Bilingual memory: structure, access, and processing

James Bartolotti and Viorica Marian

Abstract

Language and memory are closely intertwined in the human cognitive architecture. Language acquisition depends on successful memory encoding and retrieval; at the same time, language itself is instrumental for encoding and storing knowledge. For bilinguals, the need to keep their two languages functionally distinct influences memory. In this chapter, we review the structure of bilingual memory, including long-term, short-term, and phonological working memory and how they are influenced by knowledge of multiple languages. We also investigate memory access and review research on episodic memory access in bilinguals and on semantic memory access during bilingual language comprehension and production. We then examine processing in the context of existing models of bilingual language and memory. Finally, we consider how the prism of novel language learning can provide insight into the interaction between memory and language. We conclude that bilingualism changes the human cognitive architecture and affects the structure, access, and processing of language and memory.

Successful acquisition and use of language requires the storage in memory of many words, their associated concepts, and grammatical rules. Access to these items in memory is accomplished with relative ease. The process by which language is stored, accessed, and processed is remarkable, yet becomes even more impressive when bilingualism is considered. A bilingual must not only store information pertaining to two languages, but also be able to access and process linguistic information according to changing linguistic contexts. The two languages

* Preparation of this chapter was supported in part by National Institutes of Health grant RO1 HD059858 to VM. The authors would like to thank Anthony Shook, Scott Schroeder, and Sarah Chabal for comments and feedback on earlier versions of this manuscript. Correspondence regarding this chapter should be directed to James Bartolotti at j-bartolotti@u.northwestern.edu or at the Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208.
have the potential to compete for memory resources and processing capacity. One way to avoid competitive interference would be for the bilingual architecture to include two systems that store, access, and process information in a language-specific manner. However, this characterization of the bilingual as the sum of two monolingual minds is inconsistent with a wealth of research that indicates high interactivity between the bilingual’s two languages in memory and language processing (Grosjean, 1989). In the current chapter, we consider how multiple languages interact to influence memory storage, access, and processing. Whether memory is accessed selectively from one language or whether both languages are retrieved automatically has implications for long-term memory organization. Similarly, the way in which information is processed depends on what becomes activated during memory access, and how information is encoded and stored in long-term memory.

We begin by focusing on the structure of long- and short-term memory. Long-term memory is composed of substructures specific to knowledge categories, and the degree to which language-specific information is represented may differ between these categories. The manner in which information in short-term memory is processed by the working memory system is also discussed. Next, we examine how stored representations in long-term memory are accessed. Once again, we consider whether the language of retrieval affects recall success, indicating language-specific access, or whether both languages become activated automatically, indicating close integration between a bilingual’s multiple languages. To address this question we examine access to episodic memory and retrieval of semantic knowledge during recognition and production of language. The chapter ends with a focus on how bilinguals process their two languages. We review the organization of several theoretical and computational models, and their capabilities in capturing aspects of bilingual language processing, after which we explore how bilingualism affects the ability to process and encode novel information, such as during novel language learning.

**Structure of bilingual memory**

The primary division between types of memory is made according to the timescale over which information is retained. This separation has its origins in James’s (1890) *Principles of Psychology*, which differentiated between primary memory for recent experiences and secondary memory for information retained over a long period of time. The distinction received renewed attention during the 1960s with the development of
the field of cognitive psychology (Neisser, 1967) and its attempts to describe the information processing capabilities of the mind. Atkinson and Shiffrin (1968) defined short-term memory as an information-maintenance system that controlled access and encoding to long-term memory. Evidence for a separation between short- and long-term memories came from patients with amnesia, who seemed to display specific impairments to one of the two memory systems (Baddeley & Warrington, 1970; Milner, 1966). The underlying architectures of long- and short-term memory are debated, but there is agreement on their functional distinction. In this section, we will consider first the storage of language in long-term memory, and then the effect of language processing and information encoding in the short-term memory system.

**Long-term memory in bilinguals**

Long-term memory (LTM) contains stable representations of knowledge acquired over time, including explicit memory for facts and events, and implicit memory for skills, routines, and associations. Explicit and implicit memory can be dissociated and appear to involve distinct neural components (Eichenbaum & Cohen, 2001; Mishkin, Malamut, & Bachevalier, 1984; Poldrack & Packard, 2003; Squire & Knowlton, 2000; Voss & Paller, 2008). They also differ with respect to how language is involved. Explicit memory can be consciously demonstrated by verbally recounting an event or by providing an answer to a query, while implicit memory can only be demonstrated as a non-conscious change in performance due to information gained over time. Both explicit and implicit memory play important roles in language acquisition and processing (Morgan-Short, 2007); implicit memory contributes to acquisition of grammar (Ullman, 2004), but explicit memory has been the focus of more extensive study in research on bilingual cognition (e.g., Kroll & de Groot, 1997; Pavlenko, 2000).

Explicit memory can be further divided into semantic memory for general facts, including word–meaning associations, and episodic memory for events and their linguistic environment. Memory models currently disagree on the specifics of semantic and episodic memory consolidation (the process by which information is encoded and stored in LTM). For example, memory consolidation theory (Paller, 1997; Scoville & Milner, 1957; Squire, Cohen, & Nadel, 1984) maintains that both semantic and episodic memories are formed by hippocampal binding of information across neocortical sites. Over time, the paired associations between neocortical sites strengthen, and the hippocampus is relied on less to reactivate memories (McClelland, McNaughton, &
O’Reilly, 1995; Norman & O’Reilly, 2003; Paller, 1997; Rempel-Clower et al., 1996). Patients with hippocampal lesions are unable to consolidate new semantic and episodic memories but demonstrate preserved recall for facts and events prior to the injury, which are thought to be stored across neocortical sites. In contrast, the multiple memory trace theory separates the processes governing semantic and episodic memory storage. (Moscovitch et al., 2005; Nadel et al., 2000; Rosenbaum et al., 2005). Episodic memory is thought to always rely on the hippocampus for retrieval, whereas semantic memory is stored in the neocortex without hippocampal involvement (Levine et al., 2002).

If semantic and episodic memories are stored independently, as the multiple memory trace theory suggests, then it is possible that they differ in whether they can mark memories for language assignment. Episodic memories are integrative and preserve a large amount of the encoding context across modalities. Language is inescapably part of this context, which may be reflected in language-specific encoding and retrieval of episodic memories. In contrast, semantic memory may forgo linking concepts to specific languages, forming targeted connections across neocortical sites. Overall, the structure of episodic and semantic memory opens the possibility for language-specific storage in the case of episodic memory and non-language-specific storage of semantic memory. Greater separation of languages in episodic memory can allow for easier access and processing in monolingual contexts and reduced interference from the non-target language. In a semantic memory system, language identity is determined during processing, after activating items in both languages. The degree to which patterns of lexical access and processing can reveal the structure of languages in LTM storage will be considered more carefully in subsequent sections.

**Short-term memory in bilinguals**

Information that is stored in long-term memory must be accessed and transferred to short-term memory (STM) to be processed in a meaningful way to formulate output. STM is part of the working memory (WM) system, which additionally subsumes attentional and control units involved in information processing. The structure of bilingual memory places unique demands on WM, and appears to improve the efficiency with which the system operates, improving the bilingual’s ability to maintain and encode novel information.

The first issue to consider is whether STM represents a distinct neural system compared to LTM or whether the two rely on the same underlying architecture. This issue has implications for defining the structure
Bilingual memory: structure, access, processing

and functional capacity of STM. On the basis of double dissociations between LTM and STM loss in amnesic patients, multiple-store models posit distinct neural components for the two systems (Baddeley & Warrington, 1970; Shallice & Warrington, 1970; Vallar & Papagno, 2002). The working memory model of Baddeley and colleagues further subdivides WM into functional subcomponents and is shown in Figure 1.1. An independent phonological loop and visuospatial sketchpad process auditory/linguistic and mental imagery respectively, while an episodic buffer integrates across modalities and with LTM, and a central executive controls memory manipulation and attention (Baddeley, 1986, 2010; Baddeley & Hitch, 1974).

In contrast, unitary-store models propose that STM represents the reactivation of LTM representations, concluding that both systems rely on the same underlying neural architecture (Atkinson & Shiffrin, 1971; Cowan, 1988, 2000). The critical difference between multiple- and unitary-store models is in how STM capacity is represented. In multiple-store models, there is an interaction between natural item decay and the inherent limitations of the mental rehearsal mechanism – STM capacity represents the point at which information decays faster than it can be rehearsed (Baddeley, 1986, 1992). In unitary-store models, capacity is limited by the attentional network, which determines how many
items from LTM can be reactivated at any given time (Cowan, 2000). While multiple- and unitary-store models describe functionally similar representations of STM, there is not yet consensus on its structure or how to represent its limits.

One of the first descriptions of STM capacity was Miller’s seminal paper on a capacity limit of seven items plus or minus two (1956). The effective capacity could be augmented either by facilitating rapid encoding of items from WM to LTM, or by chunking items into larger conceptual units that could be reactivated more quickly. Examples of the latter involve the ability to remember many individual letters by combining them into seven or fewer words that can be quickly rehearsed, or by remembering a longer list of words by knowing that they all start with the same letter. The ease with which rehearsal strategies can be implemented, though, may have led to overestimations of actual capacity, by conflating WM and access to LTM resources. Experimental designs that effectively restrict access to LTM yield WM estimates of between three and four items (for a review, see Cowan, 2000). Evidence from studies using event-related potentials (ERP, e.g., Vogel & Machizawa, 2004) and functional magnetic resonance imaging (fMRI, e.g., Todd & Marois, 2004, 2005) additionally find that neurological responses related to memory encoding increase up to four items and reach an asymptote, supporting the behavioral evidence that suggests an upper limit on WM capacity of around four items. More restrictive accounts of memory capacity suggest that of these four items, only one can be the focus of attention at a time (Garavan, 1998; Oberauer, 2002).

Although there appear to be hard limits on WM capacity as a consequence of STM architecture, there is evidence to suggest that bilinguals outperform monolinguals in tasks designed to assess WM ability. A recent meta-analysis investigated the effect of bilingualism on several cognitive components, including WM (Adesope et al., 2010), and found an overall advantage in WM capacity for bilinguals compared to monolinguals, with a moderate effect size. One explanation is that functional WM capacity may improve as a consequence of the unique demands of bilingualism, while the architectural WM capacity (i.e. four items with one as the focus of attention) may not be affected. As noted earlier, by employing strategies that recruit LTM, WM capacity can benefit by increasing the information density of individual items held in STM. Forming these information-chunking rules is likely to involve attentional control mechanisms to selectively activate the attributes of a set of items that allow them to be grouped as a conceptual unit, while reducing the salience of their differences (Engle 2002; Kane et al., 2001). It has been shown that bilingualism can reduce age-related decline in
the ability to acquire novel task-related rules that are arbitrarily defined (Bialystok et al., 2004). This pattern of selective attention and executive control is closely linked to WM ability (Rosen & Engle, 1998), and has been shown to be associated with lifelong bilingualism in other domains, in terms of ability to inhibit irrelevant information (Bialystok, Craik, & Ryan, 2006; Bialystok, Craik, & Luk, 2008; Costa, Hernández, & Sebastián-Gallés, 2008) and task-switching capabilities (Bialystok, Craik, & Ruocco, 2006; Prior & MacWhinney, 2010).

Overall, the evidence suggests that the architecture of STM itself is inherently limited, but that functional capacity can be improved by recruiting LTM. By rapidly encoding novel items, STM space can be freed up, whereas by retrieving organizational schemas from LTM, multiple items can be chunked into a single unit of attention. The efficiency with which these processes operate will determine individual WM capabilities. Bilinguals demonstrate increased functional WM capacity, and their improved executive control is a likely contributor to this increased capacity by facilitating the communication between STM and LTM.

**Phonological working memory in bilinguals**

Phonological working memory controls the maintenance of auditory information in an articulatory loop (Baddeley, Gathercole, & Papagno, 1998). It is commonly tested by auditorily presenting a list of digits or non-words, which the participant reproduces after a brief delay. The ability to perceive and maintain novel phonological forms is a necessary step before information can be encoded and stored in LTM.

Typical implementations of phonological memory tests, however, result in cross-linguistic differences in observed WM span (Nell, 2000), where the length of individual items (i.e., digit names) can constrain WM capacity (Ellis, 1992). English digit names are shorter than Spanish and can be rehearsed faster; accordingly, English monolinguals typically show a higher digit span capacity as compared to Spanish monolinguals. However, English–Spanish bilinguals resemble neither monolingual group, and show lower English digit spans than English monolinguals and higher Spanish digit spans than Spanish monolinguals (Ardila et al., 2000). In terms of storage and processing, then, the bilingual does not represent two independent monolingual language systems, but instead reflects a processing compromise. The bilingual's stored representations in each language may differ qualitatively from monolinguals, or the way in which information is processed may differ due to the demands of managing interference between languages.
One way to address problems of cross-linguistic differences while studying bilingual WM is to investigate the effect of increased bilingual experience on WM. It has been shown that early bilinguals outperform bilinguals who acquired a second language later in life on WM tasks (Ardila et al., 2000), and highly proficient bilinguals show WM advantages as compared to less proficient bilinguals (Bajo, Padilla, & Padilla, 2000; Majerus et al., 2008; Service et al., 2002). What’s clear from these studies is that there is a positive relationship between bilingualism and WM efficiency, with additional research needed to better understand the nature of this relationship.

**Bilingual memory access**

The manner in which linguistic information is retrieved from LTM can help to inform the structure of memory storage, in particular, the degree to which storage is language specific or not language specific. Explicit knowledge is tightly connected to language, in the form of word–concept associations in semantic memory and experiences in episodic memory associated with particular language contexts. In this section, we discuss how episodic memory may be stored and accessed according to language-specific mechanisms that preserve the encoding context. We then review research on semantic memory access in bilinguals, starting with lexical access during comprehension and following with lexical access during production.

**Episodic memory access**

Episodic memories contain representations of previous events, places, and times; when these events are personally relevant, they become part of autobiographical memory. Episodic memories are experienced as vivid multisensory events (Mather et al., 2006), and as such are subjectively distinct from the recall of semantic facts about past experiences. Episodic memories become less accessible with age, such that older adults rely more on semantic facts to describe past events (Levine et al., 2002). Evidence suggests that age-related episodic memory decline can be attenuated by bilingual experience (Schroeder & Marian, 2010). As episodic memory retrieval depends in part on intact central executive functions (Baudouin et al., 2009; Troyer, Graves, & Cullum, 1994), and bilingualism has been shown to protect against decline in executive control with age (Bialystok et al., 2004), it is possible that the improved retrieval observed in bilinguals can be attributed to preserved executive function that facilitates episodic memory access.
Additionally, bilinguals may be able to take advantage of the linguistic context associated with an event to cue retrieval. Encoding specificity refers to the phenomenon of improved memory recall when contextual cues present at encoding are also present at recall (Tulving & Thomson, 1973), and is supported by fMRI studies that suggest reactivation of an event’s brain state at encoding during retrieval (Buckner & Wheeler, 2001; Danker & Anderson, 2010). Bilinguals experience life events in one of two languages; this distinction provides an additional cue that may affect episodic memory recall. In particular, the language in which autobiographical memories are cued may influence the types of memories that become available for recall. Experimental studies and evidence from psychotherapy sessions indicate that memories that share a linguistic encoding context with the retrieval context are more numerous, more detailed, and more emotional than memories from an incongruent linguistic context of encoding (Bugelski, 1977; Larsen et al., 2002; Marian & Neisser, 2000; Matsumoto & Stanny, 2006; Schrauf, 2003; Schrauf & Rubin, 1998, 2000). This pattern suggests that in contrast to language-independent access to semantic memory, episodic memories retain language-dependent information that affects retrieval.

One of the early studies investigating the effect of retrieval context on bilingual memory recall examined older Spanish–English bilingual adults who reported not speaking Spanish for the previous ten years (Bugelski, 1977). Memories were elicited with a cued-recall paradigm in which participants were asked to provide their first memory in response to a cue. Memories were elicited separately to English cues and their Spanish translation equivalents, and participants were asked to broadly designate the period of their life that the memory occurred in. Memories elicited to English words were more likely to have occurred later in life, while memories to Spanish words were more likely to have occurred in childhood. This study was one of the first to indicate that the linguistic context at retrieval influences the type of memory preferentially recalled.

As the field developed, studies on autobiographical memory in bilinguals began to take advantage of the unique situation of sequential bilingual immigrants. These are individuals who grew up monolingual, but in adulthood immigrated to a country that required them to adopt a new language. This allows one to manipulate the linguistic context in which autobiographical memories are recalled to examine whether episodes from a congruent encoding context are facilitated. There are multiple factors that contribute to a linguistic context, and they appear to have different effects on retrieval. In particular, the ambient language of the immediate environment and the language of the cue word
independent of environment both influence recall (Marian & Neisser, 2000). Of these two, the language of the surrounding environment appears to exert a more powerful effect. Autobiographical memory access is effortful and typically involves a mental search through associations starting with the cue until a target memory is retrieved. The initial cue may facilitate search by biasing the language context of the first association, while the ambient language context may have a larger effect by influencing global search parameters regardless of the initial step facilitated by the word cue.

The process by which bilinguals arrive at a target memory during a mental search helps to illuminate how memory is organized. By using an autobiographical memory protocol analysis in which participants “think out loud,” this mental search can be made explicit (Ericsson & Simon, 1993; Schrauf, 2003). One of the strategies bilinguals are able to use during a mental search is to spontaneously translate the cue word (Schrauf, 2003). Bilinguals’ ability to use this strategy suggests that when a story could not be brought to mind, participants were aware that a word’s translation equivalent might offer new mental associations. If memories are organized according to the linguistic context they were encoded in, then translating becomes a viable strategy during memory search. Another way this inner mental translation can be observed is by asking participants to report in what language a memory first came to them, before they formed a response appropriate to the retrieval context. Participants sometimes report a memory first coming to them in a language different from the language of the cue. In these cases, a language switch occurred during mental search when one language did not afford a suitable personal memory. In accordance with a language-specific encoding hypothesis, autobiographical memories that had first come to participants in their native tongue occurred at an earlier age than memories in their second language (Matsumoto & Stanny, 2006; Schrauf & Rubin, 1998, 2000).

These studies, however, necessarily confound sequential bilingualism with biculturalism. One of the ways this problem has been addressed is to examine two sets of bilingual immigrants with the same native and host countries who differ only in their age of immigration (Larsen et al., 2002). Two groups of Polish–Danish bilinguals were compared who had immigrated to Denmark at a mean age of either 24 or 34. It was found that memories that came to participants in Polish were more likely to have occurred prior to immigration, while memories in Danish were more likely to have occurred after immigration. This shift was qualitatively similar between groups, but centered around the age of immigration. Though groups differed in their cultural identities,
reflected by use of inner speech in their first and second languages, in both groups, the shift in language context drove autobiographical memory retrieval.

In summary, episodic memory in bilinguals retains language-specific information that can influence the ease with which memories are accessed. Bilingualism can protect against age-related episodic memory decline, potentially by providing a cue in the linguistic context of encoding that can facilitate retrieval in a matching linguistic context. Results across multiple studies on different bilingual populations indicate that access to autobiographical memories is enhanced depending on the overlap between the linguistic context in which access is attempted and the language that was in use at the time the memory was encoded. Compared to semantic memory access, episodic memories appear to integrate language more closely and preserve the language of encoding in the memory trace.

**Semantic memory access during language comprehension**

The idea that bilinguals are able to shift into a monolingual language mode and restrict memory access to words in a single language is appealing (Grosjean 1985, 2001), since bilinguals are able to functionally separate their two languages and converse with monolingual speakers. However, a strict language mode hypothesis where the speaker is able to restrict access to a single language is inconsistent with evidence for activation of a non-target language that results in competitive interference, even in a monolingual context (Dijkstra & van Hell, 2003; Duyck et al., 2007; Marian & Spivey, 2003a, b).

One strategy that has been used to investigate lexical access in bilinguals utilizes cross-linguistic orthographic similarity. Upon viewing a printed word, its lexical representation in LTM is accessed, but orthographic neighbors (words that differ from the target by only one letter) also become activated and can influence processing (van Heuven, Dijkstra, & Grainger, 1998). Evidence suggests that bilinguals access orthographic neighbors both in the target and in the non-target language (van Heuven et al., 2008). This suggests that lexical access is not language specific during word recognition, as bilinguals automatically retrieve items from both languages. Research on cognates provides additional evidence suggesting that information in the non-target language is accessed. In lexical decision tasks, cognates are typically identified as words more quickly than non-cognates, since word representations in both languages are accessed and facilitate lexical decision. The magnitude of this effect is dependent upon the degree to which the
orthographic representations of the cognates overlap (Dijkstra et al., 2010). Even though the task only requires participants to access words in a single language, cognate processing indicates that bilinguals activate representations in both languages.

While orthographic overlap across languages allows one to test lexical access during a monolingual task, cognates may overtly activate both languages. A strong test of non-language-specific access would restrict overt activation of the non-target language, and is possible with minimal cross-linguistic phonological overlap. In the visual world paradigm, participants view a display of several objects and are instructed to manipulate objects while their eye movements are tracked. If a participant is instructed, for example, to “pick up the candy,” participants make more eye movements to objects that overlap phonologically with the target, such as a candle, compared to non-overlapping control objects. This suggests that “candle” was partially activated due to its similarity to the target word “candy” (Allopenna, 1998; Tanenhaus et al., 2000).

This paradigm has been successfully adapted to bilingual language processing to investigate whether words in a non-target language are co-activated. It was found that when Russian–English bilinguals were instructed in Russian to pick up a marka (the Russian word for stamp), they were significantly more likely to look at a marker than at a phonologically unrelated control (Blumenfeld & Marian, 2007; Marian & Spivey, 2003a, b; Spivey & Marian, 1999). The English names of competitor objects were never spoken, yet bilinguals accessed the lexical representations as a result of hearing words in the other language. This pattern of results provides strong support for non-language-specific lexical retrieval. It has also been shown that sentence context can reduce the magnitude of between-language competition (Chambers & Cooke, 2009).

Lexical access in these studies is thought to occur due to phonological items activating word representations in both languages non-selectively. Cross-linguistic lexical access and interference can be demonstrated in the absence of input overlap, indicating interactive processing links that facilitate lexical retrieval. In a series of studies by Thierry and Wu (2004, 2007), cross-linguistic lexical access was demonstrated in English–Chinese bilinguals. Chinese is an ideographic language that lacks orthographic form overlap with English. English–Chinese bilinguals were asked to read English word pairs and make a semantic relatedness judgment. Unbeknownst to the participants, the English words’ translation equivalents in Chinese were each composed of two characters, and in half of the experimental trials, a character was
repeated within a word pair. Behaviorally, there was no effect of character repetition; but when electrophysiological responses were examined, it was observed that Chinese repetition pairs elicited a larger N400 cortical response than control word pairs. Despite input in only one language, non-language-specific lexical access was observed, suggesting that retrieval is automatic and can cross language boundaries. Further evidence was observed in a recent study using the visual world paradigm with bimodal bilinguals (hearing users of a spoken and a signed language, such as English and American Sign Language). Although their two languages lack form overlap, co-activation was observed across languages, reflecting non-language-specific access (Shook & Marian, 2010, submitted).

In summary, during language comprehension, bilinguals appear to access linguistic representations in semantic memory in a non-language-specific manner. Access is not modality specific, but occurs in both orthographic and phonological contexts. The effect of non-target language activation appears to vary with task constructs, as lexical decision is facilitated by cross-linguistic activation, but target processing is interfered with as a consequence of phonological competition across languages.

**Semantic memory access during language production**

Language production occurs under a different set of demands compared to language recognition, and these differences are likely to influence the manner in which information is accessed. During recognition, linguistic input is initially ambiguous and is integrated as it unfolds. A system that considers lexical alternatives in either language is thus able to easily adapt to changing language contexts. During production, however, language switches are determined by the speaker, and this prior knowledge may help to constrain lexical selection and minimize interference between languages. Speech production requires two steps, an initial activation of potential lexical items to be produced and a selection process that determines the item to be articulated. During the first step of lexical access, words in both languages are thought to be accessed in parallel (Colomé, 2001; Colomé & Miozzo, 2010; Costa, Caramazza, 2001).

1 The N400 is a component of the ERP signal (event-related potentials are used to assess neural responses by recording electrical activity on the scalp). The N400 occurs roughly 400 ms after stimulus onset and is characterized by a negative deflection in the waveform. It is commonly thought to be a marker of semantic processing. A larger response indicates increased difficulty, in this example caused by the covert Chinese form overlap that interfered with semantic processing.
& Sebastián-Gallés, 2000; Gollan, Montoya, Fennema-Notestine, & Morris, 2005). However, bilingual production models disagree on whether the subsequent step of lexical selection is language specific, precluding competition (Costa, La Heij, & Navarrete, 2006), or not language specific, allowing for between-language competition (Kroll, Bobb, & Wodniecka, 2006). Research from studies on cognate naming, cross-modal processing, and cross-linguistic interaction indicates that while multiple languages can interact in the bilingual mind during language production, this interaction does not result in competitive interference, as in bilingual language recognition.

As words that share cross-linguistic form and meaning overlap, cognates explicitly activate both of a bilingual’s languages. If similar but non-identical cognates compete for selection, they should be difficult to produce. Instead, research suggests that bilinguals name cognates more quickly than non-cognates (Costa et al., 2000). This suggests that rather than competing for selection, activation of a word in a non-target language can activate phonemes shared with the target word, facilitating production. Critically, in order for facilitation to be observed at all, initial lexical access must have been non-language-specific, allowing for the target word’s translation equivalent to boost articulatory processes. Evidence from bimodal bilinguals indicates that in the absence of the biological constraints on word production in unimodal bilinguals, the language system is able to co-articulate translation equivalents, using concurrent manual signs with spoken speech (Casey & Emmorey, 2008; Emmorey et al., 2008). This example demonstrates again non-language-specific lexical access during bilingual speech production.

Cross-language facilitation has also been observed during a target/distractor naming task (Colomé & Miozzo 2010). Bilingual participants viewed two superimposed red and green drawings, and were asked to name pictures of a certain color. When the name of the distractor in the non-target language overlapped phonologically with the target, target naming latency actually decreased. Even though the non-target language was never elicited and the distractor picture was clearly identified, participants appeared to spontaneously access word labels and phonological representations for the distractor in both languages, causing constructive feedback similar to that found by Costa et al. (2000) that facilitated target production.

By only examining instances of successful language retrieval, though, potential interference across languages may be overlooked. Words are not always accessed successfully during speech production, and these other instances of retrieval failure help to inform the
interaction between lexical activation and selection. Retrieval failures commonly occur in tip-of-the-tongue (TOT) experiences, which entail an inability to retrieve a word despite a speaker’s certainty that it exists in memory. The phenomenon has been well documented in monolingual speakers, particularly in cases where individuals are attempting to retrieve low-frequency words (Brown, 1991; Schwartz, 1999). During a TOT state, fragments of the target word may be accessible, including knowledge of the initial phonemes or syllables. Some TOT states are spontaneously resolved, while others can be resolved after being provided with phonological primes as cues (Abrams, White, & Eitel, 2003; James & Burke, 2000).

On the one hand, TOT states may reflect a breadth of lexical knowledge – a difficulty in accessing words may produce TOTs in individuals with large vocabularies, whereas those with lower vocabularies may simply not know the target word. On the other hand, TOTs may indicate specific failures to retrieve and a problem with lexical access. TOTs of the first sort are typically observed in older adults, and only to low-frequency words (Dahlgren, 1998; Gollan & Brown, 2006; Schwartz, 2002). Bilinguals instead show TOTs of the second sort, to words of all frequencies (Gollan & Acenas, 2004; Gollan & Brown, 2006; Gollan, Montoya, & Bonanni, 2005; Gollan & Silverberg, 2001). While superficially similar, then, TOTs in older adults are a reflection of larger vocabularies leading to more cases of attempted lexical access, whereas TOTs in bilinguals reflect a systematic disruption of lexical access.

As a result of dividing their time between two languages, bilinguals necessarily use each language less than monolingual speakers, who devote all of their communication to a single language. Frequency is an important predictor of TOTs; as seen in older adults, low-frequency words cause more TOTs than high-frequency words. Since the absolute frequency of individual words in the bilingual lexicon is lower than in monolinguals, individual items may be more difficult to retrieve. Lexical access failures in bilinguals do not appear to be driven by interference between languages, as bimodal bilinguals experience TOTs at a rate comparable to unimodal bilinguals (Pyers, Gollan, & Emmorey, 2009). The bimodal bilingual’s languages cannot interfere phonologically, but it is reasonable to expect that lower absolute word frequencies drive TOTs in a manner similar to that in unimodal bilinguals.

Additional support for the word frequency account in bilinguals comes from research on TOT incidence for retrieval of proper names. An individual’s proper name is consistent in different language contexts, and thus should be matched in usage frequencies across monolinguals and
bilinguals. It has been shown that, in fact, bilinguals experience TOTs to proper names (either famous names or personally relevant names) at a comparable rate to monolinguals (Gollan, Montoya, & Bonanni, 2005). These results suggest that TOT incidence is sensitive to word frequency, and when the sum of a word’s frequency across multiple language contexts approaches that word’s frequency in a monolingual context, lexical access is comparable.

Just as proper names overlap between different language contexts and act as high-frequency words, cognates may act as high-frequency words compared to non-cognates due to form overlap that can facilitate parallel activation. Even if cognates are represented distinctly in each language, their phonological similarity may cause both items to activate in monolingual contexts. Cognates have been found to facilitate recall and reduce TOTs to the level of monolingual speakers, but only for cognates that a bilingual knew in both languages, and thus could conceivably use as often as a monolingual would (Gollan & Acenas, 2004). Even when only non-cognates were considered, words that more bilinguals were able to translate easily had reduced TOTs as compared to words without translation equivalents.

The sum of these findings suggests that translation equivalents do not impair lexical access during production. Instead of competing, words in the non-target language actually facilitate retrieval. This is consistent with models of parallel activation of lexical items in both languages during production, which can serve to increase the relative frequency of names, cognates, and translation equivalents as compared to words that a bilingual only knows in one language. The relative frequency of words within an individual bilingual lexicon determines their accessibility and may be sensitive to the strength with which those items are encoded in memory. The other finding to come from studies on TOT retrieval failures is that as a consequence of dividing their time between two languages, items in the bilingual lexicon can act as lower frequency words as compared to their equivalents in the monolingual lexicon, leading to group differences in retrieval efficiency. Thus, bilingualism results in non-language-specific access to words in both languages during production, but this parallel access does not interfere with lexical selection. Instead, difficulties in bilingual lexical access during production seem to be caused by knowing more words than monolinguals and using them less frequently.

In summary, the degree to which lexical access is language specific versus not language specific depends on the type of information stored in memory. As a consequence of experiencing life in two languages, bilinguals encode information about the world in different languages
Bilingual memory: structure, access, processing

according to varying linguistic contexts. These memories can be accessed in either language, resulting in situations where the language of encoding and the language of retrieval are either congruent or incongruent. The degree to which congruent language contexts affect memory access informs the specificity of LTM for individual languages. Evidence from autobiographical memory recall in bilinguals shows that congruence between encoding language and retrieval language affects retrieval, suggesting that episodic memories preserve language classification in the memory trace. Semantic memory instead shows a pattern whereby lexical items in both languages become activated and can influence language comprehension and production. During both comprehension and production, non-language-specific activation appears to be automatic, supporting the idea that information in semantic memory is accessed without regard to language classification.

**Information processing in bilinguals**

Although a bilingual’s two languages co-activate to a certain degree during language processing, the two languages remain functionally distinct and can be used in different contexts. This separation between languages may emerge either from architectural constraints in the language system, such as separate lexicons, or it may result from an external control mechanism that manipulates global language activation. Several mechanisms for differentiating languages have been implemented in theoretical and computational models of bilingual language processing, and in this section we review the successes and limitations of a number of these models. In addition, we consider how novel language learning can be used to understand the emergence of functionally distinct languages. Whereas modeling work is vital in determining how proficient bilinguals may control access to their two languages, the process of novel language learning provides insight into how a functional separation between languages can develop. Further, by comparing novel language learning success in monolinguals and bilinguals, it is possible to determine how existing mechanisms for controlling language processing in bilinguals may extend to novel language learning success.

**Modeling bilingual processing**

Models of bilingual language processing have to account for the fact that although a bilingual is able to communicate effectively in a single language, both languages remain active and can potentially interfere at
lower levels of lexical processing. Current models approach this problem in different ways, depending on the domain that they are specialized for. Comparing the strengths and limitations of these separate models highlights some of the primary issues that must be considered in a theory of bilingual memory and language processing, as well as potential areas of further research. In this section, we briefly review some of the more influential models of bilingual memory and language processing, and consider how these models describe language interaction during lexical processing (for a more thorough review of selected models, see Dijkstra, Haga, Bijsterveld, & Sprinkhuizen-Kuyper, this volume).

We start with the Revised Hierarchical Model (RHM; Kroll & Stewart, 1994), an early descriptive model of bilingualism that focused on associations between languages and between lexical and conceptual representations, and prompted research in the field of bilingual language organization and development. The RHM laid the groundwork for a number of specialized models, including the Bilingual Interactive Activation model, which focuses on visual word processing and the structure of the bilingual lexicon (Dijkstra et al., 1998; Dijkstra & van Heuven 2002; van Heuven et al., 1998), and the Bilingual Language Interaction Network for Comprehension of Speech, which focuses on auditory word processing and patterns of cross-linguistic activation (Shook & Marian, in review). Language production and control are described in the Inhibitory Control model (Green, 2003), and the emergence of distinct languages during bilingual language acquisition has been depicted in the Self-Organizing Model of Bilingual Processing (Li & Farkas, 2002). The sample of models reviewed in this section provides a useful framework for considering the organization of language in bilingual memory, and the degree to which multiple languages interact during processing.

The Revised Hierarchical Model The Revised Hierarchical Model (RHM; Kroll & Stewart, 1994) includes separate lexicons for a bilingual’s two languages, but also contains direct associative links between translation equivalents (see Figure 1.2). The RHM built on the work of Potter and colleagues (Potter et al., 1984), which separated the associative links between a bilingual’s two languages and the conceptual links between words and the concepts they represent. The RHM also separates lexical and conceptual stores, with a shared conceptual store across languages that contains associations to lexical items in both languages. By varying the strength of these different connections, either between lexical items or from the individual lexicons to the conceptual store, organization of the language system can be
modeled across development in simultaneous or sequential bilinguals. In a beginning second language learner, words in the second language (L2) are most strongly associated with their translation equivalents in the native language (L1), and contain only weak connections to the shared conceptual store. Accessing meaning for a word in the L2 is thus accomplished by activating the L1 equivalent, which then accesses semantic meaning. As L2 proficiency increases, lexical items in the L2 become more strongly associated with the conceptual store, reducing the need for L1-mediated semantic access.

The RHM plays an important role in how thinking about bilingual memory has developed, as it posited a dynamically refined system that allows for both language interaction and functional independence. However, the structure of the RHM has been challenged by some recently emerging patterns in bilingual language processing, and may be less of a good fit as compared to other language models when describing interactivity within the bilingual system (Brysbaert & Duyck, 2010). In particular, the separation of L1 and L2 into separate lexicons is inconsistent with data demonstrating effects on processing of orthographic neighbors in a non-target language (van Heuven et al., 1998), reflecting cross-linguistic word form effects that are difficult to explain using separate lexicons. Additionally, predictions of the RHM regarding asymmetric priming from the L2 to the L1 due to strong direct associative links between translation equivalents are not well supported (Schoonbaert et al., 2009). The RHM has been successful in its ability to describe changes to bilingual language storage as proficiency
changes, but the basic structure of segregated language storage in the model may need further development.

The Bilingual Interactive Activation Model The Bilingual Interactive Activation (BIA) Model and its revised form, the BIA+ (Dijkstra et al., 1998; Dijkstra & van Heuven 2002; van Heuven et al., 1998), are models of bilingual visual word recognition and have been implemented computationally. The BIA and BIA+ models differ from the RHM in that they include integrated storage of the bilingual lexicon (see Figures 1.3 & 1.4). The presence of language input (orthographic only in BIA, and extending to phonological information in BIA+) non-selectively activates items that overlap with the input in either language. Non-selective language access is consistent with research showing that cognates and interlingual homographs interfere with target processing, even in a monolingual context. In the BIA+ model, to the degree that an item overlaps with the input, it will become activated regardless of its language. At higher processing levels in which a response is selected, language assignment can be enforced. The BIA+ model implements language assignment by tagging individual words in the shared lexical storage to distinct language nodes; as words in one language are activated, items with the same language tag increase in activation, while items with the opposite language tag are reactively inhibited. Task-oriented language modes can selectively activate or inhibit language tags at the response level, and in this manner, the BIA+ model relies less on the organization of language in LTM to separate languages, and instead shapes language context at the processing level. The BIA+ model has provided a useful framework for findings on bilingual semantic memory access during language comprehension. As discussed earlier in this chapter, automatic activation of both languages in response to single-language input has been supported by numerous studies, and the BIA+ model will continue to be an important resource for understanding bilingual language processing.

The Bilingual Language Interaction Network for Comprehension of Speech The Bilingual Language Interaction Network for Comprehension of Speech (BLINCS; Shook & Marian, in review) models language interaction during bilingual spoken language comprehension, in contrast to the BIA’s focus on visual input. BLINCS represents a bilingual functional architecture in which the acoustic signal perceived by bilinguals travels to a feature level, then to a phonemic level, then from there to the lexical level, and further to the semantic level (see Figure 1.5). The interaction between levels is bi-directional,
Figure 1.3 Van Heuven, Dijkstra, and Grainger's Bilingual Interactive Activation Model (BIA)
allowing for both feed-forward and back-propagation. Within levels, language-specific and language-shared representations are included, with bi-directional connections between languages allowing for competition within and across languages. Each processing level includes a self-organizing map (see Li & Farkas, 2002 for a review) that organizes according to the amount and type of dual-language input, such that structurally similar items are grouped together. Connections between levels are bi-directional and strengthen according to the co-occurrence...
of items. The model quantifies associations between items based on structural similarity regardless of language classification, as similar items occupy adjacent regions in the self-organizing maps.

The Inhibitory Control Model  Although it makes few claims about the organization of multiple languages in memory, the Inhibitory Control (IC) Model (Green, 2003) plays an important role in describing the pattern through which language selection occurs when language storage and access is not language specific (see Figure 1.6). The IC Model includes abstract language schemas, separate from the lexical-semantic system, that compete to determine relative language activations. These language task schemas are themselves moderated by a supervisory attentional system (SAS) that regulates their activity. By inhibiting the task schema for the non-target language, communication
in a monolingual context can occur by minimizing intrusions from the other language. In order to switch languages, then, and reactivate a previously inhibited language, latent inhibition must be overcome and will be associated with a processing cost (Meuter & Allport, 1999). As discussed earlier in this chapter, bilingual language experience appears to be associated with improved inhibitory control ability, leading to gains in attention and working memory that extend to novel language learning skill (Bajo, Padilla, & Padilla, 2000; Kaushanskaya & Marian, 2009b; Majerus et al., 2008; Service et al., 2002). However, while it seems clear that the Inhibitory Control Model accurately describes a method by which language activation can be adjusted, the efficiency with which the inhibitory demands are managed may vary between different types of bilinguals or multilinguals (Costa & Santesteban, 2004), and thus in its present form may not be broadly applicable to general bilingual processing.

The Self-Organizing Model of Bilingual Processing

The Self-Organizing Model of Bilingual Processing (SOMBIP; Li & Farkas, 2002) was designed to test the ability of the language system to self-organize over time in a manner that accommodates dual-language
input. It was developed based on connectionist models of language storage in monolinguals (Elman, 1990) and bilinguals (French, 1998) that self-organize according to statistical regularities in the input. SOMBIP has been tested on a mixed-language naturalistic input, and has been shown to successfully separate languages and store representations within a shared space. The self-organizing models in SOMBIP initially start by randomly activating patterns of nodes distributed in a two-dimensional space. Activated units and their neighbors adjust to activate more strongly to that same input in the future. The effect over time is an increased selectivity of the model’s response to inputs, and the emergence of functionally distinct regions that respond to specialized types of information. SOMBIP contains two mutually interconnected self-organizing models, one that activates to phonological input and one that activates to semantic input (see Figure 1.7). The architecture of the model allows for translation equivalents to become closely associated at the phonemic level due to their similar semantic representations, and for distinct semantic concepts to become associated due to phonologically similar interlingual homographs. In contrast to other models (see de Groot & Kroll, 1997), semantic concepts in SOMBIP are not shared between languages. Instead, semantic concepts in each language are represented in a shared space within a self-organizing model for semantic information. The strength of SOMBIP is its ability to functionally separate languages within a shared storage space according to co-occurrences in the input, providing a plausible mechanism by which an integrated lexicon can form in bilingual long-term memory. SOMBIP’s ability to capture aspects of bilingual language development
makes it a prime candidate to incorporate the growing body of research on novel language learning success in bilinguals.

Theories of bilingual language processing necessarily make assumptions about how language and memory are organized, and how memory access occurs; even if a theory makes explicit claims only regarding one aspect of language processing, implicit assumptions are made on how the rest of the system is organized. The advantage of a model, whether it is verbally described or computationally implemented, is that it forces the researcher to confront these implicit assumptions while they construct a theory of processing. It is only when assumptions are made explicit that they can be challenged and refined, as seen in the shift away from separate lexicons in the RHM toward an integrated bilingual lexicon as seen in BIA+, BLINCS, and SOMBIP. At this stage, a challenge for models of bilingual language processing arises due to their increasing specialization. It is becoming more important that models unify the terms and frameworks in which they are based, so that competing assumptions can be recognized and compared.

Novel language learning and processing

Models of bilingual language processing have increased understanding of the unique challenges that result from organizing and processing multiple languages within a single mind. Research on novel language learning can contribute to this discussion by exploring how the language system can change to accommodate an additional language. Novel language learning entails a reorganization of the language processing system to adapt to the new language’s rules of grammar and phonotactics, and to the new vocabulary. In addition to mastering the new language, the learner must be able to mitigate interference from more strongly represented proficient languages, and increasing processing demands. A bilingual’s years of experience controlling access to two languages may affect these learning and memory processes in ways that improve novel language acquisition. Further, late bilinguals (those who learned a second language after acquiring their first language, instead of learning both concurrently) may be able to profitably transfer skills developed during L2 acquisition to learning an L3. Novel language learning, then, can be used to compare the flexibility of monolinguals’ and bilinguals’ language processing abilities, and how well they are able to integrate novel vocabulary and grammatical rules. With recent advances in this area and without the confounding factors of earlier research (see Lambert, 1981), a consistent pattern has emerged where bilinguals learning an L3 outperform monolinguals
learning an L2, across domains spanning vocabulary (Cenoz, 2003; Cenoz & Valencia, 1994; Kaushanskaya & Marian, 2009a, b; Keshavarz & Astaneh, 2004; Sanz, 2000; Thomas, 1992; van Hell & Mahn, 1997), grammar (Klein, 1995; Sanz, 2000; Thomas, 1992), and pragmatics (Safont Jorda, 2003). The research methods are varied, and may be primarily observational, in which performance of participants enrolled in academic language courses is investigated, or experimental, in which training occurs within a tightly controlled laboratory environment, and the target language may be either naturally occurring or artificially constructed (e.g., Cenoz & Jessner, 2000; Sanz & Lado, 2007).

Observational studies on bilingual third-language learning success in classroom environments have made important contributions by demonstrating that novel-language vocabulary learning is influenced by previous bilingual experience. In one study by Sanz (2000), English language proficiency was assessed in Catalan–Spanish bilinguals and Spanish monolinguals (mean age 16.53 years) with the CELT English Proficiency Test (Harris & Palmer, 1970). A hierarchical multiple regression analysis showed that bilingual experience contributed to third language proficiency, and that this effect remained after controlling for the contribution of other factors including total English exposure and learning motivation. These results were in line with earlier research in a different community comprised of Spanish monolinguals and Basque–Spanish bilinguals, in which bilingual experience contributed to improved word learning, even after accounting for exposure and motivation to learn the target language (Cenoz & Valencia, 1994). Keshavarz and Astaneh (2004) compared English language knowledge of Persian monolinguals, Turkish–Persian bilinguals, and Turkish–Armenian bilinguals using a Controlled Productive Ability Test in English (Nation, 1990). All participants were enrolled in Iranian pre-university intermediate English language classes, and results indicated that both bilingual groups produced more words correctly as compared to the monolingual group, suggesting that bilingual experience improved the ability to learn and recall English vocabulary learned from classroom instruction.

The trade-off when studying language-learning performance in classroom settings is that the researcher often has minimal influence on the instructional materials and is unable to control the participants’ exposure to the target language outside of the instructional setting. By training monolinguals and bilinguals on a target language within a research setting, greater control over the learner’s environment is available, although the extent of language training is limited by the experimenter’s time and resources. Van Hell and Mahn (1997) investigated
how late bilinguals, who learned their L2 in early adolescence (“experienced foreign language learners”), compared to monolinguals (“inexperienced foreign language learners”) in terms of foreign-language vocabulary learning. Participants were trained to associate Spanish words (presented visually and auditorily) with native language translations (i.e., Dutch for bilinguals or English for monolinguals). Learning performance was tested using a backward translation task in which participants viewed orthographic Spanish words and verbally provided the learned Dutch or English translations. Bilinguals were found to be more accurate and faster to respond compared to monolinguals, suggesting either that skills developed while learning an L2 may transfer to foreign-language vocabulary learning, or that years of bilingual language experience influence the organization of the language system in ways that promote further learning.

Kaushanskaya and Marian (2009b), with a conceptually similar study, showed that advantages in foreign vocabulary learning also extended to early bilinguals who had acquired two languages concurrently in childhood. The researchers taught English monolinguals, early English–Spanish bilinguals, and early Mandarin–English bilinguals words in an artificially constructed language (words in the artificial language were controlled for similarity to English, Spanish, or Mandarin). At test, participants performed a backward translation task from the novel language to English. Both immediately following training and at a one-week follow-up, the two bilingual groups achieved higher accuracy compared to monolinguals, providing additional support for a general bilingual advantage for word learning in a novel language.

Bilingualism has been shown not only to influence vocabulary learning in a third language, but also to improve acquisition of a novel language’s grammar relative to monolingual speakers. Klein (1995) showed that bilinguals learning English made fewer preposition-stranding errors compared to monolinguals. Both groups made the same types of errors, but they occurred with lower frequency in the bilingual group, suggesting that bilinguals were mastering English syntactic use earlier than the monolingual learners, but were not approaching the constructions differently. The bilingual learning advantage also extends to more abstract sequence learning. In a series of studies, bilinguals and multilinguals were found to extract the underlying rules of an artificial grammar better than monolingual learners (Nation & McLaughlin, 1986; Nayak et al., 1990). Bilingual experience was also found to be associated with increased ability to segment novel words in a continuous auditory sequence using statistical probabilities that defined words (Bartolotti et al., 2011). Sequence learning is important for both word
Bilingual memory: structure, access, processing

learning and for acquisition of syntactic rules such as word order, and may be a significant contributor to bilingual foreign language learning ability.

These bilingual advantages in novel language learning are likely the result of multiple contributing factors. For example, bilinguals may have more opportunities to transfer knowledge from known languages to the new language. Cognates, which overlap in form and meaning across languages, are typically easier to learn (MacWhinney, 2007; Murphy, 2003); bilinguals will have encountered more cognates than monolinguals, by virtue of already knowing two languages. Cenoz (1997) examined cases of language transfer from Basque and Spanish to English in children, and found that children utilized both of their known languages, but preferentially transferred word knowledge from Spanish to English, regardless of whether their native language was Basque or Spanish. English is more typologically related to Spanish than to Basque, thus the children’s preference suggests that bilinguals are sensitive to which of their two known languages provides a better scaffold for L3 learning.

Other contributing factors depend not on the specific languages learned, but on the unique demands bilingualism itself places on the cognitive architecture. For example, bilingualism may increase phonological working memory ability, allowing bilinguals to sustain novel words in WM until they can be encoded in LTM (Papagno & Vallar, 1995; van Hell & Mahn, 1997). Bilingualism may also increase metalinguistic knowledge for how language operates as a system (Jessner, 1999, 2008). As a result, bilinguals rapidly acquire the understanding that the relationship between words and concepts is arbitrary; this knowledge is essential in order to associate known concepts with new words in a foreign language. Finally, bilingual experience may minimize cross-linguistic interference during novel language learning. It can be difficult to suppress aspects of a known language, such as letter-to-phoneme mappings, when they conflict with a novel language, but bilinguals are better able to minimize this conflict compared to monolinguals during novel language learning (Kaushanskaya & Marian, 2009b). In fact, it appears that advantages in resolving interference persist beyond the learning process itself and extend to novel language use. In a recent study by Bartolotti and Marian (2010, in review), monolinguals and bilinguals were taught an artificial language designed to elicit between-language competition. Specifically, words in the novel language (e.g., shundo, meaning acorn) overlapped phonologically with English competitor words (e.g., shovel). Activation of these native-language competitors was assessed with eye-tracking
and mouse-tracking. It was found that bilinguals processed the novel language more efficiently and managed interference from the native language more successfully than monolinguals.

As bilinguals learn to use and process a novel language, it is likely that they rely on a combination of the skills discussed above, including linguistic transfer, phonological working memory, metalinguistic knowledge, and inhibitory control. Studying the role that these factors play during language learning in bilinguals can provide valuable insights into the relationship between memory and language. In particular, language learning offers a glimpse into how interaction across languages occurs and how a functional distinction between languages can develop.

Conclusions

This chapter reviews the representation of and interaction between multiple languages in bilingual memory. A central question in bilingual memory research is at what level a bilingual’s two languages are differentiated, and whether information is stored, accessed, and processed in a language-specific or non-language-specific manner. This question has implications not only for bilingual research, but also for language processing in general. The bilingual’s two languages provide contrasting contexts through which the links between language and memory can be investigated. It appears that episodic memory for events, especially when it is autobiographical, retains language-specific information. Episodic memories contain vivid details relating to the encoding context of the event. A bilingual’s two languages contribute to this encoding context, facilitating memory recall in cases where the encoding language and the retrieval language overlap. Semantic memory for facts and word–concept associations, on the other hand, appears to be stored independently of language, as evidenced by parallel access to both languages during recognition or production. In addition, modeling work has shown that patterns of bilingual language processing can be captured in systems that integrate words across a bilingual’s two languages into a single lexicon, providing a plausible account of non-language-specific storage in semantic memory. Thus, it appears that memories can be stored without being explicitly tied to individual languages, but that this link can be utilized to represent the encoding context of episodic memories.

In conclusion, a bilingual system has to accommodate the storage of multiple languages, as well as the processing demands resulting from parallel language activation. The experience of acquiring and using
Bilingual memory: structure, access, processing

multiple languages improves bilinguals’ ability to encode novel information relative to monolinguals, while practice managing interlingual competition improves working memory processing and inhibitory control. These cognitive benefits result from experience managing the conflict that occurs when representations in multiple languages are activated, but only one must be selected as the output. Much of the research on bilingual memory investigates what happens when this output selection mechanism fails, as these cases allow rare glimpses into the structure, access, and processing of memory. What is remarkable is that outside of the laboratory, countless bilinguals routinely perform these cognitive feats, speaking and switching languages with ease while rarely betraying the complex processes occurring unseen within the mind.

References


Bartolotti & Marian


Bilingual memory: structure, access, processing


