functions in a non-modular mind.

Language is a part of our organism and no less complicated than it. … The tacit conventions on which the understanding of everyday language depends are enormously complicated.
Ludwig Wittgenstein, May 14, 1915

AQ1

Language is more than words—it is the amalgamation of sounds and simultaneously co-occurring visual inputs (e.g., facial features, gestures and nearby objects) that combine to mutually influence one another. The human linguistic system[1] is equipped to automatically integrate and process information from multiple modalities and to use this integrated information to enhance processing. Perhaps the most well-known demonstration of the interactivity of the language processing system is the McGurk effect

(McGurk and MacDonald 1976), which is a perceptual phenomenon in which competing auditory and visual information lead to the perception of an un-presented sound. For example, if a listener hears the phoneme /ba/ while simultaneously watching a video of a speaker's lips producing /ga/, the combination of the competing inputs often leads listeners to perceive the phoneme /da/, even though the sound /da/ was never presented. When one of the modalities is stripped away (i.e., participants receive only auditory or only visual input) participants' perceptions match the input. The McGurk effect holds even when the sources of auditory and visual information are presented with a separation in space (Jones and Munhall 1997; MacDonald et al. 2000) or with a discrepancy in temporal presentation (Van Wassenhove et al. 2007) and therefore do not appear to emanate from the same source (for a demonstration of the McGurk effect see auditoryneuroscience.com/McGurkEffect).

The robustness of the McGurk effect and other demonstrations of audio-visual integration (e.g., co-speech hand gestures: Kelly et al. 2004; processing of facial articulatory movements: Van Wassenhove et al. 2005), coupled with evidence that even infants appear able to combine auditory and visual inputs by matching facial movements with their corresponding sounds (Kuhl and Meltzoff 1982), together lend credence to the idea that integration from multiple modalities is important for, and inherent to, language perception and processing (see also the motor theory of speech perception, e.g., Liberman and Mattingly (1985), for an example of integration across motor and auditory domains). This cross-modal integration may be of even greater significance for bilingual speakers, who rely on auditory and visual cues to interpret information from multiple languages.

In the beginning stages of second language learning, visual input can aid in the acquisition of non-native phonemic contrasts (e.g., Hardison 2003; Hazan et al. 2002) and novel vocabulary (Plass et al. 1998). For example, while learning German, native English speakers are better able to remember word translations when they are presented both verbally and with a visual representation (i.e., picture or video clips; Plass et al. 1998). Visual input, in the form of facial cues (e.g., Ronquest and Hernandez 2005; Soto-Faraco et al. 2007) or head movements (e.g., Davis and Kim 2006), can also be used to determine the language of an interlocutor. In fact, bilingual infants are better than monolingual infants at discriminating between languages based on visual input alone (i.e., silent videos of speakers; Sebastián-Gallés et al. 2012; Weikum et al. 2007). The multi-modality of the language system, then, appears to be a crucial part of bilingual language acquisition and perception.

AQ2

Furthermore, the importance of an interactive language system is not limited to the beginning stages of bilingualism. Even for highly skilled, balanced bilinguals, who use both of their languages on a daily basis, there is a discrepancy between auditory perceptual abilities in their two languages. For example, French-Spanish bilinguals may fail to perceive lexical stress in Spanish (Dupoux et al. 2010), and Catalan-Spanish bilinguals may be unable to recognise certain phonemic contrasts in Catalan in spite of fully developed perceptual abilities in Spanish (Sebastián-Gallés et al. 2005). Such perceptual difficulties in a second language can be overcome by integrating information from the speech stream with visual information provided by facial articulatory gestures (Navarra and Soto-Faraco 2007). Visual input can also aid bilinguals' retrieval of lexical items (Schroeder and Marian, in prep). In speech production, visual and motor input from the use of hand gestures has been linked to better linguistic access (for a review see Krauss and Hadar 1999). In fact, bilinguals often produce more hand gestures than their monolingual peers (see Nicoladis 2007 for a review of bilinguals' use of hand gestures). Together, such findings demonstrate the positive effects of multi-modal processing within a bilingual context.

9.1. Eye-Tracking as an Index of a Multi-~~m~~Modal Language

System

There are a variety of methods that have been developed to explore the interactive nature of the language processing system, including neuroimaging (e.g., Calvert et al. 1997; Campbell et al. 2001) and electrophysiological techniques (e.g., Kelly et al. 2004; Musacchia et al. 2007); computational modelling (e.g., Dupont and Luettin 2000; Shook and Marian 2013); and behavioural methods such as cross-modal priming (e.g., Berry et al. 1997; Loveman et al. 2002) and visual word recognition (e.g., Baron and Strawson 1976). A particularly powerful method is eye-tracking, which can index participants' real-time reactions to simultaneous auditory and visual inputs.

Eye-tracking provides a means of measuring gaze and eye motion around a visual scene by exploiting the reflective properties of the eye. In the most common eye-tracking systems, infrared light, unseen by the participant, illuminates the pupil and cornea. Calculating the angle between the two yields a measure of gaze direction, which can be calibrated to a visual scene.

Current eye-tracking methods are capable of tracking eye movements at sampling rates of up to 2,000 Hz, and provide high temporal resolution that allows for the exploration of real-time language processing.

A popular technique that relies on eye-tracking to study language is the visual world paradigm (e.g., Cooper 1974; Tanenhaus et al. 1995), in which participants carry out spoken instructions to touch, move, or manipulate objects in a visual workspace that is either real or presented on a computer screen (e.g., Tanenhaus and Spivey-Knowlton 1996). Such work has informed our understanding of language processing by illustrating the incremental nature of comprehension. In other words, spoken-language processing occurs over time as auditory information unfolds, and multiple lexical candidates can be partially activated (Marslen-Wilson 1987). For example, the spoken word "candy" will activate words such as "cap," "can," and "candle," as the sounds "c-a-n-d-y" emerge over time. Eye movements are sensitive to these instances of multiple activation, and a participant who is instructed to "Pick up the *candy*" will glance at a *candle* that is also in the visual display (e.g., Allopenna et al. 1998; Tanenhaus et al. 1995). Because participants' eye movements to objects in a visual scene are closely time-locked to the auditory references to those same objects, eye-tracking, by indexing those eye movements, provides an on-line measure of how language comprehension progresses.

By exploiting these temporal dynamics of auditory language processing, eye-tracking research has been able to demonstrate that not only is language processed incrementally and influenced by relationships between linguistic features, but it is also immediately influenced by relevant visual information. For example, when items with phonologically overlapping names are simultaneously present in a search display (e.g., "candy" and "candle") and participants are instructed to interact with one of those objects (e.g., "Click on the candy"), not only do they make eye movements to the phonologically related competitor (e.g., candle), but eye movements are also slower to the target when the competitor is present (e.g., Tanenhaus et al. 1995, 1996). Whereas participants are able to identify the target an average of 55 ms *before* the offset of the target name when no competitor is present, target identification takes place approximately 30 ms *after* word-offset when there is a competitor present (Tanenhaus et al. 1995). Because auditory information is identical across competitor-present and competitor-absent trials and the only variable factor is the visual input, such studies provide evidence that relevant visual context affects the early stages of spoken-language processing.

In addition to providing good temporal correspondence that allows for the indexing of on-line language processing, eye-tracking is also recognised for its ability to be used in natural language contexts. When words are embedded in natural sentence structures, eye movements can be measured without disrupting spoken input. This allows researchers to explore real-time language comprehension (as discussed throughout this chapter) and production (e.g., Griffin and Bock 2000; Meyer et al. 1998).

9.2. Using Eye-Tracking to Explore Bilingual Multi-~~m~~Modal Language Processing

Because of its ability to index natural language with good temporal correspondence, eye-tracking has found its niche not only in monolingual language processing research but also in the exploration of bilingual language processing. Within bilingualism research, eye-tracking was first introduced to explore whether bilinguals process their two languages in parallel or separately. Parallel processing assumes that both of a bilingual's two languages are activated simultaneously; conversely, separate or sequential processing assumes that a bilingual speaker or listener has selective access only to the language system that is currently in use. While some behavioural evidence suggested that a bilingual's languages were independently activated and that interference did not occur across languages (e.g., Durgunolu and Roediger 1987; Gerard and Scarborough 1989; Kirsner et al. 1980; Ransdell and Fischler 1987; Scarborough et al. 1984; Watkins and Peynircioglu 1983), other studies provided reason to believe that bilingual language processing may occur in parallel, with both languages becoming simultaneously activated and mutually influencing one another (e.g., bilingual Stroop task: Chen and Ho 1986; Preston and Lambert 1969; Tzelgov et al. 1990; cross-linguistic priming: Beauvillain and Grainger 1987).

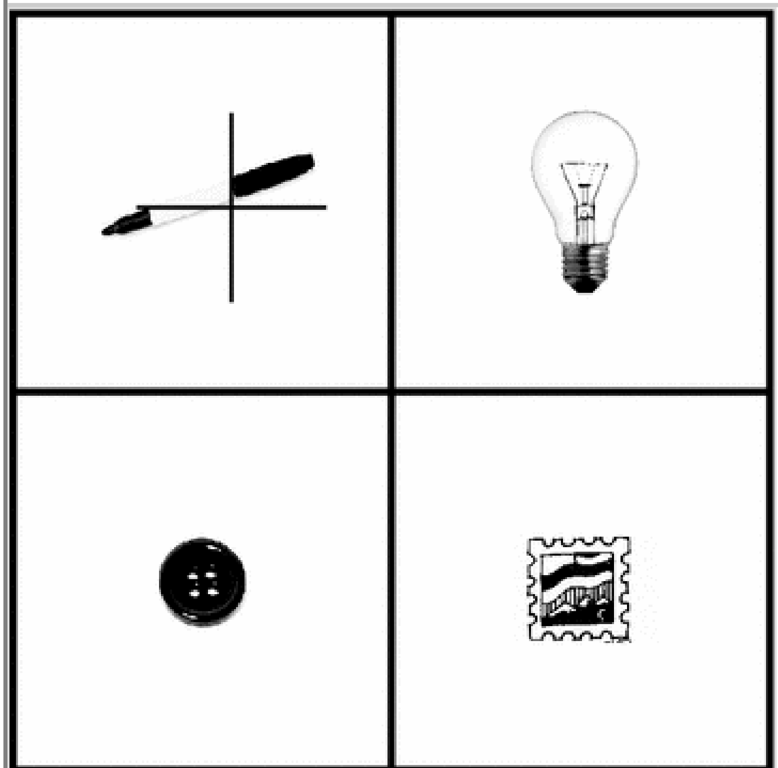
However, in tasks such as cross-linguistic priming paradigms and the bilingual Stroop (in which participants see printed colour words in one language and are instructed to name the ink colour of those words in another language, i.e., the Spanish word for blue, "azul", written in yellow ink), both of a bilingual's languages are overtly cued. It is therefore not clear whether parallel processing occurs when only one language is intentionally accessed. In the Stroop task, for example, visual input is provided in one language and production happens in the second; in cross-linguistic priming tasks both languages are visually presented. What is needed, then, is a methodology that allows for the activation of two languages to be tested while only requiring input or output in a single language.

Techniques that probe the activation of a language without overt cuing have been used to explore parallel language access in bilingual production (e.g., Colomé 2001; Costa et al. 2000; Hoshino and Kroll 2008; for a review see Kroll et al. 2006), written comprehension (e.g., Midgley et al. 2008; Morford et al. 2011; Thierry and Wu 2007; Van Heuven et al. 1998), and spoken comprehension (Marian and Spivey 2003a, b; Spivey and Marian 1999). Specifically, in spoken comprehension, eye movements are often used to index language processing in bilinguals (but see also Thierry and Wu 2007 for an example of how electrophysiological techniques can be used to explore parallel language access).

Using the visual world paradigm, Spivey and Marian (1999) provided the first eye-tracking evidence of parallel language activation during bilingual language comprehension. They tested Russian-English bilinguals in monolingual Russian sessions in which participants were requested, for example, to pick up the postage stamp ("Poloji marku nije krestika"). In competitor conditions, the target object (e.g., "*marka,*" Russian for "postage stamp") was accompanied by an object whose English name shared initial phonological features with the target (e.g., "marker"). As the auditory instructions unfolded, incoming phonological information mapped to both of the bilinguals' languages, and participants made looks to the phonologically competing "marker", even though input was only received in Russian (see Fig. 9.1). Moreover, not only do bilinguals experience between-language competition (e.g., "marker" competes with "marka") but, like monolinguals, they also experience within-language competition in each of their two languages (e.g., "candy" competes with "candle" in English, and "spichki" [matches] competes with "spitsy" [knitting needles] in Russian; Marian and Spivey 2003a). Although within-language competition is typically stronger than between-language competition (Marian and Spivey 2003b), bilinguals must still contend with multiple sources of competition—in contrast to monolinguals who only encounter competition within a single language.

Fig. 9.1

An illustration of a search display showing a Russian-English bilingual's fixations (crosshairs) on the phonologically competing "marker" (*top left quadrant*) when instructed in Russian to pick up the postage stamp (*marka* in Russian; *bottom right quadrant*). (Adapted from Spivey and Marian 1999, Fig. 1.)



Furthermore, to add to these already challenging processing demands, within-language competition and between-language competition can occur simultaneously (Marian and Spivey 2003b). When Russian-English bilinguals were presented with visual world displays that contained a target object (e.g., "speaker") paired with both a within-language competitor (whose English name was phonologically similar to the target, e.g., "spear") and a between-language competitor (whose Russian name was phonologically similar to the target; e.g., "spichki" [matches]), they made eye movements to both the Russian and the English competitors. This suggests that bilinguals simultaneously experience competing activation from lexical items that overlap phonologically both within and across their two languages. Importantly, these findings are robust and have been replicated in language pairs including Dutch and English (e.g., Lagrou et al. 2013; Weber and Cutler 2004), French and German (e.g., Weber and Paris 2004), Spanish and English (e.g., Canseco-Gonzalez et al. 2010; Ju and Luce 2004), and Japanese and English (e.g., Cutler et al. 2006).

Not only do these eye-tracking studies demonstrate that both of a bilingual's languages are activated in parallel, but they also illustrate how the surrounding visual display interacts in real time with the phonological information being received by the bilingual listener (Marian 2009). This audio-visual integration during spoken-language processing has been explored computationally using the Bilingual Language Interaction Network for Comprehension of Speech (BLINCS; Shook and Marian 2013), which is, to our knowledge, the only model to date that illustrates and predicts bilingual spoken-language activation as it occurs over time (as individual phonemes unfold) within a constraining visual environment (see Fig. 9.2). Specifically, the model receives a word~~,~~ one phoneme at a time and, after each phonemic unit, determines which words (in two languages) best match that input. Lexical units that match the phonemic input receive activation, with activation levels of each unit changing over time as additional phonemes are introduced into the model. Simultaneously, information about visual representations is integrated into the semantic level of linguistic processing, so that words with meanings that map more closely to the semantic information provided by the visual input receive a greater amount of activation. Through direct top-down links between the semantic and phono-lexical levels, BLINCS is able to simulate how objects in a visual scene affect the activation (and eventual selection) of words within the bilingual's lexicon and can make predictions that are supported by behavioural eye-tracking data (see Fig. 9.3).

Fig. 9.2

The *Bilingual Language Interactive Network for Comprehension of Speech* (*BLINCS*) is equipped to integrate visual information with unfolding auditory input to model and predict bilingual spoken-language activation as it occurs over time within a constraining visual environment. (From Shook and Marian 2012, Fig. 1.)

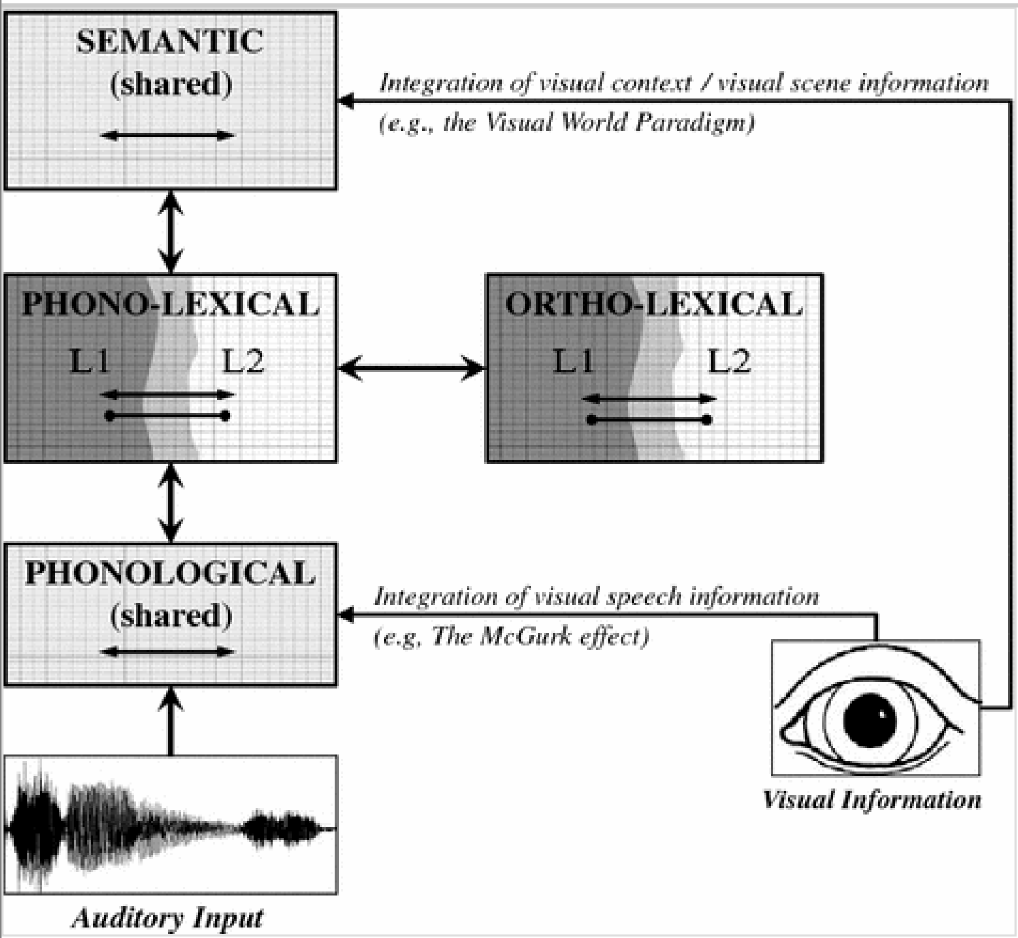
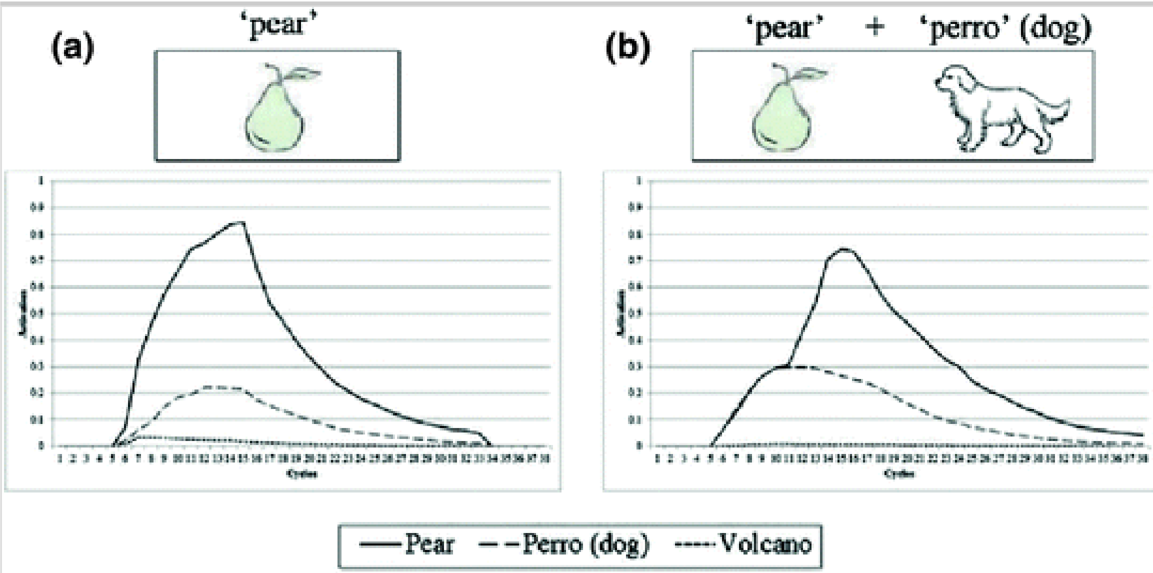


Fig. 9.3

Activation of the BLINCS model during auditory presentation of the word "pear" accompanied by visual presentation of *pear* alone (**a**) and *pear* and *dog* [*perro* in Spanish] (**b**). These figures illustrate how BLINCS allows for interactions between auditory and visual inputs within the bilingual language processing system. (Adapted from Shook and Marian 2013, Fig. 11 Panels A and B.)

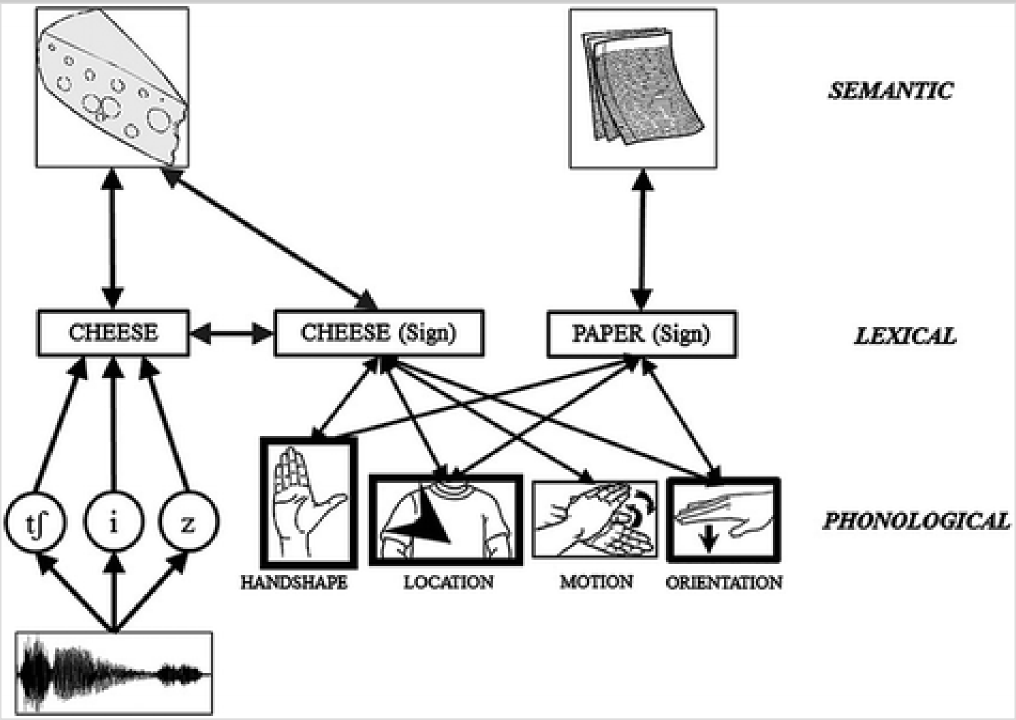


One prediction made by BLINCS is that competition and dual-language activation in bilinguals can occur both from bottom-up phonological input *and* from top-down connections.

In fact, recent eye-tracking work supports this hypothesis by demonstrating that bimodal bilinguals, who are users of a spoken and a signed language such as English and American Sign Language (ASL), experience competition between their two languages (Shook and Marian 2012). Shook and Marian instructed ASL-English bilinguals to click on a target image, such as "cheese." Embedded in the computerised display was a competitor item ~~that~~whose name overlapped ~~with~~in ASL on three of four phonological parameters—hand-shape, motion, location of the sign in space, and orientation of the palm or hand. For example, the ASL sign for "paper" overlaps with "cheese" in hand-shape, location, and orientation, whereas the pair differs in the motion of the signs. Importantly, none of the target-competitor pairs overlapped in English phonology. Participants were instructed, in English only, to click on the target item (i.e., "Click on the cheese") while their eye movements were recorded. Recall that in the classic visual world paradigm studies, unfolding auditory information leads to competition between items in the non-target language (e.g., "Click on the speaker" leads to looks to the matches, "spichki" in Russian; Marian and Spivey 2003b). However, in Shook and Marian's study, the bilinguals' languages did not share auditory phonology. Nevertheless, despite a lack of unfolding English phonological information that would lead to the activation of "paper" when "cheese" was heard, ASL-English bilinguals gave more looks to the paper than did English monolinguals, and they also gave more looks to the paper than to control items that did not overlap with the target in ASL phonology. This demonstrates that languages are activated in parallel even when bottom-up, featural information is only available for one of the languages, and provides support for a language system that includes top-down or lateral connections (see Fig. 9.4). It also suggests that information in a visual display can affect linguistic processing even if the visual input cannot be mapped directly onto the incoming language stream.

Fig. 9.4

Proposed pathways showing *bottom-up* (e.g., auditory phonological information to the lexical item "cheese"), *lateral* (e.g., the spoken lexical item "cheese" to the sign for cheese), and *top-down* (e.g., the lexical sign for cheese to the phonological feature of handshape) connections between languages. When bimodal bilinguals were instructed in English to find the "cheese," they made eye movements to the "paper," which shares phonological overlap with cheese in American Sign Language but not in English. (From Shook and Marian 2012 , Fig. 4.)



Further support for a highly interactive bilingual language system and the role of visual input in language activation comes from recent research suggesting that bilingual language access can proceed even in the absence of *any* bottom-up linguistic information. Chabal and Marian (~~2013~~ 2015) presented Spanish-English bilinguals with an image of a target object (e.g., a *clock*, "reloj" in Spanish) and asked them to search for that item in a subsequent visual display. Even though the objects' linguistic forms were never explicitly presented (i.e., there was no auditory input), participants made eye movements to objects whose names overlapped with the target in English (e.g., a *cloud*, "nube" in Spanish) and to objects whose names overlapped with the target in Spanish (e.g., a *present*, "regalo" in Spanish). The same image, therefore, was able to lead to access of both of the bilinguals' languages. This demonstrates the ubiquity of multi-modal linguistic processing by showing that visual input alone may be sufficient to spark language access. Moreover, it provides strong evidence for the parallel access of both of a bilingual's languages because the viewing of pictures activated not only English but also Spanish, which was never used within the context of the experiment (all task instructions and experimenter interactions occurred in English only).

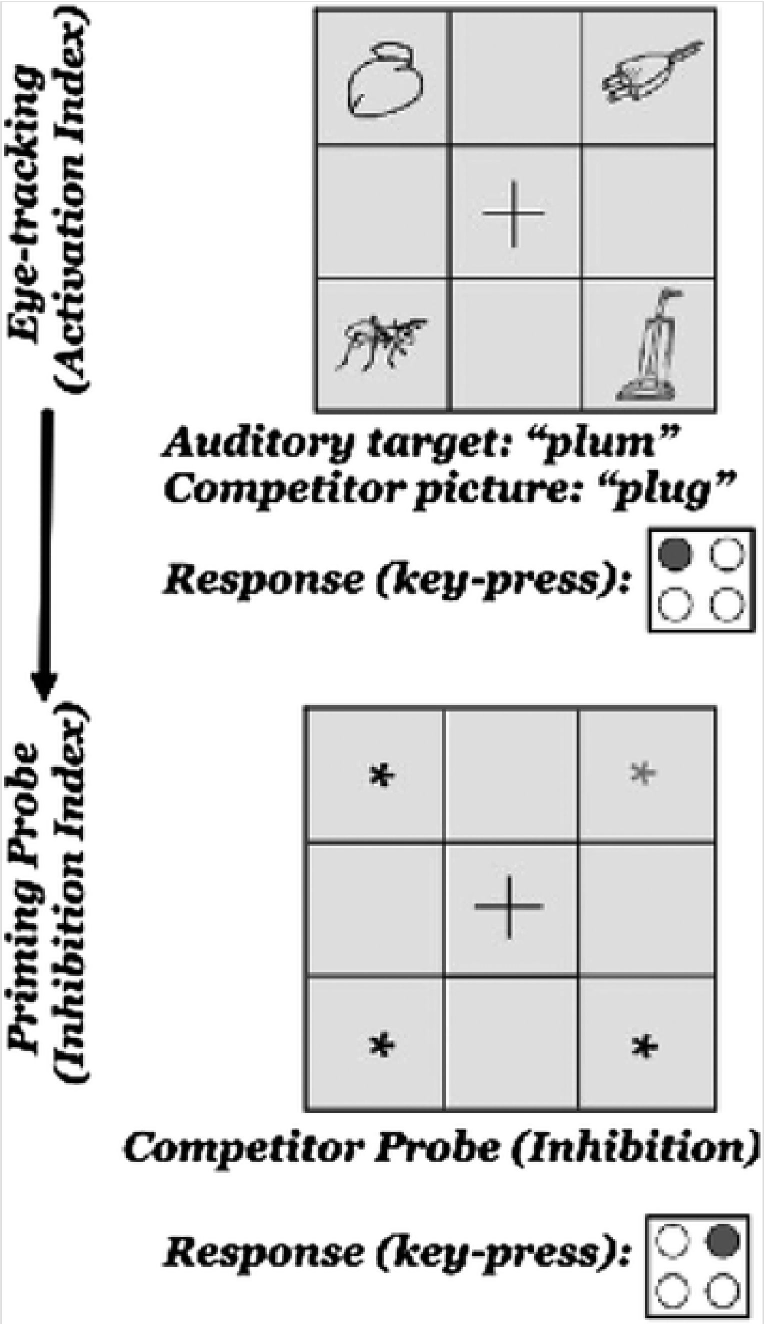
9.3. Consequences of an Interactive Bilingual Language System

One consequence of the highly interactive nature of the bilingual language system is that bilinguals must develop strategies to cope with the constant activation of both of their languages. It has been proposed (e.g., Kroll 2008) that as a result of suppressing information from the unneeded language and of attending only to the relevant language, bilinguals have enhanced executive function abilities relative to monolinguals (e.g., Bialystok 2006, 2008; Costa et al. 2008; Martin-Rhee and Bialystok 2008; Prior and Macwhinney 2009). For example, bilingual children have been found to outperform monolingual children on tasks requiring attentional control (e.g., Martin-Rhee and Bialystok 2008), and similar bilingual advantages have been observed across the lifespan (e.g., Bialystok et al. 2004; Bialystok 2008; Costa et al. 2009). It is important to note, however, that the extent of bilinguals' executive function gains and inhibitory abilities may be dependent upon a number of factors including proficiency levels across their two languages (e.g., Khare et al 2013; Singh and Mishra 2013), their amount of experience within a bilingual environment (Bialystok and Barac 2012), and the age at which they became actively exposed to both languages (Luk et al. 2011). Nevertheless, consistent with the multi-modality of the bilingual experience, executive function advantages are seen in the auditory (e.g., Soveri et al. 2010), visual (for a review see Bialystok 2011), and combined audio-visual (e.g., Bialystok et al. 2006; Krizman et al. 2012) domains.

However, the link between bilinguals' need to control interference within and across their languages and their performance on executive tasks remained largely inferential. Bilinguals were known to activate both of their languages and were known to display enhanced cognitive control abilities, but a direct connection between the two had never been empirically demonstrated. In order to explain how the processing of ambiguous auditory information (which could lead to activation of multiple words within the lexicon) can be associated with executive function, Blumenfeld and Marian (2011) tested monolinguals and bilinguals on a visual world paradigm task containing an item whose name overlapped phonologically with the name of a spoken target (e.g., "Click on the plum" while a plug was present in the search display). Following each eye-tracking trial, participants completed a negative priming task to probe residual activation or inhibition of locations that had previously contained competitor items (see Fig. 9.5). Although monolinguals and bilinguals both experienced similar competition between the phonologically overlapping items, as evidenced by eye movements to competitors, the groups differed in how they responded to the negative priming probes. Specifically, bilinguals inhibited the visual location of the auditorily received input, and the strength of this inhibition was correlated with their performance on a non-linguistic executive control task. For the monolingual group, inhibition of phonological competitors did not relate with the group's executive control abilities. By exploiting the tight link between incoming auditory information and the corresponding visual representations, Blumenfeld and Marian provided empirical support for the idea that cognitive control mechanisms can be affected by linguistic experience.

Fig. 9.5

Sample visual display from the eye-tracking and negative priming paradigms. Participants were presented with an eye-tracking trial, in which they were instructed to locate the picture of a spoken object while another object on the display competed phonologically within the same language, English (e.g., *plum*-*plug*) (*top* of figure). Next, the inhibition of the competitor item was explored by asking participants to quickly locate a shaded asterisk that was positioned where the phonological competitor had been previously (*bottom* of figure). Although both English monolinguals and Spanish–English bilinguals made looks to phonological competitor pictures, bilinguals displayed less residual inhibition of competitor locations than monolinguals, suggesting that their inhibitory processes may be more efficient. Consistent with this interpretation, bilinguals' (but not monolinguals') inhibition on the priming probes was inversely correlated with their performance on a non-linguistic executive control task. (Adapted from Blumenfeld and Marian ~~2012~~2011 Fig. 1 Panel A.)



These enhancements in bilinguals' executive function render a number of practical advantages. For example, bilinguals' ability to avoid distraction from irrelevant languages may be one reason that they are better than monolinguals at learning a new language's vocabulary (e.g., Cenoz 2003; Kaushanskaya and Marian 2009; Keshavarz and Astaneh 2004). To test this possibility, Bartolotti and Marian (2012) trained monolinguals and bilinguals to equivalent levels on vocabulary in a made-up language called Colbertian (matching the groups on novel-language proficiency ensured that effects were not due to one group learning better than the other). Interference from participants' previously known language (i.e., English) was then tested using the visual world paradigm. As in other visual world paradigm studies we have discussed in this chapter, participants were presented with a target object (e.g., an acorn, called "shundo" in Colbertian) and a distractor object whose English name either overlapped (e.g., "shovel") or did not overlap (e.g., "mushroom") with the newly learned name of the target object. As the phonological information, "shundo," unfolded, bilinguals made fewer looks to competitor items than did monolinguals. Although both monolinguals and bilinguals experienced competition between items that competed phonologically, Bartolotti and Marian's findings suggest that the groups differ in how they manage this competition. Specifically, bilinguals were better able to suppress competition from their previously known languages, which may be a contributing factor to observed language-learning benefits.

Together, Bartolotti and Marian's (2012) and Blumenfeld and Marian's (2011) studies illustrate how eye-tracking evidence can be combined with knowledge gleaned from other behavioural, neuroimaging and electrophysiological techniques to explore how and why bilinguals' multi-modal linguistic experiences shape their cognitive systems. Methods that allow for the integration of auditory and visual inputs (such as eye-tracking) will advance the field of bilingual research by facilitating the exploration of higher order cognitive processes such as language, memory, attention, and decision making.

9.4. Constraints of the Visual World Paradigm

Although eye-tracking provides bilingual researchers with an invaluable tool to index language processing and audio-visual integration within a spoken-word context, there are a few limitations (as with any behavioral or neuroimaging technique) that must be kept in mind. First, because of constraining study environments, researchers conducting eye-tracking experiments must select only a small subset of visual objects and auditory tokens when designing their studies. In reality, the human language experience is not confined to utterances concerning only the objects within our immediate visual scenes. While a number of studies using the visual world paradigm have included full sentence structures (e.g., Tanenhaus et al. 1995), and some have investigated eye movements after the relevant visual scene has been removed (e.g., Altmann 2004), eye-tracking studies continue to rely primarily on tightly controlled auditory and visual inputs.

Second, effects of phonological competition within the visual world paradigm are susceptible to subtle task and presentation manipulations. For example, the amount of time that the visual images are displayed before the onset of an auditory stimulus affects the types of information that participants are able to access about those pictures (e.g., shape, semantic, or phonological information; Huettig and McQueen 2007), and phonological competition effects can be abolished by eliminating or providing only a short preview of the search display (for a review of how the mechanisms of visual processing affect language-mediated eye movements, see Dahan et al. 2007).

9.5. Conclusions

What we learn only through the ears makes less impression upon our minds than what is presented to the trustworthy eye.

Horace

Eye-tracking with the visual world paradigm continues to be a valuable method for researchers interested in both the auditory and visual components of the language system. The use of eye-tracking to examine both the architecture of the bilingual language system and the consequences of bilingualism for broader cognitive functioning has provided crucial insight into the bilingual experience and the multi-modality of language. For example, eye-tracking research has demonstrated that a bilingual's two languages are simultaneously activated (e.g., Ju and Luce 2004; Spivey and Marian 1999; Weber and Cutler 2004) and interact in bottom-up and top-down fashions (Shook and Marian 2012). In fact, both of a bilingual's languages are activated even when neither is being used (Chabal and Marian ~~2013~~2015). As a consequence, bilinguals must contend with competition arising both within and between their two languages (Marian and Spivey 2003a, b), which bolsters their executive functioning abilities

(Blumenfeld and Marian 2011). More efficient executive functioning abilities, in turn, manifest in practical benefits such as the ability to more easily learn additional languages (Bartolotti and Marian 2012).

Such advances in the field of bilingual language processing and cognition can, in part, be attributed to the use of eye-tracking. By combining information from auditory and visual sources to closely resemble real-world, multi-modal situations, and by allowing language processing to proceed in a naturalistic context (Huettig et al. 2011), the visual world paradigm provides an ecologically valid methodology for studying language processing. At the intersection of perceptual (e.g., auditory and visual) and higher order processing, eye-tracking techniques can be used to explore how language interacts with other cognitive functions in a highly interconnected, non-modular mind (e.g., Prinz 2006).

Acknowledgments

Preparation of this chapter was supported in part by grant NICHD R01HD059858 to Viorica Marian. The authors would like to thank Anthony Shook, Scott Schoeder, James Bartolotti, Jennifer Krizman and Tuan Lam for helpful comments on this work.

References

Allopenna, P., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419–439. doi: 10.1006/jmla.1997.2558 .

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The "blank screen paradigm". *Cognition, 93*(2), B79–B87. doi: 10.1016/j.cognition.2004.02.005 .

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance, 2*(3), 386–393. doi: 10.1037/0096-1523.2.3.386 .

Bartolotti, J., & Marian, V. (2012). Language learning and control in monolinguals and bilinguals. *Cognitive Science, 36*, 1129–1147. doi: 10.1111/j.1551-6709.2012.01243.x .

Beauvillain, C., & Grainger, J. (1987). Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of Memory and Language, 26*(6), 658–672.

Berry, D. C., Banbury, S., & Henry, L. (1997). Transfer across form and modality in implicit and explicit memory. *The Quarterly Journal of Experimental Psychology A, 50*(1), 1–24. doi: 10.1080/027249897392206 .

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology, 60*(1), 68–79. doi: 10.1037/cjep2006008 .

Bialystok, E. (2008). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12*(01), 3. doi: 10.1017/S1366728908003477 .

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of*

Experimental Psychology, 65(4), 229–235. doi: 10.1037/a0025406 .

Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition, 122*(1), 67–73. doi: 10.1016/j.cognition.2011.08.003 .

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*(2), 290–303. doi: 10.1037/0882-7974.19.2.290 .

Bialystok, E., Craik, F. I. M., & Ruocco, A. C. (2006). Dual-modality monitoring in a classification task: The effects of bilingualism and ageing. *Quarterly Journal of Experimental Psychology, 59*(11), 1968–1983. doi: 10.1080/17470210500482955 .

Blumenfeld, H. K., & Marian, V. (2011). Bilingualism influences inhibitory control in auditory comprehension. *Cognition, 118*(2), 245–257. doi: 10.1016/j.cognition.2010.10.012 .

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., & David, A. S. et al. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*(5312), 593–596. doi: 10.1126/science.276.5312.593 .

Campbell, R., MacSweeney, M., Surguladze, S., Calvert, G. A., McGuire, P. K., Suckling, J., & David, A. S. et al. (2001). Cortical substrates for the perception of face actions: An fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Cognitive Brain Research*, *12*(2), 233–243.

Canseco-Gonzalez, E., Brehm, L., Brick, C., Brown-Schmidt, S., Fischer, K., & Wagner, K. (2010). Carpet or carcel: The effect of age of acquisition and language mode on bilingual lexical access. *Language and Cognitive Processes, 25*(5), 669–705. doi: 10.1080/01690960903474912 .

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism, 7*(1), 71–89. doi: 10.1177/13670069030070010501 .

~~Chabal, S., & Marian, V. (2013). Language experience modifies the processing of visual input. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*. Berlin, Germany.~~Chabal, S., & Marian, V. (2015). Speakers of different languages process the visual world differently. *Journal of Experimental Psychology: General, 144*(3).

Chen, H., & Ho, C. (1986). Development of stroop interference in Chinese-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(3), 397–401. doi: 10.1037/0278-7393.12.3.397 .

Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language, 45*(4), 721–736. doi: 10.1006/jmla.2001.2793 .

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. *Cognitive Psychology, 107*, 84–107.

Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 26*(5), 1283–1296. doi: 10.1037/0278-7393.26.5.1283 .

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition, 106*(1), 59–86. doi: 10.1016/j.cognition.2006.12.013 .

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition, 113*(2), 135–149. doi: 10.1016/j.cognition.2009.08.001 .

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*(2), 269–284. doi: 10.1016/j.wocn.2005.06.002 .

Dahan, D., Tanenhaus, M. K., & Salverda, A. P. (2007). The influence of visual processing on phonetically driven saccades in the "visual world" paradigm. In R. P. G. van Gompel, H. Fischer, Martin, S. Murray, Wayne, & L. Hill, Robin (Eds.), *Eye movements: A window on mind and brain* (p. 720). Amsterdam, Netherlands: Elsevier.

Davis, C., & Kim, J. (2006). Audio-visual speech perception off the top of the head. *Cognition, 100*(3), B21–B31. doi: 10.1016/j.cognition.2005.09.002 .

Dupont, S., & Luettin, J. (2000). Audio-visual speech modeling for continuous speech recognition. *IEEE Transactions on Multimedia, 2*(3), 141–151. doi: 10.1109/6046.865479 .

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: Stress "deafness" in simultaneous French-Spanish bilinguals. *Cognition, 114*(2), 266–275. doi: 10.1016/j.cognition.2009.10.001 .

Durgunolu, A. Y., & Roediger, H. L. (1987). Test differences in accessing bilingual memory. *Journal of Memory and Language, 26*(4), 377–391. doi: 10.1016/0749-596X(87)90097-0 .

Gerard, L. D., & Scarborough, D. L. (1989). Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 15*(2), 305–315. doi: 10.1037/0278-7393.15.2.305 .

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*(4), 274–279. doi: 10.1111/1467-9280.00255 .

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24*(04), 495–522. doi: 10.1017/S0142716403000250 .

Hazan, V., Sennema, A., & Faulkner, A. (2002). Audiovisual perception in L2 learners. In *Proceedings of ICSLP* (pp. 1685–1688).

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition, 106*(1), 501–511. doi: 10.1016/j.cognition.2007.02.001 .

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57*(4), 460–482. doi: 10.1016/j.jml.2007.02.001 .

Huettig, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica, 137*(2), 138–150. doi: 10.1016/j.actpsy.2010.07.013 .

Jones, J. A., & Munhall, K. G. (1997). The effects of separating auditory and visual sources on audiovisual integration of speech. *Canadian Acoustics, 25*(519), 13–20.

Ju, M., & Luce, P. A. (2004). Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science, 15*(5), 314–318. doi: 10.1111/j.0956-7976.2004.00675.x .

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review, 16*(4), 705–710. doi: 10.3758/PBR.16.4.705 .

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*(1), 253–260. doi: 10.1016/S0093-934X(03)00335-3 .

Keshavarz, M. H., & Astaneh, H. (2004). The impact of bilinguality on the learning of English vocabulary as a foreign language (L3). *International Journal of Bilingual Education and Bilingualism, 7*(4), 295–302. doi: 10.1080/13670050408667814 .

Khare, V., Verma, A., Kar, B., Srinivasan, N., & Brysbaert, M. (2013). Bilingualism and the increased attentional blink effect: Evidence that the difference between bilinguals and monolinguals generalizes to different levels of second language proficiency. *Psychological Research, 77*(6), 728–737. doi: 10.1007/s00426-012-0466-4 .

Kirsner, K., Brown, H. L., Abrol, S., Chadha, N. K., & Sharma, N. K. (1980). Bilingualism and lexical representation. *The Quarterly Journal of Experimental Psychology, 32*(4), 585–594. doi: 10.1080/14640748008401847 .

Krauss, R. M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. In R. Campbell & L. Messing (Eds.), *Gesture, speech, and sign* (pp. 93–116). Oxford: Oxford University Press.

Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences of the United States of America, 109*(20), 7877–7881. doi: 10.1073/pnas.1201575109 .

Kroll, J. F. (2008). *Juggling two languages in one mind*. Psychological Science Agenda: American Psychological Association. 22.

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism, 9*(02), 119. doi: 10.1017/S1366728906002483 .

Kuhl, P., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*, 1138–1141.

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2013). Interlingual lexical competition in a spoken sentence context: Evidence from the visual world paradigm. *Psychonomic Bulletin & Review,20(5), 963-972*. doi: 10.3758/s13423-013-0405-4 .

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1–36.

Loveman, E., Van Hooff, J. C., & Gale, A. (2002). A systematic investigation of same and cross modality priming using written and spoken responses. *Memory, 10*(4), 267–276. doi: 10.1080/09658210143000380 .

Luk, G., De Sa, E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, *14*(4), 1–8. doi: 10.1017/S1366728911000010 .

MacDonald, J., Andersen, S., & Bachmann, T. (2000). Hearing by eye: How much spatial degradation can be tolerated? *Perception, 29*(10), 1155–1168. doi: 10.1068/p3020 .

Marian, V. (2009). Audio-visual integration during bilingual language processing. In A. Pavlenko (Ed.), *The bilingual mental lexicon: Interdisciplinary approaches* (pp. 52–78).

Marian, V., & Spivey, M. J. (2003a). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics, 24*(~~0~~2), 173–193. doi: 10.1017/S0142716403000092 .

Marian, V., & Spivey, M. J. (2003b). Competing activation in bilingual language processing: Within- and between-language. *Bilingualism: Language and Cognition*, *6*(2), 97–115. doi: 10.1017/S1366728903001068 .

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1–2), 71–102.

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *11*(01), 81–93. doi: 10.1017/S1366728907003227 .

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*(2), B25–33.

Midgley, K. J., Holcomb, P. J., VanHeuven, W. J. B., & Grainger, J. (2008). An electrophysiological investigation of cross-language effects of orthographic neighborhood. *Brain Research, 1246*, 123–135. doi: 10.1016/j.brainres.2008.09.078 .

Morford, J. P., Wilkinson, E., Villwock, A., Piñar, P., & Kroll, J. F. (2011). When deaf signers read

English: Do written words activate their sign translations? *Cognition, 118*(2), 286–292. doi: 10.1016/j.cognition.2010.11.006 .

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America, 104*(40), 15894–15898. doi: 10.1073/pnas.0701498104 .

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological Research, 71*(1), 4–12. doi: 10.1007/s00426-005-0031-5 .

Nicoladis, E. (2007). The effect of bilingualism on the use of manual gestures. *Applied Psycholinguistics, 28*(03), 441–454. doi: 10.1017/S0142716407070245 .

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology, 90*(1), 25–36. doi: 10.1037//0022-0663.90.1.25 .

Preston, M. S., & Lambert, W. E. (1969). Interlingual interference in a bilingual version of the stroop color-word task. *Journal of Verbal Learning and Verbal Behavior, 8*(2), 295–301. doi: 10.1016/S0022-5371(69)80079-4 .

Prinz, J. J. (2006). Is the mind really modular? In R. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 22–36). Oxford: Blackwell.

Prior, A., & MacWhinney, B. (2009). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13*(~~0~~2), 253. doi: 10.1017/S1366728909990526 .

Ransdell, S. E., & Fischler, I. (1987). Memory in a monolingual mode: When are bilinguals at a disadvantage? *Journal of Memory and Language, 26*(4), 392–405. doi: 10.1016/0749-596X(87)90098-2 .

Ronquest, R., & Hernandez, L. (2005). *Lip-reading skills in bilinguals: Some effects of L1 on visual-only language identification*. *Research on Spoken Language Processing, Progress Report No. 27* (pp. 219–226). Bloominton, Indiana.

Scarborough, D. L., Gerard, L. D., & Cortese, C. (1984). Independence of lexical access in bilingual word recognition. *Journal of Verbal Learning and Verbal Behavior, 23*(1), 84–99. doi: 10.1016/S0022-5371(84)90519-X .

Schroeder, S. R., & Marian, V. (in preparation). Audio-visual interactions during memory encoding in monolinguals and bilinguals.

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*(2), 240–255. doi: 10.1016/j.jml.2004.11.001 .

Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science, 23*(9), 994–999. doi:

10.1177/0956797612436817 .

Shook, A., & Marian, V. (2012). Bimodal bilinguals co-activate both languages during spoken comprehension. *Cognition, 124*(3), 314–324. doi: 10.1016/j.cognition.2012.05.014 .

Shook, A., & Marian, V. (2013). The bilingual language interaction network for comprehension of speech (BLINCS). *Bilingualism: Language and Cognition*, ~~(September)~~*16*, ~~1–21~~304-324. doi: 10.1017/S1366728912000466 .

Singh, N., & Mishra, R. K. (2013). Second language proficiency modulates conflict-monitoring in an oculomotor stroop task: Evidence from Hindi-English bilinguals. *Frontiers in Psychology, 4*~~(June)~~, 322. doi: 10.3389/fpsyg.2013.00322 .

Soto-Faraco, S., Navarra, J., Weikum, W. M., Vouloumanos, A., Sebastián-Gallés, N., & Werker, J. F. (2007). Discriminating languages by speech-reading. *Perception and Psychophysics, 69*(2), 218–231. doi: 10.3758/BF03193744 .

Soveri, A., Laine, M., Hämäläinen, H., & Hugdahl, K. (2010). Bilingual advantage in attentional control: Evidence from the forced-attention dichotic listening paradigm. *Bilingualism: Language and Cognition*, *14*(03), 371–378. doi: 10.1017/S1366728910000118 .

Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*(3), 281. doi: 10.1111/1467-9280.00151 .

Tanenhaus, M. K., & Spivey-Knowlton, M. (1996). Eye-tracking. *Language and Cognitive Processes, 11*(6), 583–588.

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634. doi: 10.1126/science.7777863 .

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K. M., & Sedivy, J. C. (1996). Using eye movements to study spoken language comprehension: Evidence for visually mediated incremental interpretation. In *Attention and performance XVI: Information integration in perception and communication* (pp. 457–478). Boston: The MIT Press.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences of the United States of America, 104*(30), 12530–12535. doi: 10.1073/pnas.0609927104 .

Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(5), 760–771. doi: 10.1037/0278-7393.16.5.760 .

Van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39*(3), 458–483. doi: 10.1006/jmla.1998.2584 .

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America, 102*(4), 1181. doi: 10.1073/pnas.0408949102 .

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*(3), 598–607. doi: 10.1016/j.neuropsychologia.2006.01.001 .

Watkins, M. J., & Peynircioglu, Z. F. (1983). On the nature of word recall: Evidence for linguistic specificity. *Journal of Verbal Learning and Verbal Behavior, 22*(4), 385–394. doi: 10.1016/S0022-5371(83)90246-3 .

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50*(1), 1–25. doi: 10.1016/S0749-596X(03)00105-0 .

Weber, A., & Paris, G. (2004). The origin of the linguistic gender effect in spoken-word recognition: Evidence from non-native listening. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, 1446–1451.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-faraco, S., Sebastián-gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science, 316*, 1159.

[1] The vocabulary that we use to describe functions such as language, cognition, and perception (e.g., “linguistic system”) implicitly identifies entities as distinct from one another. However, language, cognition, and perception are not distinct modules, but rather are part of a highly interactive network. In order to understand how this system operates, we use math symbols (e.g., computational models) or verbal labels (e.g., words) to describe particular functions of the network. The inclusion of these terms does not preclude interactivity, but rather gives us a way to describe functions and refer to concepts either verbally or symbolically.