7 Laboratory Designs and Paradigms: Words, Sounds, and Sentences

Judith F. Kroll, Chip Gerfen, and Paola E. Dussias

7.1 Introduction

In the past 10 to 15 years there has been a notable increase in the number of studies that take a psycholinguistic or cognitive approach to bilingualism. Although there is a long history of research on issues such as whether there is a critical period[1] for second language acquisition (see Birdsong, 1999a, for a review), it is only recently that cognitive scientists have begun to see that bilingualism is the norm for most of the world's population. Because cognitive science seeks to identify universal properties of thought, the bilingual has become a model subject of study rather than a marked case. Researchers have come to see that studies of bilingual cognition provide critical evidence regarding the principles that constrain or permit interactions across cognitive systems. At the same time, the development of a set of laboratory tools to investigate language performance and its neurocognitive basis has enabled a new experimental approach to bilingualism that is informed by studies of cognitive processing and brain function, in addition to the linguistic approaches that have traditionally characterized bilingual research.

What do bilinguals tell us about cognition? And what can cognitive approaches tell us about bilingualism? In the sections that follow we introduce readers to some of the laboratory designs and paradigms that are commonly used in experimental studies of bilingualism. The methods that we describe have been used to ask how a bilingual manages the presence of two languages in a single mind. If the two languages were entirely independent of one another, then the question might not be as pressing. However, as we will see in the discussion that follows, there is a great deal of evidence that suggests that the bilingual is not two monolinguals in one

1 The term 'critical period' refers to a time in early childhood, typically assumed to be prior to the onset of puberty, when individuals are hypothesized to be sensitive to the input of the languages to which they are exposed in a manner that allows native-like acquisition. Although there is agreement that early exposure results in superior language acquisition, there is little agreement about its basis.

(Grosjean, 1989). Instead, the recent experimental research demonstrates that the bilingual's two languages interact closely. These interactions influence the way in which bilinguals understand words spoken and read in each language, how speech is produced with or without a foreign accent, and how sentences are comprehended and produced when the grammar of the two languages is similar in some ways and distinct in others. What is remarkable is that the observed interactions are not restricted to the second language, but affect the native language as well. The methods that have been developed to examine language processing in bilinguals have been used to explore the scope of these interactions and the constraints that are imposed by the structure of the specific languages themselves. Although it might seem that a language system that has to cope with two sets of competing alternatives might suffer in some respects, the recent evidence suggests that bilingualism confers benefits to cognition by virtue of forcing the development of cognitive skills to negotiate the activity of the two languages (e.g., Bialystok, 2005) and that there may even be structural consequences for brain organization (e.g., Mechelli, Crinion, Noppeney, et al., 2004).

In this chapter we review three major areas of research activity in experimental psycholinguistics. Section 7.2 examines the way in which bilinguals recognize words when they are spoken or read in each language and when they produce words in the language in which they intend to speak. Section 7.3 addresses speech to ask how the sounds of each of the bilingual's two languages are processed when they are heard or spoken. Section 7.4 examines sentences to ask how the grammatical structures and preferences associated with each of the bilingual's languages are affected by the presence of both languages. Within each of these topics our review will necessarily be brief, but we hope to illustrate the logic of the experimental approach in a way that will provide a useful guide to the primary literature.

7.2 Words

When a bilingual hears, reads, or speaks a word in one of his or her two languages, is the other language also active? A great many studies of visual and auditory word recognition have investigated the question of whether the bilingual lexicon is an integrated representation across the word forms of both languages or whether words in each language access independent representations, one for each language. A full discussion of the theoretical alternatives associated with this debate is beyond the scope of the present chapter, but there are a number of summaries of this work available in recent articles and chapters (e.g., Dijkstra & Van Heuven, 2002; Kroll & Dussias, 2004; Costa, 2005). In brief, most of the evidence suggests that lexical access is non-selective, in that alternatives in both languages appear to be activated in parallel when words are processed in one language alone.

How can we draw the conclusion that lexical access occurs in parallel across the bilingual's two languages? Here we describe three paradigms that will serve to illustrate the logic of research on the bilingual lexicon: (1) visual lexical decision,

(2) eye tracking, and (3) the picture-word Stroop task. These paradigms have been used, respectively, to examine visual word recognition, spoken word recognition, and spoken word production.

7.2.1 Visual lexical decision

In lexical decision, a string of letters is presented on a computer screen and the participant must decide whether it forms a real word or not. Typically the participant presses a "yes" button when the letter string is a real word and a "no" button when it is not, and the speed and accuracy of his or her decision is recorded. When the letter string is a real word, the word can be common and familiar (e.g., cat) or a word that is only infrequently seen (e.g., obtuse). It can be a word whose spelling resembles many other words (e.g., cat looks like hat, mat, rat, etc.) or few, or a word that is concrete and easy to visualize (e.g., cat) or abstract and hard to visualize (e.g., obtuse). When the letter string does not form a real word, it is typically a possible word in that is pronounceable and follows the spelling rules of the language (e.g., blart). Using non-words that are possible words in the language ensures that the participant cannot use spelling or phonology alone to make the lexical decision; the mental lexicon itself must be accessed to determine whether the word is known. The task has been used extensively in psycholinguistic research within a single language to examine lexical access (e.g., Balota, 1994).

For bilinguals, lexical decision provides a context in which a set of factors can be manipulated to determine whether only one or both languages are active when a string of letters is presented. The logic in many of the bilingual studies is to exploit similarities that exist across languages in orthography or phonology. For example, in languages such as Dutch and English there are a significant number of translation equivalents that are identical or very similar in their spelling patterns. These translations are called cognates and provide a clever means to determine whether bilinguals are able to function monolingually in performing a task such as lexical decision. In Dutch and English the words 'bed' and 'hotel' are cognates because they have the same spelling in both languages. Other cognates, such as 'tomaat' in Dutch and 'tomato' in English are similar, but not identical, in the two languages. If a bilingual can access a word in one language without contacting the other, then lexical decision performance for cognates should be no different than lexical decision for words that are unambiguous. Thus, a Dutch-English bilingual performing lexical decision exclusively in English (i.e., are these strings of letters words in English or not?) should not be influenced by the fact that cognates have representations that are similar in Dutch (see figure 7.1 for an illustration of the task). The results of many experiments (e.g., van Hell & Dijkstra, 2002) show that bilinguals are in fact faster to decide that a string of letters is a real word in one language alone when it is a cognate.

A related type of experiment uses interlingual homographs, or words that have similar orthography and/or phonology in two languages, but different meanings. For example, in Dutch, the word 'room' means cream, as in cream for your coffee. If a Dutch-English bilingual can effectively switch off his or her Dutch when

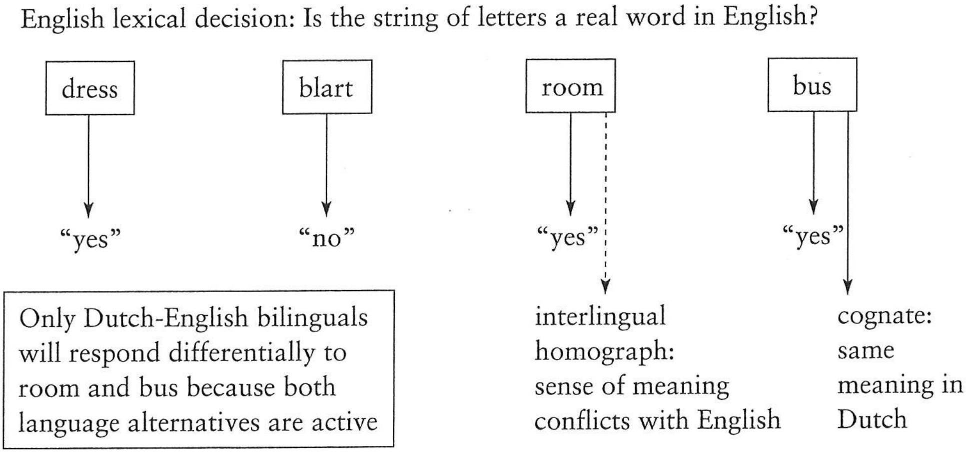


Figure 7.1: An illustration of a lexical decision task performed in English exclusively but including words that are ambiguous with respect to language membership. For a Dutch-English bilingual, the word *room* is an interlingual homograph, meaning room or cream. The word *bus* is a cognate, with the same meaning in Dutch and in English.

reading in English, then a word like 'room' should be processed no differently than any other English word that does not have this special relation to Dutch. The alternative sense of the word 'room' should not intrude. However, many experiments have shown that the unintended language does affect lexical decision performance (e.g., Von Studnitz & Green, 2002). When lexical decision is performed in the second language (L2), there is interference from the unintended sense of the word in the first language (L1). However, Dijkstra, Van Jaarsveld, and Ten Brinke (1998) have shown that when the lexical decision task is altered slightly to be a language-general task (what they call generalized lexical decision), there is facilitation for interlingual homographs because under these conditions any activated sense of the word is sufficient to make a "yes" response that the string of letters is a word.

A criticism of the logic of these word recognition studies is that both languages are necessarily active by virtue of the participant's knowledge that the experiment is about their bilingualism. Grosjean (2001) has argued that when bilinguals are in "bilingual mode" with both languages active to some degree, there will necessarily be evidence for cross-language interactions of the sort that have been reported. A number of recent studies have attempted to address this criticism and to evaluate the effect of the participant's expectations by keeping participants in a strictly "monolingual mode" in one language alone. For example, van Hell and Dijkstra (2002) recruited Dutch university students to participate in an experiment in Dutch exclusively. Unbeknownst to the participants, some of the items in the experiments were cognates in Dutch and English or in Dutch and French. They showed that there was facilitation in recognizing cognates related to controls even when English and French words were not present in the experiment and there was no explicit instruction regarding any language other than Dutch, their L1. The result is striking because Dutch was the native and dominant language of these bilinguals and one might expect that they would be able to function independently in their L1, yet the L2 and L3 affected their performance in the task.

7.2.2 Eye tracking

Monitoring the movements of the eye while a person reads visually presented text is a task that has often been used to infer the processes that underlie skilled reading (see section 7.4 below on processing sentences). Recent studies of spoken word recognition have also used eye movements to track the pattern of eye fixations when a listener hears a word while looking at a display of objects whose names bear some similarity to the phonology of the spoken word. This paradigm, developed initially in the domain of spoken word recognition in the native language to test the seriality of lexical selection mechanisms (e.g., Allopenna, Magnuson, & Tanenhaus, 1998), has been extended to investigate the parallel activation of words in both of the bilingual's languages when they hear a word in one language alone (e.g., Marian & Spivey, 2003; Ju & Luce, 2004).

In this task, the participant, wearing a head-mounted eye tracker, is seated in front of a display that contains four objects (either real objects or pictures on a computer display). The person is instructed to fixate on a central point on the screen until he or she hears a spoken target word. In the computerized version of the task, the participant clicks on the picture that corresponds to the spoken word. The critical manipulation in these studies is the presence of objects whose names sound like the spoken word either in the language presented or in the bilingual's other language. To illustrate, we use the materials in Spanish and English from the Ju and Luce (2004) study (see figure 7.2). Here, the spoken target word is *playa* (meaning beach in Spanish). The correct response is to click on the picture of the beach. However, one of the distractor pictures shows a pair of pliers (in Spanish the word for pliers is *alicate*). If a bilingual can perform this task in one language exclusively, then the presence of the pliers should have no effect on performance because the Spanish word *alicate* bears no resemblance to the word *playa*. However, if both language alternatives are activated in parallel, then the English word *pliers*, which is phonologically similar to *playa*, should intrude momentarily and the pattern of eye fixations should reveal that participants are more likely to glance at the picture with a phonologically similar name than at other control pictures. A series of experiments by Marian and colleagues (e.g., Marian & Spivey, 2003) has shown that Russian-English bilinguals appear to activate both language alternatives. However, Ju and Luce modified this claim to show that whether evidence for parallel activation was obtained depended on the acoustic properties of the spoken word. If the word was spoken with Spanish-appropriate voice onset times, bilinguals were less likely to fixate pictures with names that were phonologically similar in English. For present purposes, the main point of this illustration is to demonstrate the sensitivity of the cognitive system to processes that reveal themselves over time and across modalities. Here, the pattern of eye fixations corresponds to the nature of the lexical information that is activated when a spoken word is heard. The process of perceiving spoken language can of course be studied within the auditory

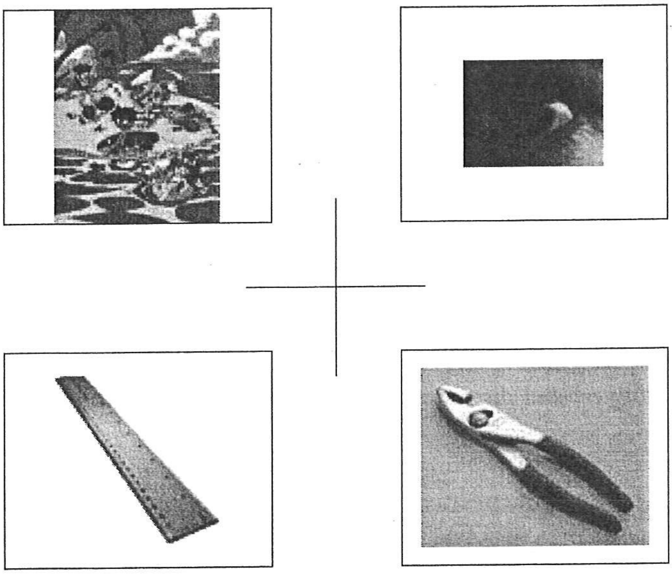


Figure 7.2: An illustration of the eye-tracking paradigm used to study cross-language activation in spoken word recognition. The materials are adapted from Ju & Luce (2004). Here a Spanish-English bilingual hears the word *playa* and must click on the appropriate picture of a beach scene. The question is whether the bilingual glances briefly at the picture of the pliers which is an English word that is phonologically similar to the Spanish word *playa*.

domain alone (see section 7.3 below on phonology and Grosjean & Frauenfelder, 1997, for a review).

7.2.3 *Picture-word Stroop*

The final example we describe to illustrate how lexical processing has been studied in bilinguals focuses on the way in which bilinguals plan spoken utterances to produce a single word in only one of their two languages. Language production has been far less studied than comprehension, in part because it is difficult to devise tasks that encourage speakers to produce uniform utterances. As a consequence, most of the early research on language production relied on patterns that could be inferred from large corpora of speech errors (see Poulisse, 1999, for an example of a speech error analysis for L2 learners). Although errors are informative with respect to the constraints that guide speech planning, they do not provide a sensitive means to examine the planning process as it unfolds over time when speech is produced accurately. A solution to this problem has been to invent tasks that simultaneously constrain spoken utterances and provide a method for asking what sort of information is available to the planning process at different points in time prior to articulation.

The picture-word Stroop task has been used extensively in recent studies of monolingual and bilingual language production to examine the time course of planning and to evaluate alternative models of the planning process (e.g., Levelt, Roelofs, & Meyer, 1999). The task is a variant of the color-word naming task first described by Stroop (1935). In the original Stroop task, participants named the color of the ink in which printed words appeared. The Stroop effect is the interference that is induced when a color word appears in an incongruent color (e.g., the word *blue* printed in red ink). In the picture-word task, a picture is presented and the participant is asked to speak its name aloud as quickly as possible. At some point just prior to or following the presentation of the picture, a word distractor is presented either auditorily or visually. The participant is told to ignore the word and name the picture. By manipulating the relation of the distractor word to the picture's name and the timing of its presentation, it has been possible to map out the time course of speech planning. Generally, there is interference for semantically related distractors when they are presented early in the planning process and facilitation for phonologically related distractors when they are presented late in the planning process.

The overall pattern of distractor effects in picture-word interference suggests that the phonology of the spoken utterance can only be encoded once the meaning of the intended utterance is specified. A debate in this area of research is whether the sequencing that allows speech planning to proceed from meaning to phonology is a strictly serial and encapsulated process or one that reflects interactions across different levels of information (e.g., Dell & O'Seaghdha, 1991; Levelt, Roelofs, & Meyer, 1999). Although a full discussion of the theoretical background is beyond the scope of the present chapter, for the purpose of extending the methods used into the bilingual domain, we consider briefly the focal issue towards which the research has been designed.[2]

Like research on bilingual word recognition, the question in bilingual word production has been whether alternatives in the non-target language (i.e., the language not spoken) are active during the planning of an utterance (see Costa, 2005, for a review of this literature). Unlike word recognition, production is a process that is initiated by a conceptual event (e.g., a picture to be named, a word to be translated, an abstract idea to be spoken), so it might seem that in the course of conceptualizing the intended utterance only the language to be produced would be active. Although there is debate in the literature about the selectivity of language production (e.g., Costa, 2005), a great deal of evidence suggests that words in both of the bilingual's languages are active at least to the level of abstract lexical representations and perhaps to the point of actually specifying the phonology associated with the translation.

How can the picture-word Stroop paradigm be used to inform the debate about whether alternatives in the bilingual's other language are active when they intend to speak in one language only? A number of studies have varied the language of the

[2] The issue of whether processes are encapsulated refers to a longstanding debate in psycholinguistics concerning the modularity of language (e.g., Fodor, 1983). The basic question is whether certain language functions (e.g., parsing a sentence into its grammatical components or retrieving the meaning of a word) are separate from other cognitive representations and goals, or are guided by them.

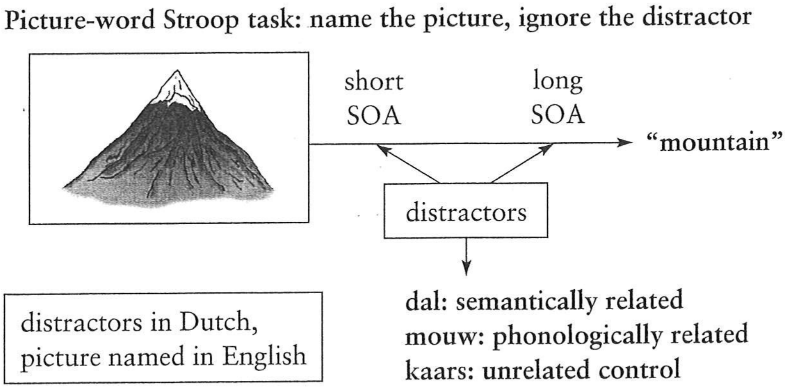


Figure 7.3: An illustration of the cross-language picture-word Stroop task. The materials are adapted from Hermans, Bongaerts, De Bot, & Schreuder (1998). Here a Dutch-English bilingual names a picture in English and attempts to ignore distractor words presented in Dutch.

distractor and the language in which the picture is to be named to investigate this issue (e.g., Hermans, Bongaerts, De Bot, & Schreuder, 1998; Costa, Miozzo, & Caramazza, 1999). An illustration of the paradigm adapted from Hermans et al. (1998) is shown in figure 7.3. Here a Dutch-English bilingual is asked to name a picture of a mountain as the word *mountain* in English. The distractor is presented at the same time as the picture, at a brief delay following the picture, or after a longer delay. The interval between the presentation of the picture and the onset of the distractor is known as the stimulus onset asynchrony, or SOA. In this example, the distractor is the word *dal* in Dutch which means valley in English and is therefore semantically related to the word to be spoken but presented in the non-target language. By comparing the time it takes bilinguals to name the picture when it is accompanied by a semantically related word, like *dal*, a phonologically related word like *mouw*, which sounds like the first syllable of 'mountain' in English but means sleeve, or an unrelated control word, like *kaars*, which means candle in English, it is possible to estimate what sort of information is active at any given moment in time before the word 'mountain' is actually spoken. Hermans et al. found a similar pattern of results in picture naming in the L2 regardless of whether the language of the distractor was L1 or L2. Semantically related distractors produced interference relative to unrelated controls and the semantic interference was greatest early in the time course of speech planning. Phonologically related distractors produced facilitation relative to unrelated controls, and the facilitation was greatest late in the time course of speech planning. Costa, Miozzo, and Caramazza (1999) reported similar results for Catalan-Spanish bilinguals. The finding that distractors in the non-target language also produce interference and facilitation in picture naming suggests that, as seems to be the case for bilingual word recognition, lexical access in bilingual speech production is initially language-non-selective, with alternatives activated in both languages in parallel. Other production paradigms have produced evidence that converges with these general conclusions. These include speaking the translation of individual words (De Groot, 1992; Kroll & Stewart, 1994) and monitoring the phonemes in the names of pictures (Colomé, 2001).

7.3 Sounds

A longstanding issue in the study of bilingual phonology is the question of how L2 phonetic categories are both produced and perceived. Much of the research focuses on late bilinguals, i.e., bilinguals who have learned their second language near or past puberty and who, in many cases, have lived for extended periods of time in the L2 environment. As is well known, late L2 speakers often differ in their production and perception of phonetic categories from native speaker norms. From the perspective of the researcher, the phonology and phonetics of bilingualism provide a fertile testing ground for exploring hypotheses about the critical period for language acquisition, for examining issues of neural plasticity throughout the development of L2 proficiency, for probing how L2 learning is constrained in comparison to L1 learning and/or by the phonological system of the L1 (see Flege, 2003, for a review), and for understanding the generally complex issue of accentedness in L2 speech production and perception (see Piske, MacKay, & Flege, 2001, for a review). While a lengthy discussion of the various theoretical alternatives associated with the issues noted here is beyond the scope of this chapter, in the rest of this section we will examine a number of particular studies that both exemplify a range of techniques employed and inform the theoretical questions noted above.

7.3.1 Production

Production tasks are commonly employed as a means of assessing bilingual speech for numerous theoretical goals. These can differ in terms of both elicitation technique and size of the speech sample elicited. In most cases, the task involves having participants read aloud either target phrases (e.g., Flege, 1987; Moyer, 1999), single words (e.g., Flege & Eefting, 1987; Moyer, 1999), or larger chunks such as paragraphs (e.g., Moyer, 1999), and recording their speech production for subsequent analysis. Some studies have also employed repetition techniques in which participants listen to and then repeat experimental items produced by a native speaker either immediately (e.g., Markham, 1997) or after a delay designed to minimize the chances of direct imitation from sensory memory (e.g., Piske, MacKay, & Flege, 2001). Finally, other studies have employed less controlled techniques designed to elicit speech tokens under increasingly natural conditions, such as asking participants to talk about events in their lives (e.g., Moyer, 1999). In broad strokes, the recorded data are subjected to two kinds of analyses, depending on the goals of the research. The first involves carrying out acoustic analyses of the bilingual production data and comparing the results with measurements of native speaker controls. The second involves using native speakers to rate the L2 productions of the participants, an approach used extensively in studies assessing L2 accentedness (see Piske et al., 2001, for an overview of rating techniques).

Flege (1987) provides a useful example of a typical production study and illustrates how acoustic measurement techniques can be used to address questions of neural plasticity and the critical period hypothesis, i.e., the hypothesis that language learning (or, in our case, speech learning) is rigidly constrained by a critical period in maturation ending around puberty (see DeKeyser, 2000, for a review). Flege examined the L1 and L2 productions of English/French and French/English bilinguals for the French and English alveolar stop /t/ and for the /i/ and /u/ vowels in English and the /y/ and /u/ vowels in French. For reasons of space, we focus here on the production results for /t/.

Both French and English have the voiceless stop /t/ in their phoneme inventories. In English, the category /t/ (like other voiceless stops) is realized phonetically as a long lag or aspirated stop [t^h] when initial in a stressed syllable (such as /tu/ *two*). In French, however, /t/ is produced as a short lag or unaspirated stop [t] in the same position (as in *tous* /tu/ 'all'). The term *lag* refers to voice onset time (VOT), i.e., to the interval of time between the moment when the closure of the stop consonant is released and the onset of voicing in the following vowel. VOT can be measured with great precision from a display of the acoustic waveform, as can be seen in figure 7.4, made with the Praat acoustic analysis software (Boersma & Weenink, 2005).

Flege (1987) measured the VOTs of the stop /t/ for three groups of late English/French bilinguals whose experience with French varied (US students 3–6 months removed from a 9-month abroad program in France; French professors whose L1 was English and who were living in an English-language context; and L1 English speakers living in Paris for an average of 11.7 years) and for a group of late French/English bilinguals living in an English context (French women living in Chicago for an average of 12.2 years). Their data were compared to production data from monolingual English and French speakers collected under the same experimental conditions. As is standard in such phonetic studies, multiple repetitions of the critical phonetic context were produced by each speaker to allow for the calculation of a reliable mean VOT value for each participant. Data were elicited in two conditions. In Condition 1, participants read seven phrases in isolation. All phrases were matched for the initial phoneme sequence /tu/ by employing the same

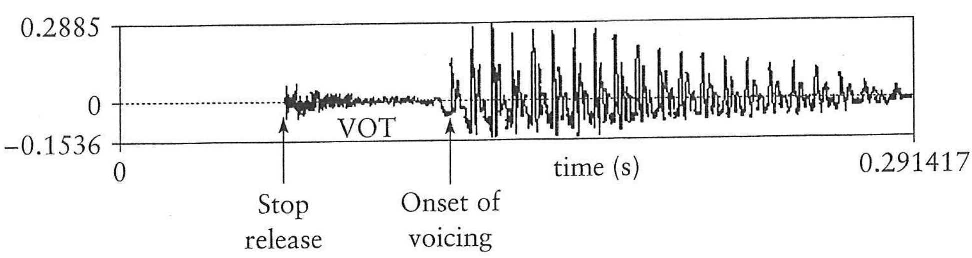


Figure 7.4: An illustration of how VOT is measured as the time interval between the release of the stop (as indicated by the arrow to the left) and the onset of voicing in the vowel following the stop. This is the acoustic wave for a token of the English CV syllable [p^ho].

initial word. The English phrases began with the word *two* (e.g., two little boys), while the French phrases began with the word *tous* (e.g., *tous les soldats* 'all the soldiers'). In Condition 2, participants were prompted to use each of the phrases in an original sentence while being cued by the written phrases from Condition 1. VOT was measured for each instance of initial /t/ in each condition, and a mean was calculated for each speaker by condition. After Speaking Condition was found to be non-significant for VOT duration, a mean of the two means was calculated for each speaker and used in the analysis.

Of particular interest here are the results for two of the groups, the Americans living in Paris and the French women living in the United States. Specifically, Flege's results show that for the bilingual L1-English/L2-French speakers, VOT durations in English are significantly shorter than those of monolingual English controls. Likewise, the bilingual L1-French/L2-English group produces the French stops with significantly longer VOTs than do the French monolingual controls. That is, in these bilinguals, the VOT targets for /t/ in their L2s (longer in English and shorter in French) are reshaping the realization of their respective /t/ categories in their L1s. Here we can see how bilingualism again provides a fundamental tool for testing theories of language learning. The logic of the argument in this case is that a strong critical-period hypothesis incorrectly predicts that late learning of an L2 should not reshape the phonetic space of an L1. By contrast, theories which do not assume that the neural plasticity necessary for speech/language acquisition declines precipitously after a critical period predict that sufficient experience in an L2 (such as longtime residency in an L2 context and constant use of the L2) may affect even aspects of the phonetic system of the L1.

7.3.2 Perception

As with production, the perception of phonetic categories in an L2, particularly by late L2 learners, often diverges from the perception of the same categories by native speakers. Though theoretical models differ in the specifics of their approaches to the problem, there is a general consensus that L2 perception is filtered by knowledge of the L1 phonological system (see, e.g., the Perceptual Assimilation Model, Best, 1995, and the Speech Learning Model, Flege, 1988, 2002). Much research indicates that one particular way that L1 learning shapes the perception of L2 categories is that L1 acquisition involves perceptual tuning to the phonetic properties necessary for producing and perceiving phonological distinctions in the L1. Over the course of maturation, this tuning leads to a warping of the acoustic space through which subsequent L2 learning is filtered (e.g., the Native Language Magnet, Kuhl, 2000).

A number of experimental paradigms have been deployed to examine questions of both how and how well L2 phonetic categories are perceived. For example, Iverson, Kuhl, Akahane-Yamada, et al. (2003) utilize three different tasks in examining how language experience with the L1 shapes the perception of non-native categories. Specifically, they examine how Japanese listeners differ from native English listeners in their perception of the English /l/ vs. /r/ contrast. For their stimuli, Iverson et al. created a set of 18 CV syllables, synthesizing a continuum from /ra/ to /la/ in English by systematically varying two spectral properties, the frequencies of the second (F2) and third (F3) formants of the initial liquid consonant. They tested their participants in three ways: via the collection of identification and goodness ratings, via similarity scaling, and via a discrimination task.

The identification and goodness tasks involved having participants listen to a stimulus item – items were presented twice in this experiment in two randomized blocks – and identify it with a phonetic category in their native language. Upon identification, participants were also asked to rate each token on a scale of 1 (bad) to 7 (good) as an exemplar of that category. In similarity-scaling tasks, participants are also asked to provide ratings. In this case, however, participants were presented aurally with pairs of stimuli and asked to rate their similarity on a scale of 1 (dissimilar) to 7 (similar). Stimuli were presented in a single randomized experimental block of 306 trials, with every pair of the 18 stimuli items presented in both possible orders. Finally, AX discrimination tasks consist of asking participants to listen to a pair of stimuli and make a determination as to whether they have heard the same or different items. In this experiment, stimuli were presented in a single randomized block of 480 trials, consisting of 48 same-pair and 48 different-pair trials for each pair of stimulus items. The different condition pairs differed in this task only along the dimension of the third formant (F3).

For Iverson et al. (2003) the similarity-scaling and discrimination tasks yielded converging results. Japanese listeners are, erroneously, well tuned to changes along the F2 dimension, while American listeners are finely tuned to changes along the F3 dimension that signal the phonetic category boundary between English /r/ and /l/. Iverson et al. argue that the identification and goodness ratings are relevant in that they provide a means of explaining why Japanese listeners differentially tune to spectral components of the stimuli. Specifically, as F2 falls Japanese listeners begin to identify stimuli as belonging to the Japanese /w/ category instead of as belonging to /r/. That is, their experience with Japanese has shaped their acoustic space in such a way as to lead them to attend to cues such as the F2 difference signaling the contrast between /r/ and /w/. By contrast, the F3 changes to which English speakers are highly tuned fall within a single Japanese category, /r/, and Japanese listeners thus show a reduced sensitivity to change along this dimension.

If L2 perception is constrained by the filter of L1, the question then arises of how malleable the system is over the course of L2 learning. Escudero and Boersma (2004), tested Spanish speakers' perception of the English /i/ ~ /ɪ/ contrast using a synthetic /i/ ~ /ɪ/ continuum which varied both the frequency of F1 and the duration of the synthetic vowel in an experimental design similar to that of Iverson et al. (2003). This contrast is notoriously difficult for Spanish learners of English (cf. Bradlow, 1995; Fox, Flege, & Munro, 1995). Interestingly, Escudero and Boersma tested two groups of L1 Spanish speakers living in the L2 environment: a group living in a Scottish English environment where the /i/ ~ /ɪ/ distinction is realized primarily by differences in the frequency of the first formant (F1) of the vowel – a property inversely related to vowel height – and a group living in southern England, where the dialect primarily uses duration differences in realizing the difference between the categories. In the Escudero and Boersma study, participants listened to each stimulus token in isolation and performed a forced-choice task, selecting between a picture of a *sheep* (indicating the /i/ category) and a picture of a *ship* (indicating the percept of the /ɪ/) category. The task is different from the Iverson et al. (2003) identification task in that participants in this study were forced to choose between L2 categories (indirectly in the form of pictures) rather than identifying the stimulus item with an L1 category. Importantly, the results indicate that although they did not exhibit native-like perception with respect to the L1 English speaker controls, more advanced L2 learners adopted strategies that involved tuning to the properties of the dialect in which they were immersed, thus exhibiting a good degree of plasticity in their developing sensitivity to dialect-particular acoustic cues.

A striking contrast with the cases of L2 plasticity that we have noted in both production and perception can be found in the research conducted on the /e/ ~ /ɛ/ contrast in Spanish/Catalan bilinguals – a phonemic distinction present in Catalan but lacking in Spanish. In a series of experiments (Sebastián-Gallés & Soto-Faraco, 1999; Bosch, Costa, & Sebastián-Gallés, 2000; Pallier, Colomé, & Sebastián-Gallés, 2001) researchers have employed a range of techniques and found that highly proficient yet Spanish-dominant early Spanish/Catalan bilinguals perform in a non-native fashion in tasks involving the processing of the Catalan /e/ ~ /ɛ/ distinction, in comparison with the superior performance of Catalan-dominant early Catalan/Spanish bilinguals. Two of these studies are particularly useful in that they allow us to review additional experimental approaches found in the bilingual perception literature.

Sebastián-Gallés and Soto-Faraco (1999) employ a modified version of the gating technique (Grosjean, 1980, 1988) to probe for differences in the way that highly proficient Spanish-dominant, i.e. Spanish/Catalan, bilinguals process the Catalan /e/ ~ /ɛ/ distinction (among other contrasts) in comparison with the performance of highly proficient Catalan-dominant, i.e. Catalan/Spanish, bilinguals. In a gating experiment, an aurally presented stimulus, usually a word, is played for participants in successively larger increments. In this sense, the participant's exposure to the stimulus is gated. At each gate, participants must make a forced choice between possible forms and then rate the confidence with which they have made their choice. In the Sebastián-Gallés and Soto-Faraco study, gated stimuli consisting of one of two minimally distinct non-words in Catalan were presented while written pairs of the non-words were shown on a computer screen. During each gating trial, participants chose one of the displayed forms and rated the confidence of their choices on a 1 to 9 scale. The authors analyzed their results in terms of two key points: (1) the *isolation point*, the gate at which participants correctly identified a target word with no further change in their subsequent choices upon hearing larger chunks of the gated form; and (2) the *recognition point*, the gate after which participants expressed a confidence rating of 8 or higher in their choice. The results indicate that despite the fact that the Spanish/Catalan bilinguals are all highly proficient Catalan speakers who had acquired Catalan in their early childhood, these speakers are less efficient, i.e., they need significantly more acoustic information than did the Catalandominant bilinguals to successfully complete the gating task. Arguably, then, the gating task provides a fine-grained way of distinguishing in a fairly precise manner between even highly proficient bilingual groups. In a larger sense, Sebastián-Gallés and Soto-Faraco claim that their results demonstrate that the tuning effect of L1 categories (in this case, driven by the lack of an /e/ ~ /ɛ/ distinction in the L1 Spanish) may persist deeply into L2 acquisition, even when an individual's L2 is learned early, involves intensive exposure, and is used extensively.

With the exception of the gating task (the status of which is ambiguous; see, e.g., Grosjean, 1996), the techniques most often used in bilingual phonetic perception studies involve off-line tasks. Pallier, Colomé, and Sebastián-Gallés (2001) provide a useful example of how on-line tasks can also be used to address the issue of how the phonological system of L1 filters the perception of L2 phonetic categories. In continuing to examine the difficulty that highly proficient Spanish/Catalan bilinguals have in perceiving the Catalan /e/ ~ /ɛ/ distinction, Pallier et al. employ a medium-term auditory repetition priming technique. The auditory repetition priming technique is a variation on an auditory lexical decision task. Specifically, it involves presenting participants with both spoken words and non-words and asking them to make a decision as quickly as possible regarding whether the presented stimulus is a word or not. The task is called an auditory repetition priming technique, because some of the words and non-words in the stimuli list are presented twice. A general finding of this task is that real words are responded to more rapidly when seen for a second time (i.e., they are primed), while response times for non-words are not faster when presented a second time. Pallier et al. capitalize on this effect by including minimal pairs of Catalan words such as [pere] 'Peter' and [perɛ] 'pear.' The logic of their experimental design is that if listeners process such minimal pairs as acoustically different and thus distinct lexical items, no priming effects should be found for them. On the other hand, if listeners hear such forms as homophones by failing to perceive the difference in their final vowels, priming effects are expected. Their results are consistent with the other studies in which Spanish-dominant Spanish/Catalan bilinguals do not perform in the same fashion as Catalan-dominant bilinguals. Specifically, the Spanish-dominant group differs significantly from the Catalan-dominant group in that the former exhibit a facilitation effect for minimal pairs of Catalan forms that is of the same magnitude as the facilitation effect found for real repetitions of identical forms. This indicates that Spanish-dominant participants are not appropriately tuned to the spectral differences cueing contrasts such as /e/ vs. /ɛ/. At the same time, the overall reaction-time data corroborate the authors' claim that they are testing highly proficient bilinguals, given that the Spanish-dominant group did not differ significantly from the Catalan-dominant group in either their response times or their error rates in the lexical decision task for Catalan words. Methodologically, these results are interesting for our purposes in that they show how converging results can be obtained with a variety of tasks involving both behavioral, on-line techniques, gating, and off-line tasks such as discrimination and identification.

7.3.3 *Imaging*

Recent advances in imaging techniques have also been brought to bear on many of the questions addressed above. Though a thorough review is beyond the scope of this chapter (for further discussion see chapter 8 in this volume), we will discuss here Winkler, Kujala, Tiitinen, et al.'s (1999) event-related potentials (ERP) study of phonetic category perception in native, naive non-native, and proficient L2 speakers of Finnish (see Handy, 2004, for an overview of ERP experimental designs, approaches, and applications). Broadly speaking, ERP is a functional brain-scanning technique that allows for the non-invasive measuring of brain activity during cognitive processing. Electrodes are attached to the scalp in order to measure ongoing electrical activity as an electroencephalogram (EEG). Event-related potentials are calculated as averages of electrical activity in the brain that are time-locked to the presentation or to the response of particular stimuli. The experimental approach taken in Winkler et al. employs a design in which a "standard" binaurally presented stimulus (a synthesized Finnish /e/ vowel) is played repeatedly to participants (82.5 percent of the time) and occasionally interrupted by one of two "deviant" stimuli (either the Finnish vowel /æ/ or the Finnish vowel /y/, also both synthesized). The experiment tests for the elicitation of what is known as a mismatch negativity event-related potential (MMN) during the processing of the deviant stimuli. Research has shown the MMN potential to be associated with bottom-up, preattentional phonetic processing (see Näätänen, 2001, for an extensive review). Of most relevance here is that elicitation of the MMN reflects the perception of change along a particular phonetic dimension, in this case as a function of change from the repeated "standard" to the "deviant" stimulus.

Winkler et al. (1999) tested the perception by Hungarians of a Finnish vowel contrast /e/ ~ /æ/ that falls in the acoustic space of a single vowel category in Hungarian. Given the preattentional nature of MMN elicitation, they reasoned that an ERP study of the perception of non-native contrasts would provide a useful mechanism for exploring the issue of brain plasticity in the late learning of a second language. Specifically, they hypothesized that if Hungarian speakers are unable to perceive the vowel contrast, they should also fail to elicit MMN potentials on deviant trials. By contrast, the elicitation of MMN potentials on deviant trials would indicate a deep, low-level sensitivity to the acoustic difference between the two vowel categories in Finnish. Importantly, they found that MMNs were not elicited for the naive Hungarian speakers when exposed to the Finnish vowel contrast, i.e. in response to the presentation of the deviant stimuli /æ/ vowels. In keeping with the performance of a different group of naive speakers on an off-line discrimination task, the ERP data indicated that these naive speakers were simply not perceiving the difference between the Finnish vowels but rather were perceiving both /e/ and /æ/ as tokens of a single vowel category. By contrast, the relatively proficient Hungarian L2 speakers of Finnish displayed a clear MMN response to the presentation of the deviant tokens – a response pattern, in fact, that did not differ significantly from that of native Finnish-speaking control participants. These results are a bit of a conundrum when compared to the apparent non-plasticity characterizing the early and highly proficient Spanish/Catalan bilinguals' performances on an array of tasks as described above. For the late L2 Hungarian speakers, the results, when taken together with the non-elicitation of MMN responses in the naive group, strongly suggest that late L2 learning is characterized by continued brain plasticity at the very lowest levels of phonetic perception. Finally, from a methodological perspective, imaging studies are interesting in that they show how non-behavioral tasks can add to our arsenal of experimental paradigms for testing questions of bilingual phonetic processing.

7.4 Sentences

When we try to comprehend sentences in our second language (and, for that matter, in our first), we face many uncertainties about how the people or objects referred to are connected to one another. This is so because when our eyes move along the printed text in a left-to-right fashion, the information needed to establish correct dependencies between word strings is not yet available. In other words, we need to wait.

So what does the L2 reader do under these conditions of uncertainty? Given that L2 speakers approach the task of sentence processing with a fully developed processing system from their L1, one may ask what representations are created while speakers process written text in their L2, what types of information are used in constructing them, and when these representations are formed. It is reasonable to imagine that during the earlier stages of L2 learning, L2 speakers rely, at least partially, on sources of information from their first language (e.g., lexical information encoded in verbs, such as verb argument structure) to construct a licit syntactic construction (i.e., *parse)* in the L2. And one would expect that as language proficiency increases, sentence processing in the L2 should approximate that of monolingual speakers of the target language.

Experimental work in L2 sentence comprehension has investigated these questions using an array of psycholinguistic techniques, ranging from the very simple to the more highly sophisticated and powerful (Juffs & Harrington, 1996; Weber-Fox & Neville, 1996; Frenck-Mestre & Pynte, 1997; Hahne & Friederici, 2001; Dussias, 2003; Felser, Roberts, Gross, & Marinis, 2003; Fernández, 2003). This rapidly growing body of work suggests that, when the L2 learner parses sentences in the L2, his or her performance is sometimes strikingly close to that of native speakers, but other times it is not. The most compelling type of evidence in support of the first claim comes from studies that have used event-related brain potentials (ERPs) while speakers are exposed to sentences that vary systematically with respect to particular semantic characteristics. Monolingual English speakers and L2 speakers of English faced with the sentence *The scientist criticized Max's event of the theorem* will be, by and large, equally sensitive to the semantic anomaly contained within it (Weber-Fox & Neville, 1996). At the same time, apparent discrepancies between L1 and L2 speakers have been obtained in the ambit of syntactic processing, providing support for the second claim.

Methodological advances in psycholinguistics have provided the community of researchers interested in L2 sentence comprehension with valuable information about the experimental techniques commonly used to advance our understanding of the psychological processes underlying sentence comprehension, as well as with rich and remarkably detailed evaluations of what each technique can and cannot reveal about comprehension processes. In this section, we consider the methods that have been most commonly used to investigate L2 sentence comprehension. Because researchers are most often interested in tracking L2 sentence processing as it unfolds in real time, we will limit our discussion to a family of techniques that have come to be known as *on-line* methods.

7.4.1 Self-paced reading

Without a doubt, self-paced reading has been the on-line method most widely used in L2 sentence comprehension research. In this task, a stimulus sentence is presented on a computer screen, segmented into words or phrases commonly referred to as *displays*, which are presented one at a time. Typically, the participant initiates the experiment by pressing a *trigger* (e.g., a pedal, or a key on a button box or a computer keyboard). This action brings up the first display. Participants read the display, press the trigger to request the next display, and continue performing the same routine until they reach the end of the experiment. In this task, the measure of interest is the time that participants spend reading a critical display (i.e., the time that has elapsed between successive trigger presses), compared to a control condition.

Self-paced reading tasks have been extensively used in the L2 parsing literature to investigate how the L2 parser proceeds in the absence of lexical constraints, as is the case of adjunct phrases or modifier phrases. In one such study, Dussias (2003; see also Felser, Roberts, Gross, & Marinis, 2003; Fernández, 2003; Papadopoulou & Clahsen, 2003) employed the task with Spanish-English and English-Spanish bilinguals to investigate their attachment preferences for structures of the type NP1-of-NP2-RC (e.g., *El perro mordió al cuñado de la maestra/ que vivió en Chile/ con su esposo/* 'The dog bit the brother-in-law of the teacher (fem.) who lived in Chile with her husband'). All sentences were segmented into three displays – as indicated by the forward slashes in the example above. When the first sentence was requested, the first display of an item appeared centered on the screen and the clock started. The participants read this display and then pressed a key to request the second display. The time that elapsed between the onset of the first display and the request for the second display was recorded. Additionally, the first display was replaced by the second display, and the clock started again. This sequence of events repeated itself until the end of the sentence was reached. The critical comparison in this study was the reading time for the last display; however, reading times for displays 1 and 2 were also compared to ensure that there were not significant differences between them. The findings revealed that the control groups (Spanish and English monolinguals) showed the conventional bias for high attachment and low attachment (respectively) reported in the literature. The English-Spanish bilinguals did not exhibit any preference for high or low attachment when processing the ambiguous sentences, but, remarkably, the Spanish-English speakers showed a consistent preference for low attachment when reading sentences in their first and second languages, suggesting that the parsing routines used to process the L2 had an impact on the processing of the L1, and that the methodology did not distort the cognitive processes that are linked with the detection of the syntactic ambiguity being studied. Like the studies reviewed earlier on recognizing words and speech sounds, these results suggest a high degree of plasticity and interaction across the bilingual's two languages.

In one variation of the self-paced reading task, dubbed the *moving-window programme* (Just, Carpenter, & Woolley, 1982), the display moves from left to right with each trigger press to allow the words of the sentence to occupy the same position in the screen that would surface if the sentence had been displayed as a whole. All letters, apart from the letters of the word in current view, are replaced with dashes (or equivalent markers). In the reading moving window paradigm, the text can be presented in a *non-cumulative* fashion (i.e., as each successive word in the sentence is prompted, the previous one disappears) or *cumulatively* (previously read words remain on the screen as new ones are added). Because the cumulative version has the disadvantage that participants may press the trigger to display all the words in a sentence, and only later initiate the actual reading task, researchers typically favor non-cumulative displays over cumulative ones.

One advantage of the moving window task is that it allows for the collection of word-level reading times, thereby allowing the experimenter to identify the specific loci of processing difficulty. To illustrate, Juffs and Harrington (1996) compared a full-sentence presentation task with a non-cumulative moving window task to examine how Chinese learners of English processed sentences such as *Who did Ann believe ____ likes her friend?* and *Who did Ann believe her friends like ____?* The sentences differed in that the first one is assumed to require extraction of the *wh* element from a subject site (indicated by ____), whereas the second requires extraction from an object site. Juffs and Harrington predicted that subject extraction sentences ought to present more difficulty for the parser than object extraction sentences, because the former would force the parser to reanalyze the *wh*-gap several times before finally arriving at a complete analysis of the sentence. Although the overall findings supported the claim that extraction from a subject site was more difficult than extraction from an object site, the different techniques produced somewhat different results. For example, no significant differences were found between subject and object extractions from finite clauses in the full-sentence condition, whereas these effects emerged in the moving window condition. Moreover, the Chinese learners had proportionally more difficulty than the monolingual English group in judging ungrammatical sentences in the moving window condition than in the full-sentence condition, suggesting that the increased processing demands of the moving window task placed a greater burden on the participants' available cognitive resources.

One of the criticisms leveled against self-paced reading in all its forms is the likelihood that syntactic parsing may be influenced by the type of segmentation employed by the experimenter (Gilboy & Sopena, 1996; but see Mitchell, 2004, for a counter-argument). Gilboy and Sopena (1996) found that relative clause ambiguity resolution was affected by whether the sentences were broken into large segments (e.g., *El perro mordió al cuñado de la maestra/ que vivió en Chile/ con su esposo/*) or smaller segments (e.g., *El perro mordió/ al cuñado/ de la maestra/ que vivió en Chile con su esposo*). A second objection raised against the task is that it relies on a secondary task (a button, a key, or a pedal press) to produce the dependent measure. These and other factors (see, e.g., Mitchell, 2004) have led researchers to favor methods that provide a richer body of data than the single latency that results from self-paced reading. In the next section, we discuss a few of the measures that have allowed researchers to determine with more precision the existence, locus, and time course of processing difficulty.

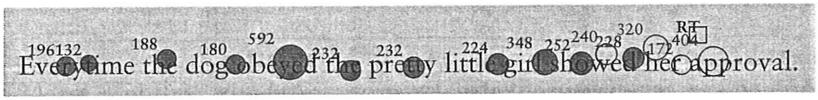
7.4.2 *Eye movements*

Eye-movement records have become a very popular technique in the study of sentence comprehension because they provide an on-line measure of processing difficulty with high temporal resolution, and do not require additional tasks (e.g., button or pedal presses) to yield the dependent measure. An additional advantage of eye-movement records is their high ecological validity. For example, eye movements are a normal characteristic of reading, the reader is free to move back and forth along the printed lines of text, and the text under examination need not be segmented into unnatural displays.

The existence of a large body of literature in experimental psychology that studies eye movements to answer questions about language processing has helped us to better understand the cognitive processes involved in reading. For example, we know that readers extract useful information from a restricted area of the text, usually spanning about 4 characters to the left of a fixation and about 15 characters to the right of the fixation (McConkie & Rayner, 1975). This maximum region from which information is extracted is referred to as the *perceptual span*. We also know that our eyes do not move smoothly along a line of printed text, but rather advance in short jumps called *saccades*. The average English reader makes about three to four saccadic movements in a second, each lasting between 20 and 40 ms. When a word is brought into fovea by a saccade, it is fixated for an average of about 225 ms, though a reader's fixation patterns over a text vary greatly depending on the linguistic characteristics of the words (Pollatsek & Rayner, 1990; Carreiras & Clifton, 2004). For instance, a word's lexical frequency affects its first fixation duration and gaze duration even when length is controlled (Just & Carpenter, 1980; Rayner & Duffy, 1986). Also, the predictability of a word from prior context influences the first fixation duration and the gaze duration on that word (Balota, Pollatsek, & Rayner, 1985), as well as the time it takes to incorporate it into the representation that the reader is constructing for a particular sentence.

What dependent variables are available to the investigator when collecting eye-movement records? For any critical region or regions of interest, a number of measurements can be distinguished. The earliest measure is *first fixation*, defined as the first time the eyes land on a region (whether a single word or a string of words). This measure appears to be sensitive to word frequency. The next measure is *first-pass time*, and refers to the sum of all fixations in a region, from first entering it until the eyes exit to the left or right of the region. On regions with only one word, first-pass time equals *gaze duration* (e.g., Rayner & Duffy, 1986). First-pass time has been found to be most informative in revealing detections of syntactic anomalies.

A. First-pass fixations



B. Re-fixations on the critical region

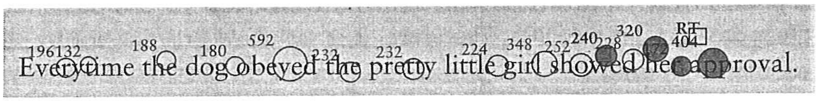


Figure 7.5: An illustration of eye-movement records while Spanish-English speakers are reading a structurally ambiguous sentence. The materials are adapted from Frenck-Mestre & Pynte (1997).

We note here that for both gaze duration and first-pass time, most researchers exclude trials in which the region is initially skipped. Another commonly used measure is *second-pass time*, which refers to the time spent reading a region after leaving the region (in other words, excluding first-pass time or after an initial skip of the region). Finally, *total time* is the sum of all fixations in a region (effectively, the sum of first-pass time and second-pass time). In addition to the measurement of time, another useful dependent measure is the *probability of a regression*, defined as the percentage of regressive eye movements (leftward movements in a language like English) out of a region. This index is usually restricted to first-pass regressions.

Figure 7.5 provides an illustration of an actual eye-movement record of a highly proficient Spanish-English bilingual reading a structurally ambiguous sentence (see Frenck-Mestre & Pynte, 1997, for a discussion of how French-English and English-French bilinguals process this ambiguity). Arrows indicating the trajectory of the eye have been omitted to simplify the image (fixation duration values appear to the left of the fixation). The ambiguity in this construction arises because the noun phrase 'the pretty little girl' can be interpreted either as a complement of the verb 'obeyed,' or as the subject of the ensuing clause. A reader who commits to the first interpretation will be forced to revise the attachment decision once the eyes reach the disambiguating region 'showed.' We observe for the first word in the sentence (i.e., 'every') a fixation on the letter *r*, with a duration of 196 ms. Given that no other fixations occurred on this word, first fixation and gaze duration equal 196 ms. The reading proceeds fairly smoothly, until the participant reaches the disambiguating region (i.e., 'showed'). First fixation on this region occurs on the letter *s*, at a duration of 348 ms. The two subsequent left-to-right fixations fall on the letter *o* and the letter *d*. These are sequenced fixations, with duration values of 252 ms and 228 ms respectively. Because all three fixations occurred before the eye was launched to another region in the sentence, gaze duration for this region equals the sum of the three fixations (828 ms). The next fixation occurs at the word *her*, and lasts 320 ms. The participant then launches a regressive movement back to the disambiguating region, which lands on the letter *e* and lasts 228 ms. In this case, then, second regression time equals 228 ms, and the total time spent reading the region is 1156 ms. It is worth noting at this point that processing difficulty at the disambiguating region occurred during early stages of cognitive processing as indexed by first-pass reading times. This finding could easily have been missed if the data had been collected with self-paced reading, because initial analysis and reanalysis cannot readily be distinguished. Returning now to our example, we note that the last word of the sentence is fixated twice, for 404 ms and 172 ms (576 ms). Generally, the last word in a sentence will show elevated fixation durations because it is the point in the construction where the sentence can be comprehended as a whole (Just & Carpenter, 1980). Therefore, it is standard practice not to place the region of interest at sentence-final position. Likewise, the first-word position of the sentence is a poor region for analysis as this region is skipped more frequently than other regions of the sentence.

In spite of the richness of information that can be obtained from eye-movement data, eye-movement records have been used less extensively in the study of L2 sentence parsing for a number of reasons (a notable exception is the work by Frenck-Mestre and colleagues). For one, eye-tracking equipment is very costly to purchase and to maintain, and can be technically demanding. In contrast, self-paced reading studies are easy to implement and relatively inexpensive. Virtually any experiment can be set up on a standard desktop or laptop computer, and experiment-generated software is available for different platforms at a modest cost. In addition, many of the signature results found with eye-tracking measures have been obtained using self-paced reading (Mitchell, 2004).

7.4.3 Event-related potentials (ERPs)

As noted previously, ERPs are small voltage changes measured at the surface of the scalp, which reflect brain activity that is triggered by sensory stimuli or cognitive processes. An ERP consists of positive and negative voltage peaks, referred to as *components*. In ERP studies, participants listen to or read text while electro-encephalographic recordings are taken from different positions on the scalp. With this methodology, changes in ambient conditions such as lighting are kept at a minimum, and blinks are discouraged as the resulting waveforms can obscure the time course of linguistic processing. By varying information-processing requirements through the use of different tasks, qualitatively different ERP patterns have been found to correlate with particular aspects of language processing. For instance, Kutas and Hillyard (1980) demonstrated that sentences ending in a word that could not be semantically integrated into the prior sentence context ("He spread the warm bread with socks") elicited a negative-going waveform peaking at around 400 ms after the onset of the presentation of the critical word; therefore difficulty with semantic integration is associated with an *N400*-component. A second component, the *P600*, is a positive waveform with an onset at about 500 ms, which has been correlated with syntactic anomalies of various types (Osterhout & Holcomb, 1993).

One particular strength of ERP methodology over other techniques that are based exclusively on reading is that it allows a natural way of studying how linguistic material is processed when it is presented in an auditory modality (Mitchell, 2004). In this respect, ERP measures have been used successfully in L2 sentence-processing studies to determine whether the specific semantic and syntactic subprocesses engaged during L2 language comprehension are different for second language speakers as compared to native speakers. For example, Hahne (2001) compared semantic and syntactic processing in proficient second language learners of German who are native Russian speakers. ERP responses to auditory stimuli containing semantic and syntactic anomalies were recorded. Similarly to previous findings (Weber-Fox & Neville, 1996), the differences in processing semantic incongruities between native and L2 speakers were only quantitative, but there were qualitative differences with regard to syntactic processing between the two groups, suggesting that the L2 learners did not process or integrate syntactic information into the existing phrase structure in the same way as native listeners. In contrast to the reading studies described above, which show that structural processing of sentences in one language are affected by the presence of the other language, the ERP evidence suggests constraints in the degree to which the syntax of the L2 can be processed in a native-like manner (see MacWhinney, 1997, for another view of cross-language interactions in sentence processing, and Tokowicz & MacWhinney, 2005, for evidence that the ERP record may provide a sensitive means to detect the formation of syntactic representations in the L2 during early stages of acquisition).

7.5 Summary

In this chapter we introduced a subset of the laboratory methods that have been used to investigate the way in which bilinguals and second language learners recognize words, understand and produce speech, and process sentences in each of their languages. As noted earlier, our review is hardly exhaustive, but we have attempted to illustrate the methods that are representative of experimental approaches to bilingualism. In the process of doing so, we hope to have shown how these tools can be used to infer the nature of the cognitive processes that bilinguals bring to the tasks of comprehension and production in their two languages. We have also tried to provide a glimpse into the theoretical debates that guide this research. A list of laboratory designs without the theoretical foundation would be misleading, because it is these questions about how the mind accommodates the presence of two languages that lead us to the methods that we use. We invite the reader to sample the primary literature on experimental approaches to bilingualism. We also append below a section on resources that may provide useful information for laboratory investigations of bilingualism. We believe that this approach will inform not only theories of the bilingual mind but also cognitive and language science more generally.

Acknowledgments

The writing of this chapter was supported in part by NSF Grant BCS-0418071 and NIH Grant MH62479 to Judith F. Kroll and by NIH Grant HD50629 to Paola E. Dussias. We thank Natasha Tokowicz for helpful comments on an earlier version of the chapter.

Further reading and resources

There are a wealth of tools available to students new to experimental laboratory research. We list below a number of programs that are commonly used by psycholinguists to implement the sorts of experimental paradigms we have reviewed in this chapter. We also provide information on databases that may be useful in generating experimental materials. Students interested in pursuing laboratory research are well advised to take courses in experimental design and statistics. There are many introductory texts on each of these topics. The resources listed below are intended to supplement a basic introduction to research design and statistical methods. Although some of the techniques reviewed in the chapter (e.g., eye tracking and acoustic analysis) require additional training that cannot be easily accomplished without immersion in a laboratory setting, others (e.g., lexical decision and picture naming) can be sampled in web-based experiment programs that are readily available.

Programs for experimentation and analysis

Boersma, P. and D. Weenink (2005) Praat: Doing phonetics by computer (Version 4.3.22) (computer program). www.praat.org/.

Cohen, J. D., B. MacWhinney, M. Flatt, and J. Provost (1993) PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25, 257–71.

Forster, K. I. and J. C. Forster (1999) *DMDX* (computer software). Tucson: University of Arizona.

PST (Psychology Software Tools, Inc.). *E-prime*. www/pstnet.com/.

Websites for on-line experimentation

There are a number of websites where you can participate in actual experiments or try out demonstrations of psycholinguistic phenomena. Here are a few of those sites:

psych.hanover.edu/research/exponnet.html
psychexps.olemiss.edu/
www.york.ac.uk/res/prg/

Useful databases for psycholinguistic research

Note: The Psychonomic Society has recently established an archive that contains many useful databases: psychonomic.org/archive/; and the Max Planck Institute for Psycholinguistics in the Netherlands maintains a database of relevant corpora: www.mpi.nl/world/corpus/index.html/.

Balota, D. A., M. J. Cortese, K. A. Hutchinson, et al. (2002) The English lexicon project: A web-based repository of descriptive and behavioral measures for 40,481 English words and non-words. Available at elexicon.wustl.edu/.

Buchanan, L. and C. Westbury (2000) Wordmine Database: Probabilistic values for all four to seven letter words in the English language. Available at www.wordmine.org/.

Davis, C. J. and M. Perea (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods, Instruments, and Computers*, 37, 665–71.
Prado, M. (1993) *Spanish False Cognates*. Chicago, IL: NTC Publishing Group.
Sebastián-Gallés, N., M. Martí Antonín, and F. Cuetos Vega (2000) *Léxico informatizado del Español*. Barcelona: Edicions de la Universitat de Barcelona.
Snodgrass, J. G. and M. Vanderwart (1980) A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 174–215.
Tokowicz, N., J. F. Kroll, A. M. B. De Groot, and J. G. Van Hell (2002) Number-of-translation norms for Dutch-English translation pairs: A new tool for examining language production. *Behavior Research Methods, Instruments, and Computers*, 34, 435–51.

A guide to writing experimental reports

Publication Manual of the American Psychological Association (2001). 5th edn. Washington, DC: American Psychological Association.