Psycholinguistics

Dominiek Sandra
University of Antwerp

Anyone who is interested in finding out what the state of the art is in a particular science wants the answer to a limited set of questions. What are the major goals of this discipline? Which theoretical models are generally accepted on the basis of the available research data or are, at least, at the top of researchers’ minds when thinking about their discipline? At a less general level, which are the most pressing questions on the current research agenda, those where the majority of scientists turn their attention to? Finally, how do researchers in this discipline collect their data, i.e., which methods and techniques do they have at their disposal for studying the phenomena they are interested in?

For instance, people who like to learn more about physics would be interested in the general insights physicists hope to achieve. They would also want to be informed on the theoretical models that current-day researchers adopt on the structure of matter, the universe, etc. Next, they would like to be informed on the questions that matter most to current-day physicists and, hence, determine the recurring themes at conferences and in scientific journals. And, finally, they would be interested in finding out which methodologies and techniques researchers use to collect the data their theories are based on.

The goal of the present chapter is an attempt to answer these questions for the field of psycholinguistics. However, before tackling each of these questions in turn, I will backtrack a few decades and give a short sketch of how psycholinguistics became a research domain of its own, at the intersection between linguistics and experimental psychology (Section 1). After that, I will briefly describe the general goals that psycholinguists try to achieve (Section 2). In Section 3, I will give an overview of the dominant models that flourish in the field and form a theoretical divide between researchers. Next, I will offer a description of the general methodological paradigms that researchers rely on and their underlying rationale (Section 4). Then, I will focus on the research techniques that are most frequently used in the methodological arena of experimentation (Section 5). Taken together, these sections offer a broad characterization of the field and form the first part of this chapter.

The second part consists of what I hope to be representative reviews of the specific psycholinguistic research that has been performed with respect to each of the four language skills. There will be a section on written language processing (Section 6), focusing on both perception and production aspects, more particularly, the recognition process of written words (Section 6.1) and the spelling of these words (Section 6.2). This will be followed by a section on spoken language processing (Section 7), where a distinction will also be made between the perception and production aspects of speech (Sections 7.1 and 7.2, respectively). The order in which these four topics are treated does not in any way reflect the relative importance of these four research streams within the discipline, but merely results from my own activity in the field. As I have always been studying the processes of reading and spelling and, hence, am only familiar with the topics in speech recognition and production through journals and conference lectures, I feel more confident starting with written language processing. In the discussion of each of the four research domains, I will focus each time on the important theoretical issues, the techniques that have been used to approach them, and the current answers to these questions or, alternatively, the controversies they have given rise to.

1. The birth, adolescence, and adulthood of psycholinguistics

Let’s start with a very general characterization of the field: the purpose of psycholinguistics is to achieve insight into the mental infrastructure that makes language use possible. The fact that psychologists ultimately want to find out what makes language use possible is obviously reminiscent of Chomsky’s credo (already half a century ago) that the human mind is designed to use language. If we had no species-specific language acquisition device (called a language module by Fodor, 1983) children could never internalize the grammar that is implicitly encoded in the sentences they hear in their environment. Thus considered, psycholinguistics is part of the large Chomskyan legacy, as so many other subdisciplines on the linguistic territory.

Chomsky’s impact on the current linguistic landscape can hardly be overestimated, irrespective of the correctness of either his core proposal that humans are endowed with a mental grammatical template guiding their language acquisition, or his formal methodology for studying language from this perspective. Not only did his view result in a radical paradigm shift in theoretical linguistics and a formal model of language that underwent a series of revisions over time, it also spawned a boost in studies of formal languages (Levy, 1974), an active interest in the search for language universals (Greenberg, 1963), the study of language acquisition (Bates, 1974; MacWhinney, 1975; Snow & Ferguson, 1977), and so much more. However, within linguistics, itself Chomsky’s theorizing not only introduced a novel paradigm for doing research in theoretical linguistics, it also influenced the field by fueling fierce controversies about the real nature of language. In the wake of these discussions, full-blown novel approaches to language saw the light, such as sociolinguistics (Labov, 2006, 2nd ed.), pragmatics (Grice, 1968; Brown & Levinson, 1978), cognitive linguistics (Langacker, 1987) and psycholinguistics (Fodor, Beiser, & Garrett, 1974). Hence the above claim that the major goal of psycholinguistics was initially inspired by the Chomskyan mentalist approach to language.
However, the first excitement about Chomsky's belief that language provides a window on the human mind, thus allowing linguists to draw the map of the language module in our cognitive system, and especially the initial enthusiasm that some of his proposals might also be relevant for language processing soon dampened. The well-known psychologist George Miller (1962) had been quite positive about the future perspectives when he wrote in an issue of *American Psychologist*: "I believe that one of the best ways to study a human mind is by studying the verbal systems that it uses. Such a program is not only important, but immediately possible." He had done some experimental work along these lines as well (Miller & Mckeen 1964). However, when a number of experimentalists made it their goal to test the so-called *psychological reality* of a number of key concepts in Chomsky's generative grammar, this time of great expectations came to an end within ten years after the publication of Chomsky's ground-breaking book *Aspects of the Theory of Syntax* (1965).

The pioneers in this endeavour, Merrill Garrett, Jerry Fodor, and Thomas Bever carried out a number of important experiments, trying to find out how well the mental processes involved in language use mapped onto the theoretical operations in generative linguistic theory. They did not. In their co-authored classic book (Fodor et al. 1974) their central message was that the major formal operations that are performed on linguistic representations in Chomsky's framework (transformations) did not correspond to the processes that language users mobilize in sentence processing. For instance, sentences that required a longer chain of formal transformations between deep and surface structure in generative grammar at the time did not necessarily take more processing time in behavioural experiments. Hence, George Miller's theory of derivational complexity, postulating such a correspondence, was proven wrong. These and other experimental findings indicated that Chomsky's theoretical notions were merely formal tools for describing, in a technical language, the implicit knowledge structures of language users (e.g., knowledge about sentence complexity), without any commitment to a psychological theory of representation and processing. Hence, they could not be treated as concepts that are isomorphic to language users' actual mental structures or processes in language use. As a matter of fact, Chomsky (1965) himself had warned against such a misinterpretation of his theory:

"When we speak of a grammar generating a sentence with a certain structural description, we mean simply that the grammar assigns this structural description to the sentence. When we say that a sentence has a certain derivation with respect to a particular generative grammar, we say nothing about how the speaker or hearer might proceed, in some practical or efficient manner, to construct such a derivation. These questions belong to the theory of language use – to the theory of performance." (p. 9, my emphasis)

Note that it is ironic to see history repeating itself, albeit with a strange twist. Exactly the reverse error regarding the psychological reality of linguistic analyses was made by (some) theoretical linguists working in the domain of cognitive linguistics. This time some of these linguists themselves rather than experimental psychologists strongly suggested that the outcome of their armchair semantic analyses of a word's polysemy offered a picture of distinctions made in the human mind (for a methodological critique on this line of reasoning, see Sandra and Rice 1995; Sandra 1998).

At any rate, even though early psycholinguistics had its roots in generative linguistics, it soon reached adolescence and became more and more an independent discipline, defining its own research questions. Importantly, this does not mean that all distinctions made in linguistics were suddenly thrown away. In the psycholinguistic study of both lexical and sentence processing, theoretical concepts like morphemes and phrase structures are still used. As a matter of fact, it is hard to imagine how one could study language processing without making reference to the major language elements that have been identified by linguists at all. How could one formulate hypotheses about the role of the elementary building blocks of words and sentences if one cannot make use of the terminology developed by linguists?

This, of course, does not necessarily entail that all psycholinguists work on the assumption that each linguistic distinction is somehow reified in the mind, as an object of representation that plays a role at some level of language processing. Some researchers use the linguistic terminology to make clear what they are talking about, but do not make the additional assumption that each linguistic concept necessarily has a representational status in the language user's mind. As a matter of fact, this is the conviction of many connectionists (see Section 3), who often define a linguistic construct as an emergent property of a processing system that discovers regularities and subregularities in the language data on the basis of its capacity to 'perceive' correlations between several kinds of elements. Notions like 'regularities' (a different concept than 'rules') and 'correlations' indicate that these are statistical approaches to language and its 'units', which sharply deviate from a linguistic perspective. Therefore, rejected the idea of morphemes as independent linguistic objects, defining them as emergent properties instead. In his view, 'morphemes' (in written words) merely reflect the fact that the two letters that cross a morpheme boundary typically have a lower co-occurrence frequency than the pairs of letters within morphemes, which causes emergent patterns in the data. In his later convergence theory, morphemes are described as graded phenomena resulting from correlations among orthography, phonology and semantics, i.e., which again betrays a statistical approach.

Despite the fact that many psycholinguists still make frequent use of linguistic terminology and that quite a few also believe that there are mental correlates for (some) linguistic distinctions, one cannot deny that there has been a divorce between the two disciplines. This divorce resulted from the early, naïve assumption that psycholinguistics must test the psychological reality of theory-dependent notions, i.e., a tool for validating or falsifying a linguistic theory. Current-day psycholinguists work on the assumption that it is not their task to test whether a theoretical linguistic framework that is the fashion of the day can be translated into claims about the way in which words and sentences are mentally represented and/or processed. One may wonder whether such a goal is even feasible in principle.
as each group of researchers targets quite different goals. Even though linguists studying language may consider language as a window on the human mind, either because they think language is a separate faculty in the human cognitive system or because they believe that language and basic cognitive principles are deeply interconnected, it is not their aim, nor could it be (by definition), to design a theory on how their theoretical distinctions are implemented in the mind, neither in terms of representations nor in terms of processes.

2. Major goals

The preceding paragraph already contained some hints with respect to this issue. We saw that, at a very broad level, psycholinguistics can be defined as the study of what is going on in the human mind during language use. Obviously, this is such a general statement that it is barely informative. Hence, a definition at a much more specific level is required, which I propose to be the following: psycholinguistics is the discipline studying the mental structures and processes that language users rely on (i) when they are confronted with a particular type of language material (e.g., reduced relatives, particular word types), (ii) while being engaged in one of the four forms of language use: speaking, listening, reading, or writing, and (iii) while being in one of the three possible states of language knowledge: a state of language acquisition, the mature state of the experienced language user, or the state of language malfunction or desintegration (the latter being the result of either an innate deficit, e.g., developmental dyslexia, or an acquired brain trauma, e.g., aphasia). Iets make this specific definition more comprehensive by breaking it down into its components.

According to the above definition the discipline can be conceptualized as the collection of cells in a 2 by N by 4 by 3 four-dimensional structure, whose dimensions correspond to the type of mental 'object' in focus (representations vs processes), the type of language phenomenon under study (e.g., lexical vs syntactic structures; there are many, hence, N), the type of language use (the four language skills) and the language user's stage of language proficiency (acquisition, mature, malfunction/desintegration). This gives rise to 24'N different lines of research, making psycholinguistics a very broad enterprise. This picture can be somewhat simplified by leaving out the factor of language proficiency. The study of how children acquire their native language and the study of how a form of language use is affected when part of its neural substrate malfunctions can certainly shed light on properties of the mental infrastructure underlying experienced language use, more particularly, by identifying the nature of the process/representation that eventually evolves into a mature state or by investigating the nature of the process/representation that desintegrates. However, the term 'psycholinguistics' is generally applied to the study of the mental apparatus that supports language use in its mature state. Researchers of language acquisition and language disabilities each form their own community, with its own scientific journals and conferences and with its own study methods, even though they fall under the umbrella name of psycholinguistics.

This leaves us with three defining factors: type of mental 'object', type of language phenomenon, and type of language use. The latter two factors need no comment, so that psycholinguistics turns out to be the study of the mental representations and processes that are involved when one uses a particular language phenomenon while being engaged in a particular language skill. Hence, the only remaining question is: what exactly is meant by a mental representation and a mental process?

The meaning of the term 'representation' can perhaps best be explained by morphologically decomposing the word into its stem and prefix: re-presentation. Literally, a representation presents something again, in a different form. Taking examples from daily life, one could say that a photograph visually represents a person or scene at one moment in the past. A particularly useful example is music, because music can be represented in a variety of ways. The sequence of musical notes on paper visually represents the sounds of a song in the form of a representational system of visual symbols. The recording of that song in the form of tiny horizontal deviations in the small spiral-formed groove of an old vinyl plate is a representation of that song in a physical code system, which can be mechanically transformed into the audio signal. The recording of the song on an audio CD involves two levels of representation: the deepest level is the physical encoding, in the form of a sequence of pits and 'lands' in the long, extremely small spiral track, a representation that is subsequently transformed into another representational format, that of a digital code. The digital information is not encoded in a simple one-to-one fashion, such that a pit is a 0 and a land is a 1 (or vice versa). Rather a change between a pit and a land or between a land and a pit represents a 1, whereas no change represents a 0. The latter code is finally transformed into the audio signal of the music.

Mental representations (of language in this case) are something similar: they represent the information that has to be stored both in a particular code (comparable to the pits and lands on a CD) and in a particular medium (comparable to the polycarbonate plastic of a CD). Their medium, of course, is formed by the neural structures of the brain. The representational code used at this deepest level of representation is unknown. Whereas people invented the code made up of pits and lands on CDs and, hence, understand it (at least, the engineers do!), the neural code that is used to represent language is the result of biological evolution and is far from understood yet. Hence, psycholinguists' hypotheses and claims with respect to the representation of words and syntactic patterns make no reference to the encoding of language in the neural hardware of the brain (not even in studies where brain imaging techniques are used).

To continue the analogy with the representation of music on a CD, any psycholinguistic hypothesis or theory pertains to a representational level that is comparable to the digital representation into which the physical representation on a CD is transformed, a level intermediate between what can be perceived in the outside world, i.e., the physical signal that must be encoded, and its actual encoding in a physical carrier. For instance, the mental representation of a word is situated between its physical representation in brain activity and its physical representation in a recording on an audio CD.
the world (comparable to the physical signal of the song). From this perspective, mental representations have no real existence, as they refer neither to the physical aspects of the linguistic signal nor to its neural representation. They are like representations in a virtual world in-between these two tangible worlds. However, intangible as they are, they are necessary scientific constructs if one wants a vocabulary for talking about language representation at all. By forming a virtual interface between the physical realization of language and its encoding in the brain, mental representations make it possible to discuss representational issues in terms of ordinary words, for instance, by making reference to common linguistic distinctions like phoneme, morpheme, semantic relationship, noun phrase, etc. In order to discover constraints on representations (ultimately the neural ones) one will have to define them in a terminology that belongs to the reality that is being represented and is, hence, understandable.

Thus considered, mental representations are like icons on a modern computer: they form an interface between the world that everybody understands and can describe with ordinary words (e.g., a desktop, a map, a file) and the physical world of the machine that only computer specialists understand. The icons enable computer users to think in everyday terms about the way their computer works without having the slightest idea about the processes of physical memory retrieval and processing on machine codes that are triggered by double-clicking an icon (or feeding the need to understand these technicalities).

Similarly, psycholinguists can use mental representations when constructing models of a mental activity, like language processing, without having to understand the physical reality at a deeper representational level.

The virtual nature of mental representations does not in the least devalue the psycholinguistic enterprise, as the progress that has been made over the past four decades has shown. On the contrary, being framed in the vocabulary of ordinary language, these representations make it possible to formulate hypotheses in terms of concepts that researchers can understand. Thus mental representations drive research that makes it possible to identify constraints on language processing. Once a constraint is identified, each level of representation will have to observe it, as the representation at one representational level is mapped onto a different kind of representation (i.e., it is recoded), while preserving the information it encodes. For instance, when participants’ behavior in an experiment (e.g., response speed in some task) reveals that they are sensitive to morphological structure when reading words, this finding indicates that this structure plays a role at a particular level of mental representation and, ultimately, also at the level of neural representation to which the mental representation provides an intelligible interface.

Given the above definition of a mental representation, it is easy to define the meaning of the term ‘mental process’. Mental processes are operations on mental representations. Ultimately, these operations can also be translated into brain processes, although the remark with respect to mental representations can also be made with respect to mental processes. When the processes are described with respect to the virtual in-between level mentioned above, they will be intelligible, as they will make reference to words in our familiar language. Many psycholinguists conceive of a mental process as a procedure that maps one type of mental representation onto a different one.

Thus the above statement that the general purpose of psycholinguistics is an attempt to characterize the mental infrastructure behind language use can now be formulated in a much more transparent terminology. It is the attempt to discover the sequence of mental representations and mental processes operating on these representations when a language user is processing a particular language phenomenon using a particular language skill (e.g., speech perception). For instance, which mental representations and processes mapping one representation onto the other are involved when a reader recognizes a morphologically complex word or a word that has the same form but a different meaning in two languages (e.g., English room, which has the same form in Dutch, where its meaning is ‘cream’)? Which mental representations and processes are mobilized when a reader reads a sentence like The horse raced past the barn fell (where fell is the verb of the main clause and raced the verb of a reduced relative)?

3. Major theoretical models

Psycholinguistic models used to be of the box-and-arrow type, which means that researchers derived from their data a number of plausible processing stages, whose names they used as labels for the boxes in their model, and drew arrows between these boxes, thus indicating that information was transmitted from one box to the other. Such models are obviously nothing more than graphical representations of the representations and processes that are described by the researcher. For that reason they are also called ‘verbal models’.

Since the time of easy access to personal computers and their high calculating power, many verbal models have been replaced by computational ones. Indeed, a model is much more powerful if it can simulate experimental data on human language behaviour. When attempting to achieve that goal one is forced to make all one’s theoretical assumptions explicit, even those that one has perhaps not thought about but are necessary to build an operational model. Only then, it is possible to translate the verbal model into a computer programme, which can then apply the processes it has been programmed to execute to its input and generate an output. The degree of match between the model’s output and the human data is an index of the model’s success.

Working from such a perspective can only be beneficial for the advancement of the scientific discipline. As a matter of fact, translating one’s theory into a testable model is the normal state of affairs in many sciences. Meteorological models, for instance, are considered successful to the extent that they can predict the weather for the coming days relatively well on the basis of an algorithm whose critical parameters (temperature, air pressure, etc.) are continually updated on the basis of information from weather stations around the globe.
Models on language processing differ along three dimensions, each of which can be formulated as a question: (i) What is the nature of the processes that retrieve stored information? (ii) What is the nature of the mental representations themselves? (iii) Does the model leave room for abstract rules, which are by necessity symbolic, as rules make reference to abstract linguistic categories (e.g., *Christen-ed to V* for regular past tense formation of verbs)? I will discuss issues (i) and (ii) together, for the simple reason that assumptions on processing and representation often go hand in hand.

3.1 The nature of mental processes and representations

Models differ considerably in the nature of the processes that are responsible for the retrieval of stored information and in the nature of their storage principles. I will illustrate this by taking models of visual word recognition as a paradigmatic example.

3.1.1 The early models

In this domain, two dominant verbal models of lexical access were developed around the seventies of the twentieth century, each based on a particular metaphor for information retrieval. Chronologically, the first model was John Morton’s (1969) logogen model, whose basic units are so-called logogens (etymologically, the term ‘logogen’ is derived from the Greek words logos and genos and, hence, means something like the birth of a word). Logogens form an interface between the physical word (e.g., a printed word) and all linguistically relevant properties of the word they represent. Basically, a logogen is a recognition device that collects knowledge about the presence of ‘its’ word in the outside world. Its mode of operation is based on the way a neuron works. A neuron, the basic processing unit in the neocortex of the brain, receives small electrical impulses from other neurons, which raise its state of activation. When a particular activation threshold has been reached, the neuron responds by generating an output of electrical activity itself (it is said to ‘fire’), which serves as the input to other neurons. The technical properties of a logogen largely correspond to this: (i) it has an activation level, which is directly proportional to the match between the word it represents and the word that the reader is fixating and (ii) it has a recognition threshold, which is the level at which sufficient activation has been collected for the logogen to ‘fire’. At that moment it has recognized the word it represents and makes all linguistically relevant information (meaning, pronunciation, etc.) available to the language processor. A logogen’s threshold level of activation is a direct function of the word’s frequency, such that words that are frequently used have a lower threshold, i.e., are recognized sooner, than lower-frequency words. Thus the model explains one of the most stable effects in the word recognition literature: the frequency effect (see below).

The other model is Ken Forster’s (1976) search model. Whereas Morton’s model takes neuronal functioning as a metaphor for lexical access, Forster’s model is inspired by an entirely different sort of metaphor: that of a sequential search through a database. Forster makes the comparison with the way a book is retrieved from the correct shelf in a library. First (at the time of the model’s conception) one searches the reference to the book’s physical location in a collection of index cards, each containing the bibliographical description of a book and a code referring to the book’s location in the library. These cards are alphabetically ordered to make the search process as efficient as possible. Then, one uses the reference number on the index card to physically locate the book and retrieve it from its shelf. This two-step process also occurs in Forster’s search model of lexical retrieval. The model distinguishes between two types of files: access files and a master file. The former contain modality-specific representations for each word (orthographic ones for visual word recognition, phonological ones for auditory word recognition, syntactic–semantic ones for word production) and a pointer to an address in the master file. This master file contains the proper lexical representations, which store all lexically relevant information for each word (its pronunciation, spelling pattern, syntactic properties, meaning). For instance, the recognition of a written word is supposed to proceed as follows: First, the word’s orthographic pattern is looked up in the orthographic access file, which is searched in a sequential fashion until a match is found. This search process is efficient because it implies an ordered search, not on the basis of alphabetical order (as in a library) but on the basis of descending frequency, such that the access codes for high-frequency words appear earlier in the search sequence than those for low-frequency words. Thus the search model accounts for the word-frequency effect as well. Once the word’s access code has been found, a reference to the appropriate address in the master file also becomes available. There the processor finally retrieves all lexical properties of the word.

3.1.2 Interactive-activation models

The basic process behind the logogen model, activation, has survived in many later and current-day models, although the architecture of these models has become much more elaborate and has been made so explicit that they can be transformed into computational models. The activation concept lies at the heart of McClelland and Rumelhart’s (1981) and Rumelhart and McClelland’s (1982) interactive activation (IA) model, the first computational model for written word recognition that was explicitly designed to simulate experimental findings and, hence, to put the validity of its architectural distinctions to the empirical test.

In this model, the input of a written word starts activating information at the lowest level of representation, the representation of letter features, which pass on their activation to all representations of letters containing these features. Once a letter representation has accumulated so much activation that its threshold is exceeded, it activates all word representations containing that letter in that specific position (e.g., *--- will activate lake, tall, task but not mist, step or rust*). Finally, at the word level, the same principle applies: each lexical representation accumulates activation that it receives from letters appearing in the appropriate word position until the recognition threshold of one representation is exceeded and the word is recognized.
Two additional remarks on activation must be made. The first concerns the nature of the effect that activation brings about. That effect can be both excitatory or inhibitory. The former notion refers to the fact that activation brings the state of the representation closer to its threshold, whereas the latter refers to the opposite: adding activation has the effect of bringing the state of the representation closer to its activation baseline. Note that these two concepts are also borrowed from neuroscience: some neurons make the neurons they project to more active, whereas others inhibit other neurons’ activity level. Whereas the notion of inhibition might sound strange at first, it is easy to make intuitive sense of it: if, for instance, we were not able to disregard or ignore e.g., inhibit much of the incoming sensory data, our processing system could not make any selection between relevant and irrelevant information.

In models of the IA type, representations between levels are connected either by excitatory or inhibitory links, depending on whether the information represented at the lower level is part of the information represented at the higher level or not. For instance, an active representation of the letter f---in word-initial position of four-letter words will increase the activation level in the word representations for fake, feel, fill, firm, feel, etc. but reduce the activation level in representations for words like soft and wolf (non-initial f), or cake, bear, lock, (no f), etc. As far as connections between representations at the same representational level are concerned, the effect of activation is inhibitory (known as lateral inhibition). This makes intuitive sense, as one cannot simultaneously recognize two different letters in word-initial position or two different words in the same letter string. The result of lateral inhibition is that all word representations that are activated to some extent (as the result of sharing letters in the same position) will exert an inhibitory effect on each other’s activation level. However, as this inhibitory effect is proportional to their own activation level, the representation that reaches the highest level of activation on the basis of the letter-to-word connections will ultimately be the only one to remain active and thus win this ‘competition’.

A second property of the activation concept in IA models concerns the directionality of the activation stream, which is bi-directional. This is a defining characteristic of these models, as they have been named after it: interactive activation models. Hence, in IA models, activation is not only sent forward, i.e. to a representation at a later processing level, but also flows back to the representation it came from, more particularly, through a feedback connection. All bottom-up connections are matched by top-down connections between the same representations. For instance, letter representations activate word representations, which feed their activation back to the letter representations they have been activated by (inhibiting others at the same time), thus further increasing their own activity level. This interactivity between levels made it possible to explain why people recognize the same letters better in words than in nonwords or random letter sequences when the visual stimulus is barely visible (extremely short presentation, immediately overwritten by other visual stimulus), the so-called word superiority effect. At any rate, the result of the interactive activation process is that activation loops emerge in the model, which serve to quickly filter out the lexical representation that matches the input, in cooperation with the process of lateral inhibition. Once the model has reached this stable state, the word has been recognized.

The IA architecture has been the inspiration for several specific models in the domain of visual word recognition, both in the study of the mental lexicon of monolinguals (Grainger & Jacobs 1996) and bilinguals (Dijkstra & Van Heuven 1998, 2002), in the domain of auditory word recognition (Elman & McClelland 1984; McClelland & Elman 1986), and in the area of speech production (Dell 1986).

3.1.3 Connectionist models

Connectionist networks are quite similar in their mode of operation to IA models. As a matter of fact, it is the logic behind IA models that has caused the emergence of connectionism, only a few years after the two seminal papers by McClelland and Rumelhart. Moreover, the very same researchers were behind the two equally seminal books on what they called parallel distributed processing (PDP), but which undoubtedly marked the emergence of connectionism (McClelland, Rumelhart, and the PDP Research Group 1986, Rumelhart, McClelland, and the PDP Research Group 1986). Connectionism has become highly influential in current-day psycholinguistic modeling. In contrast to IA models, connectionist models have no built-in functional architecture (like letters feeding input to words). Their main strength, according to the proponents, is their ability to discover the systematicity between two types of representations, for instance orthography and phonology, and to internaly represent it without the necessity to appeal to linguistic concepts and rules (see below).

The basic architecture of a connectionist model is a series of three layers, each consisting of a set of nodes: an input layer, an output layer, and a so-called hidden layer, which is situated in between the previous two. Both the input and output layers are fully connected to the hidden layer, which means that each node in these layers is connected to each node in the hidden layer. Each connection is associated with a weight determining how much of the activation that is transmitted by the sending node is actually received by the receiving node. The hidden layer is required because the systematicity in mapping representations at the input layer to representations at the output layer is generally not of the one-to-one type (if the latter were the case, a set of direct connections between two layers would suffice). The lack of one-to-one mappings is typical for language. The relationship between orthography and phonology represents a prototypical instance of one-to-many mappings (e.g., in English the letter e sounds differently in words like the, bed, care, effect, hypothesis, etc.).

The two essential processes in connectionist models are activation and backpropagation. We are already familiar with the notion of activation. Backpropagation is a supervised learning technique that makes it possible for the connectionist system to learn the regularities that are implicit in the training pairs. Before a connectionist network can function on its own, it must first go through a training/learning phase, in which it is presented with a series of training examples (e.g., a collection of word pairs such as map/marker, wrap/writing, etc.). During this phase, the network learns to associate an input pattern (e.g., the word “map”) with an output pattern (e.g., the word “marker”). The network then adjusts the weights of the connections between the input and output nodes to minimize the difference between the actual output and the desired output. During the backpropagation phase, the network takes a previously unseen input pattern and uses the weights of the connections to generate an output pattern. If there is a difference between the actual output and the desired output, the network adjusts the weights of the connections again to correct the error. This process continues until the network is able to correctly associate an input pattern with the desired output pattern.
a long list of input-output pairs, i.e., examples of correspondences. In the initial internal state of the network all connection weights between nodes are set at a random value. The input will pass through the network as a function of these weights, which will determine the representation that is formed at the output layer. As nothing has been learned yet this output will certainly be wrong. At that point the concepts of 'backpropagation' and 'supervised learning' come in. The idea is conceptually simple: for each node in the output layer a measure of divergence between the observed and the desired output is calculated (the error) and each of these errors is used to adjust the settings of the weights between layers. Technically, this is a backwards propagation of the error, which explains why the technique is called 'backpropagation'. These adjustments make it possible that the model performs slightly better the next time, i.e., that it learns. This kind of learning is supervised because there is someone who provides the model with the correct response and, hence, makes the process of backpropagation possible.

The above-mentioned process of generating very wrong to partially wrong outputs and adjusting the connection weights will continue for a long time. However, at the end of a long training session the system will be able to produce the correct output for each input it has been given in the training set. Moreover, it will be able to generalize this knowledge, i.e., to correctly apply it to instances that it has never encountered before. What has happened? What certainly has not happened is that the system has learned paired associations. Rather it has adjusted its weights for a large number of connections in such a way that, collectively, these weight settings implicitly represent the systematicity that is implicit in the input-output pairings. Nowhere in the system can one discover what the system exactly represents. There are no nodes for words or letters or whatever linguistic units. The nodes, connections, and numerical weights come linguistically unlabelled. And even though the system behaves as if it has discovered the rules for mapping an input onto an output, it has only induced the inherent regularities without explicitly representing rules. Illustrative examples of the connectionist paradigm are to be found in publications by Rumelhart and McClelland (1986), Seidenberg and McClelland (1989), Seidenberg and Elman (1999a) and Seidenberg, MacDonald, and Saffran (2003), although many others could be cited. The titles of the latter two papers “Networks are not hidden rules” and “Are there limits to statistical learning?” clearly reveal what the essence of the connectionist endeavour is: accounting for language behaviour on the basis of statistical learning rather than the learning of rules. Connectionist models represent the correlational structure behind input-output pairings, no rules.

In current-day psycholinguistics, IA and connectionist models are almost exclusively used to model participants’ results in real-time experiments — and quite successfully I must add. However, they are quite different types of modeling. Although both originate in the idea that activation is the central concept for mapping an input onto an output, their internal architecture is radically different. IA models make use of an architecture in which the basic representational levels are identified (e.g., letter features, letters, words) revealing the hypothesis of the model’s designer about the types of representations that are functionally relevant for the task. Connectionist models rely on an architecture that makes it impossible to discover this functional model in the designer’s mind. As a matter of fact, the designer does not have such a functional model in mind, first of all because all knowledge is distributed throughout the model, and second because the researcher’s only concern is to make a language process work without labeling what it does to what kind of representation and without taking existing linguistic concepts for granted (as these may only exist in the linguist’s analytical mind).

Note that both model types can represent the same kind of knowledge (e.g., how graphemes are mapped onto phonemes). However, IA models will reveal how they accomplish this job whereas connectionist models will not. In line with this architectural difference one talks about localist representations (IA networks) and distributed representations (connectionist networks). A distributed representation is literally distributed across the set of nodes, their interconnections, and the weights in the system. Some researchers explicitly opt for an IA model because they consider it their primary goal to discover the internal structure of the language processing device that they are studying, i.e., identify its functional components and label them, rather than finding a solution that solves a mapping problem (e.g., between orthography and phonology) but in their view is uninformative at the theoretical level (a set of weights). For instance, Coltheart et al. (2001) write: "We are adherents of Old Cognitivism, and so our main interest is in the internal structure—the functional architecture—of human cognitive systems [...] our view is that the past quarter of a century of empirical and theoretical research on reading has provided us with good reasons for proposing a particular architecture for the human reading system, and our preference is to rely on this body of literature, rather than on backpropagation, for ideas about what this architecture might be.”

3.1.4 Exemplar models

A final class of models that has potential use for psycholinguistics but is (too) seldom appealed to for addressing psycholinguistic issues is the class of exemplar-based learning models. Several such models exist: Skousen’s Analogical Model (Skousen 1989; Skousen, Lonsdale, & Parkinson 2002), Nosofsky’s Generalized Context Model (Nosofsky 1988), and the 1MBL model by Daelemans and Van Den Bosch (2005). They have all been developed with the goal in mind: the classification of a novel input on the basis of its similarity to a set of stored exemplars. Like connectionist models they set out from a ‘blank slate’ and must go through a training phase. Both model types differ considerably, however, in what they do with the training exemplars. Connectionist models ‘forget’ all individual exemplars but induce the implicit regularities and sub-regularities and encode them in the weights between nodes, thus representing patterns of systematicity in the training data. In contrast, exemplar models do not forget individual exemplars but keep track of each one of them by storing them in a memory component. Nor do they encode the regularities in the training
data; rather the result of their learning process is the weight of each feature in the feature-encoded input for the purpose of classification (e.g., regular vs. irregular past tense). Despite these differences, each type of model can generalize what it has learned in the training phase to novel instances. However, as their learning product radically differs, they accomplish this generalization task in totally different ways. Connectionist models 'squeeze' the novel input through their network of weighted connections, producing an output at their output layer. Exemplar-based models use their feature weights for calculating a distance between the novel input and each stored exemplar in their memory component, determine the set of most similar exemplars, and perform the classification task on the basis of the dominant class in this nearest neighbours set. To accomplish this, they make use of a metric for determining similarity, a mechanism for determining the set of nearest neighbours, and a decision mechanism for categorizing the novel input as an instance of category A or B (e.g., "takes one type of plural or another"; see Keuleers, Sandra, Daelemans, Gillis, Durieux, & Martens 2007).

3.2 Rules or no rules: That's the question

The question whether rules in language have their mental counterpart in the human mind has been a lively topic of discussion over the past couple of decades. This has been especially the case in the area of inflectional morphology, where the question arises with respect to regularly inflected verbs. Linguistically, the morphological structure of regular verbs can easily be described by (simple) rules of the type "add -s to stem" or "add -ed to stem". From an intuitive perspective it is almost self-evident that language users store such simple rules in their processing system for language. On second thought, however, one must admit that rules are descriptive tools used by grammarians, which need not correspond to what language users mentally represent. Moreover, rules are by definition abstract, and it is unclear at what level of abstraction language users represent their knowledge of a language. So, there is a real empirical question here, touching one of the essential questions about the nature of human cognition: how abstract are mental representations?

This question inspired the psychologist Steven Pinker to write his book *Words and Rules: The Ingredients of Language* (1999), in which he argues that symbolic rules are required to account for the experimental data on regular and irregular past tense formation in English. In his view, regular forms are generated by rules whereas irregular forms are represented in an associative memory component, such that subregularities within the irregular domain, like the alternation between /i/ and /ai/ (sing-sang, ring-rang, drink-drunk, sink-sank but bring-brought, think-thought, link-linked), can be captured. Besides grammatical rules, there may also be rules that map one type of representation onto another, such as the sublexical correspondence rules that have been proposed for transforming a grapheme sequence into a phoneme sequence. The Dual Route Cascade model proposed by Coltheart, Rastle, Perry, Langdon, and Ziegler (2001) is a model that makes use of such rules.

However, not all models make use of rules. As has become clear in the paragraph on connectionist models, this model type can capture regularities in a set of training data and, hence, display rule-like behaviour, without actually representing rules that operate on linguistic categories (symbols). The only thing to be found inside their architecture are connection weights. There have been attempts from researchers with a rule-based view to counter the connectionist denial of rules by arguing that what is encoded in connectionist models are rules in disguise or hidden rules. For instance, according to Marcus (1999b), "Seidenberg and Elman have not gotten rid of the rule; they have simply hidden it." However, connectionist protagonists have rejected such critiques and argued that Marcus misinterprets his own experimental data: "Rather than showing that rule learning is 'there from the start' (4), the findings in Marcus et al.'s report indicate that infants are able to encode multiple types of statistical regularities" (Seidenberg & Elman 1999a, p. 433). Moreover, from the perspective of these connectionists, researchers like Marcus try to save the existence of rules in the mind by changing the definition of the word: "What has actually happened, as Marcus' comments illustrate, is that the concept of 'rule' is being altered to conform to the properties of connectionist networks" (Seidenberg & Elman 1999a, p. 288). Clearly, the question whether the language processing system actually represents rules or whether such rules are merely easy, descriptive devices that some researchers 'project' onto the human mind has been far from solved.

4. Major methodologies

The purpose of this section is not to describe all possible experimental tasks that are used in psycholinguistic experiments. First, there are so many of them that the reader would soon be bored. Second, there is little sense in describing such a task outside the context of a specific experiment. However, in an introduction to the field of psycholinguistics, it is useful to present an overview of the general methods and experimental techniques that are used in this type of research, as these are the basic instruments in the psycholinguist's toolkit. At this general level, a distinction can be made between three methodologies: corpus research, experimentation, and simulation.

4.1 Corpus research

The analysis of a corpus of language material that forms a representative sample of the phenomenon under study has been used with much success in a variety of psycholinguistic areas, particularly in areas where the measurement of the online process of language use, i.e., real-time measurement, is either unnecessary for the topic of investigation or hard to achieve. Three domains where corpus research has been especially useful are the study of children's language acquisition, the early studies of the speech production process, and the study of the spelling process.
In the study of language acquisition, researchers collect a sample of a child's language output in a particular period of development and attempt to infer its lexical and grammatical knowledge by studying the frequency and distribution of particular word types or grammatical constructions in the corpus. Negative evidence, in the form of the absence of a word or construction at a particular developmental stage or in the form of frequent errors of a particular type, can be as informative as positive evidence, i.e., the frequency and distribution of particular language elements in the corpus (see also the chapter by Gillis and Ravid in the present volume).

The study of speech production, especially in the early period (the eighties of the previous century), has also benefited substantially from corpus studies, more particularly, analyses of large collections of speech errors. A careful and systematic study of these errors allowed researchers to identify the language elements that are treated as separate units by the speech production system and even make statements about the relative ordering of different representational stages in the production of a sentence (Fromkin 1971, 1973; Garrett 1975).

Similarly, the study of spelling errors, both the ones that occur and their distribution and the ones that could occur but do not, has enabled researchers to derive conclusions with respect to the representations and processes causing the errors. Researchers have followed the same rationale as those studying speech errors: when an error can only be explained by assuming the derailment of a particular mental process or the assumption of a particular mental representation, this can be taken as indirect evidence for the existence of this process or representation (Largy, Fayol, & Lemaire 1996; Sandra & Fayol 2003; Sandra, Frisson & Daems 1999).

Note that in each of the three domains just mentioned, techniques have been developed to study (some) processes through the use of online measures (see below).

4.2 Experimentation

This is an entirely different way to assemble data on a psycholinguistic question. The current volume contains a separate chapter on experimentation, in which two important sets of knowledge for successfully performing an experiment are discussed: the basics of experimental design and the basics of statistical significance testing. The latter part includes a clarification of the general rationale behind significance testing, which forms the foundation of any statistical test, but also describes the logic of a small number of frequently used tests in psycholinguistic research. Although a lot of attention is spent in that chapter on the technical aspects of the experimental method, there is little discussion of the actual methodologies that are used in current-day psycholinguistics.

Before turning to a brief review of these methodologies, two general remarks with respect to experimental methodology are in order. First, each experiment is an attempt to understand a particular phenomenon by making a highly controlled design for collecting data on that phenomenon. In this design, researchers manipulate one or more factors that they assume to affect a mental representation or the access speed to that representation (e.g., high-frequency vs. low-frequency words; words that appear in many derivations and compounds vs. words that do not, etc.). At the same time, they match the words in all conditions that are compared to each other on all other factors that could affect the measurements (e.g., word length, number of words differing in one letter, i.e., so-called neighbours, etc.). Second, an important distinction must be made between so-called offline and online measures. An offline measure does not tap into an ongoing language process but only registers the outcome of that process. Hence, it is not a measurement of that process in real time, whereas an online measure is examples of offline measures are the number of errors made on a particular word type in an experimental task, or the ratings on a seven-point scale of the semantic relationship between two words. Examples of online measures are the time it takes to make a particular decision on a word, to fixate a word during sentence reading, to produce the name of a picture, etc. The dynamics in the output of brain imaging techniques are, obviously, also an instance of online measurement.

From my own perspective, all available online methodologies can be sorted into two general categories: chronometric studies and brain imaging, respectively tapping the participants' processing speed and brain dynamics while they engage in a language task (e.g., read a word, listen to a sentence). Note that the category of chronometric studies comprises a large variety of experimental tasks. However, here we will focus on what the experimenter measures rather than on the task that the experimental participant has to perform.

4.2.1 Chronometric studies

Reaction time (RT) measurement is the oldest technique that attempts to tap into the processes underlying language use. It is based on the straightforward assumption that a more complex mental processing task takes more time to finish and that an estimation of this time (the termination of a mental process is obviously not directly observable) can be obtained by having experimental participants make a response to a stimulus (e.g., a word, a sentence). When the hypothesized type of complexity is experimentally manipulated, through the presentation of complex and less complex items, and this turns out to have a significant impact on the RTs, the experimenter concludes that the manipulated variable must be part of a processing model. For instance, if one hypothesizes that the occurrence frequency of a word in written texts determines the time it takes to access its lexical representation in the mental lexicon (expecting shorter access times for high-frequency words) this should be reflected in the time that elapses between the presentation of a word and participants' overt response to it, for instance, a button press indicating their decision that the presented item is an existing word rather than a made-up pseudo-word like blank. If responses are indeed faster on high-frequency words than on low-frequency words, one can conclude that frequency is a determinant of lexical access.

RT measurements bifurcate into two different methods for measuring participants' access speed to a mental representation: those including a conscious decision on the part of the participant and those without. The former are often called self-paced reading or eye-movements studies, whereas the latter are largely restricted to the measurement of the time it takes to identify a word as correct or incorrect. This approach is, at least in theory, unbiased to the speed of any internal processes since the participants are not influenced by their own reading speed.
of the participant (e.g., lexical decision: "Is the word on the screen an existing word or a pseudo-word?") and those without such a decision component, which instead rely on a behavioural index that is assumed to correlate strongly with the ongoing mental activity. A large majority of psycholinguistic experiments makes use of the first method. However, despite the huge popularity of this type of chronometric research and the fact that many data patterns obtained through its use have turned out to be replicable, it has a disadvantage. The introduction of a conscious decision process inevitably adds an extra component to the RT measurement, which causes 'noise' in the RT data. Indeed, the time that is needed to make a decision is not a constant; it varies from one participant to the other and, within an individual participant, from one item to the next. Hence, the registered RTs by no means directly reflect the effect of the design factors on the processing aspect one wants to capture (e.g., lexical access), as they do not only covary with these factors but also with this decision component. Crucially, the contribution of the latter component to the RTs cannot be removed because it is not manipulated by the experimenter but represents a source of random variability.

In addition, a decision component obviously introduces a short period between the termination of the targeted mental process and the participant's actual behavioural response (e.g., button press), the period within which the decision is made. During this period, extra processing of the language stimulus can continue, creating an often invisible trap for researchers: the possibility that some effects can originate during this temporal window rather than during the preceding phase of mental processing that forms the focus of the experiment. For instance, one wants to find out whether people recognize the word mouse faster after just having read cat than after having read pot (to test the presence of associative relations in the mental lexicon), one should consider the possibility that participants may notice this relationship after having accessed the lexical representation for mouse. If they indeed become aware of this relation during the post-access stage, they can rely on it when making their decision (e.g., "A relationship between two items can only mean that they are both words. So, this must be a word."). Hence, the observation of faster responses in the condition with associatively related targets would in principle be ambiguous, as it could arise at two different loci in the processing course: during lexical processing and after this processing has been finished. Obviously, only the former locus is of theoretical interest and can inform researchers on the structure of the mental lexicon (in the present case, that the first word preactivates the second one through an associative connection).

Several researchers have worried from early on (already in the eighties of the previous century) that tasks including a decision component may not reliably reflect the true processes under study — for the simple reason that language users do not make conscious decisions during natural language processing — and, hence, cause a contamination of the true measurements. For that reason, they have looked for techniques that more directly tap into the mental processes under study. The registration of participants' eye fixations during language processing has become a very popular alternative to the classical RT studies (at least, in the domain of reading, and also to some extent in the study of speech perception, see Sections 6 and 7). This technique does not involve any conscious decision; it is only used to monitor a participant's eye fixations while that participant is performing a normal language task. The description of an eye fixation pattern involves questions as: "Where do the eyes 'land'?, "How long does it take before they jump to a different place (technically: make a saccade)?", "When do they make a regression and to which position?". Eye tracking is based on the assumptions (i) that the eyes do not fixate a visual stimulus (like a written sentence) randomly but only fixate those parts that draw the person's attention because they are momentarily relevant (e.g., the word they are currently processing) and (ii) that they spend no longer on each part than necessary. Thus eye fixations are assumed to reflect quite accurately what is occupying the mind at that particular point in time and are, hence, considered to provide a clear window on mental processing. By studying how long the eyes fixate a particular type of stimulus relative to a control stimulus one can draw inferences with respect to the importance of the underlying factor for mental representation and processing.

4.2.2 Brain imaging

Brain imaging is a relatively new technique in the field of psycholinguistics. It has become more and more popular in the first decade of the new millennium. Several researchers attempt to pick up with this new technique and at some research centres, a lot of research on language (and other forms of cognitive and perceptual) processing is performed by making use of various brain imaging methodologies. A well-known centre is the Donders Institute for Brain, Cognition and Behavior in Nijmegen, which was opened in 2002 and where research on language and other forms of processing is done in close collaboration with the near Max Planck Institut für Psycholinguistik and staff members of the Radboud University.

Three often used techniques in the area of brain imaging are ERP (Event-Related Potentials), PET (Positron Emission Tomography) and fMRI (Functional Magnetic Resonance Imaging). In ERP research one attempts to identify changes in the EEG signal that are time-locked to the onset of a particular type of linguistic input, which represents a condition on a variable under study. The significance of some of these time-locked changes in the signal has given rise to a relative consensus among researchers. For instance, the so-called N400 is a negative deflection (hence, the N) in the EEG graph that typically occurs 400 ms following a semantic violation (e.g., The zebra was chased by the mice and ran away in panic), whereas the so-called P600 is a positive deflection (hence, the P) in the EEG pattern that occurs 600 ms after several kinds of grammatical errors, like agreement errors (e.g., The man buy a new car), or atypical grammatical constructions, like garden path sentences (e.g., The broker persuaded to sell the stocks was tall).

Whereas ERP research is based on the measurement of the brain's electrophysiological activity, PET and fMRI scans are instruments that measure changes in the brain's
metabolism, i.e., its rate of energy consumption. A PET scan, which can be made upon the injection of a small quantity of radioactive substance, visualizes brain regions where the neurons use more energy and, hence, where there is more blood flow to provide the necessary oxygen. Thus PET scans indicate which areas are actively involved in (one component of) the task that must be executed. An fMRI scan, which is based on a non-invasive technique (no injection), reflects the rate at which the neurons in different brain areas consume oxygen by measuring the magnetic properties of hemoglobin in the bloodstream (which change as the result of oxygen consumption). Thus the fMRI technique can also be used to visualize active brain regions in response to a particular processing task, like language processing.

4.3 Simulation

Simulation does not measure language users' products (e.g., error frequencies as in corpus research) or aspects of their mental processing (e.g., response speed, changes in the brain's electrical or metabolic activity in an experiment). Rather, its purpose is to use the tools from the field of computational linguistics to the data patterns obtained through experimentation or corpus research, in an attempt to reproduce (simulate) the human data. When trying to do so, researchers rely on a computational model of language processing that has been accurately specified in terms of the nature of its data structures (representations) and computational processes. The extent to which the model succeeds in mimicking the functional data determines how well the theoretical assumptions on which it is based are compatible with the nature of the mental representations and processes in real language processing. Thus a simulation is an important tool for assessing the validity of a theoretical framework for data interpretation. The implementation of this framework in the form of a computational model can lead to its rejection or serious revision.

As a computational model represents the chief aspects of that part of reality that is modeled, it should also be able to make predictions with respect to that reality, in this case, language behaviour. This is the ideal to strive for, but thus far few attempts have been made to accomplish it: build a model that cannot only account for observed effects in psycholinguistic experiments on a post hoc basis but can also predict effects that should be observed in an experiment if the model represents a correct theory on the object under investigation.

Thus, once again, I would compare a computational model for language to a meteorological model, whose knowledge base contains all variables and interactions between these variables that are known to affect the weather and which continually receives updated inputs with respect to these variables from places on a network around the globe. Such a model cannot only explain the current weather but also predict the weather that is still to come. In the same vein, adequate computational models of language should not only be able to explain the observable data patterns in past experiments but also make fairly reliable predictions with respect to observations that still have to be made with respect to a particular question. This marriage between psycholinguistics and computational linguistics is known under the name computational psycholinguistics (Crocker 1996; Daelemans & Van Den Bosch 2005; Dijkstra & de Smedt 1996).

5. Major research techniques

The concepts of 'methodology' and 'technique' are closely related and sometimes used interchangeably. For the sake of exposition I have opted to make a distinction between them. This distinction is, admittedly, debatable but it only serves a pragmatic purpose. It is based on the fact that empirical research on language can differ with respect to (1) the nature of the data that are collected (e.g., corpus data, RIs in an experiment, a simulation output) and (2) the specific 'trick' that is used for collecting these data. This distinction corresponds to my use of the words 'methodology' and 'technique', respectively. The notion 'trick' may refer to a multitude of issues, as will become clear in this section: the way in which the materials are presented, the relationship between successive items, the use of a familiar effect as a benchmark for studying the effect of another factor, etc.

My description of this terminological distinction may be too abstract to be informative. To make it more concrete, I will briefly describe three categories of experimental techniques that are frequently used in the psycholinguistic study of the mental lexicon (each of these 'parent' techniques has produced several descendants over the years). Note that this is not an exhaustive list. These particular techniques enjoy a high popularity score because they have proved to be quite successful in testing hypotheses across a large variety of research topics, both in psycholinguistics and in other areas of experimental psychological research (human perception, other forms of cognition, memory, emotion, ...).

5.1 Using single words to discover important representational factors

A very popular technique for investigating the mental lexicon is to mix a series of unrelated words (sometimes among pseudo-words) and present them one by one at the centre of a computer screen, where they remain until the participant has made a response. Obviously, these words have not been chosen on a random basis. Several of them represent one of the conditions on a variable that is almost impossible to detect by the participants. Examples of such variables are the frequency of a word, the frequency of the stem in a derivation, a word's age of acquisition (i.e., the estimated age at which the word was acquired), a word's so-called family size, (i.e., the number of morphologically related words in which that word appears: derivations and compounds), etc. I will demonstrate the rationale behind this technique by means of the variable that has probably been used more often than any other when studying questions relating to the mental lexicon: occurrence frequency.
Frequency is a particularly potent factor in our memory system. Our memory seems to keep track of the number of times we have experienced or done something. Situations and simple stimuli (like words) that we have often encountered are easier to recognize than comparable situations and stimuli that we have encountered less often. We all recognize a friend or family member we have often met in the past far more easily than a person we have seldom met but who is comparable in all other respects (height, weight, looks, ...). Similarly, the more often we have performed a certain action, the easier we find it to perform that same action on a next occasion, as we all know from our own experience. For instance, learning to ride a bike or drive a car is, for most children and young people, a slow learning process but becomes a form of automatic behaviour once they have repeated the activity sufficiently often.

These are simple observations. How the brain accomplishes this feat—the explanation behind the observations—is an interesting issue of scientific research and dispute. However, we will not enter into this discussion here, as the very fact that frequency differences in past exposure have observable implications for current behaviour is sufficient to warrant the study of the frequency factor at the level of the mental representations involved in language processing.

The first and least exciting question is whether frequency differences in our experience with words also affect the accessibility of their memory representations. They do. Words with a high occurrence frequency (high-frequency, HF, words) are recognized faster in a lexical decision task than words with a low occurrence frequency (low-frequency, LF, words), all other things being equal (Forbach, Stanners, & Hochaus 1974; Scarborough, Cortese, & Scarborough 1977). The fact that representations in the mental lexicon, a memory store for words, are sensitive to the same factor that also affects the accessibility of other mental representations (e.g., human faces) is not really surprising, as it can be anticipated that there will be general principles at work in the cognitive system, affecting the strength of all kinds of memory traces (representations). However, our knowledge that this factor strongly correlates with a word's access speed to the mental lexicon, turns it into a powerful diagnostic when studying more complex issues of lexical access, as will become clear below.

One question in the literature on visual word recognition that has attracted a lot of research attention concerns the issue of so-called morphological decomposition: are derivations and/or compounds decomposed into their constituent morphemes before access to their lexical representation can take place? From a computational perspective, this is quite plausible, certainly in the case of derivations. Derivations with the same stem might all be accessed through the representation of this shared stem. Indeed, there are two frequency-related reasons why the letter string of a derivation could 'fall apart' into its morphemic constituents, even before the word itself is recognized. First, the frequent recurrence of the letter clusters of prefixes and suffixes in several derivations (e.g., re-, un-, de-, -son, -ness, -less) and second, the recurrence of the stem in several derivations (and compounds) increase the frequency of the embedded morphemes, creating local islands of relatively high-frequency letter patterns within the word, which may make the morphemes 'jump out' of the letter string as it were.

Given the above line of reasoning and the reliability of the frequency effect in monomorphemic words, making them a trustworthy index of lexical access speed, it should come as no surprise that frequency manipulation has been used as a technique for investigating the hypothesis of prelexical morphological decomposition (Taft 1979). The logic behind Taft's experimental design was both simple and impeccable. It was also based on a clear rationale: (i) if a prefixed derivation is automatically morphologically decomposed, (ii) if its stem is used as the word's code to provide access to the mental lexicon (where the whole word is stored) and (iii) if frequency determines the speed of accessing such codes which is an established fact, then one can use frequency manipulation of the stem to test the prelexical decomposition hypothesis. As two types of frequencies are associated with derived words—the frequency of the derivation itself and the frequency of its constituent morphemes—Taft selected two sets of derivations that were contrasted on their stem frequency (HF vs LF stem) but matched on their whole-word frequency (they were also matched on length). The hypothesis of prelexical morphological decomposition predicts faster recognition times for derivations with a HF stem, which is what Taft found in a lexical decision task.

He observed the same phenomenon for inflected word forms: items matched on form frequency and letter length but contrasted on the summed frequency of all inflected word forms containing their stem (which he called their base frequency) were recognized faster when they contained a HF base than a LF one. Thus the technique of frequency manipulation provided evidence in favour of a process of automatic prelexical decomposition of (prefix-)derived and inflected words and, hence, suggested a model of the mental lexicon in which all stem-related affixed words are accessed through a shared stem representation.

A lot of work on the role of morphology in visual word recognition has been done since Taft's thirty-year-old experiments, of course (for a review of this literature, see Diependaele, Grainger, & Sandra 2009). The point of describing Taft's experiments is to demonstrate how a familiar effect, like word frequency, can be used as a diagnostic tool for addressing a new research question.

5.2 Priming

In contrast to using isolated words, investigators have also tackled many research questions by appealing to a technique that is explicitly based on the presentation of pairs of words. The words in these pairs are related on a dimension of theoretical interest (e.g., phonological or associative relatedness). This technique is known as priming and has become highly popular over the course of the years, in many areas of experimental psychology.
In psycholinguistics, priming is a technique that is used to study the effect of one word or syntactic construction (the prime) on the recognition or production of a subsequent word or syntactic construction (the target). This effect can be either positive, in which case the prime is said to facilitate target recognition or production, or negative, in which case the prime is said to inhibit target recognition or production. Whatever the nature of the effect, when researchers observe that primes of a certain type affect the recognition or production of targets of a certain type, they take this as evidence that these prime and target types access shared or interconnected mental representations, i.e., that the theoretical dimension on which they are linked plays a role at the level of lexical organization. Note that this is a very reasonable assumption: the access to one representation can only affect the access to another if the two overlap to some extent (in the extreme case, are identical) or are somehow structurally connected in the memory system representing language. Hence, when the priming technique is used with language materials, it opens unique opportunities for studying the internal structure of the mental lexicon.

Priming is an umbrella term, as it comes in a large variety of flavours, each suited to a particular purpose. I will not attempt to describe all possible variants of the priming technique but will emphasize three dimensions with respect to which each specific priming technique can be situated. These dimensions may not be exhaustive but certainly cover a broad range of applications of the technique in the literature.

The first dimension is the relationship between prime and target. The nature of this relationship is, of course, the main reason why the technique is used at all. As already mentioned, when a particular prime-target relationship causes priming effects, the underlying rationale of the technique supports the conclusion that this relationship is somehow encoded in the structure of lexical representations, their connectivity patterns, or the nature of the access mechanism. For instance, if one wants to find out whether the process of word recognition is affected by phonological factors, i.e., whether lexical access is sensitive to phonologically similar words in the mental lexicon, one can use phonological priming. If responses to targets in prime-target pairs like *mat-cat* are faster or slower than to the same targets in matched control pairs lacking a phonological relationship, like *pin-cat*, this can be treated as evidence that the lexical access process is sensitive to phonological similarity.

Psycholinguists are interested in linguistically relevant relationships, like phonological, morphological, and semantic ones, but also focus on relationships that are of little linguistic relevance, like the orthographic similarity or the associative relationship between words. Note that associatively related words need not be semantically related, although both relationships tend to co-occur (e.g., *bread-butter*). Words like *monkey* and *banana* are not semantically related—one refers to an animal, the other to a piece of fruit—but have a strong associative relationship: the word *monkey* makes many people immediately think of the word *banana*, given the frequent co-occurrence of monkeys and bananas in reality. The fact that psycholinguists study both linguistic and non-linguistic relationships demonstrates their main focus is neither on the structural dimensions of language nor on the psychological reality of linguistic distinctions. Rather, their goal is to uncover the internal structure of the language processing system, i.e., to describe the nature of the mental representations and processes that make language use possible. For instance, in the study of the mental lexicon, the ultimate purpose is to find out which dimensions determine the connectivity patterns among words and, hence, have an impact on lexical processing. Some of these dimensions may be linguistic in nature, others not at all. As mentioned in the first section of this chapter, current-day psycholinguists no longer pay lip-service to linguists by verifying their theories. Rather, they have a well-defined goal of their own (see also Chapter 1 of this volume).

A second dimension along which priming techniques considerably differ is the visibility of the prime. The first series of priming studies made use of a long-distance priming paradigm, in which prime and target words were both clearly visible and were separated by a number of intervening experimental trials (e.g., 48 trials in Fowler, Napps, & Feldman 1985) sometimes even by two days (Scarborough, Cortese, & Scarborough 1977). Using this technique, one discovered, for instance, that identical word repetition always leads to faster responses on the second presentation of a word but that this facilitation is larger for LF words than for HF ones (Forster & Davis 1984). One also discovered morphological priming effects from derived word primes on their stem targets (appearance-appear; see Stanners, Neiser, & Hall 1979).

However, one soon hit upon an annoying possibility: the data of this type of priming paradigm were likely to be seriously contaminated by the contribution of so-called episodic memory factors (Jacoby 1983; Whittlesea & Dorken 1993). Episodic memory is a memory ‘compartment’ where we supposedly store all our personal experiences, for instance, what we ate this morning but also the fact of having seen the word *appearance* a few sentences ago. The contribution of episodic memory in long-distance priming is, for instance, suggested by the observation that the divergence in repetition effects for HF and LF words in a lexical decision task is paralleled by a similar asymmetry in a recognition memory test (“Did this word occur in the list you have just seen?”), where participants have to rely on episodic memory. When participants have to indicate whether they recognize a word as an instance from the study list, they perform better on LF words like *echo* than on HF ones like *man*. This suggests that LF words leave stronger traces in episodic memory than HF ones, which might indicate that LF words require larger processing demands for memory encoding (Rachal & Reder 2006). If repetition priming effects are mediated by episodic memory traces, the observed priming asymmetry for LF and HF words would be readily explained (but see Kinoshita 1995 for an argument against a common origin of frequency effects in recognition memory and repetition priming).

However, the most compelling reason for questioning the validity of these priming data for the study of the mental lexicon was a theoretical one. If priming effects can last for several minutes and even days, the consequence would be that the activation decrease in
within a short period of time each lexical representation would be in such a high state of activation that recognition could occur on the basis of very little stimulus information. This, in turn, would annihilate the frequency effect, a prediction that is squarely at odds with the observation that this is one of the most stable effects in the literature. At this point, the technique of masked priming came to the rescue.

The masked priming technique was first introduced into the field by Humphreys and colleagues (Evett & Humphreys 1981, see also Humphreys, Besner, & Quinnan 1988; Humphreys, Evett, Quinnan, & Besner 1987) but became a very popular technique ever since the seminal work by Forster and Davis (1984). The technique owes its name to the use of a stimulus, the mask, that makes it impossible for participants to discover the identity of the prime, like a true mask conceals the identity of the person who is wearing it. Technically, this is accomplished in the Forster and Davis paradigm by presenting a sequence of three visual events: a series of hash marks (#####), which is presented for 500 ms, immediately followed by a very short presentation of the prime in lowercase letters (60 ms or less), which is in turn immediately followed by a target in uppercase letters. The target remains on the screen until the participant's response. The combined action of the temporal relationships between the stimuli and their superposition causes the impression that there are only two 'events' on the screen: a series of hash marks, followed by an uppercase letter string. Participants are surprised to hear after the experiment that on each trial a stimulus in lowercase letters had been 'sandwiched' between the hash marks and the target.

Despite the invisibility of the primes at a conscious level, many studies on different topics have demonstrated reliable and replicable effects with this technique. Most importantly for the present discussion, Forster and Davis demonstrated that the subliminal, i.e., masked, presentation of a prime eliminates the episodic effects that plague the priming paradigm with visible primes. When using a set of LF and HF words they found the typical interaction between identical repetition and frequency when the primes were visible - a larger facilitation effect for the LF words - but not when the primes were masked. In the latter case LF and HF words caused equally large repetition effects. Importantly, these could not be explained as the result of orthographic priming. Indeed, neither pseudo-word repetition (#####-flurp-FLURP) nor one-letter-different prime words (#####-race-FACE) caused reliable priming, which should have been the case if the nature of the effect were orthographic instead of lexical.

A third dimension that can be used to classify priming paradigms is the language modality of prime and target. Researchers who are interested in visual word recognition will obviously present their targets visually but have the choice to present their primes either visually or auditorily. The same choice situation occurs in research on speech perception, where auditory targets are used. When prime and target belong to different modalities (visual vs auditory), the technique is called cross-modal priming. Cross-modal priming has been used in a number of studies (e.g., Diependaele, Sandra, & Grainger 2005; Grainger, Diependaele, Spinelli, Ferrand, & Farioli 2003; Marslen-Wilson, Tyler, Wacksler, & Older 1994; Zwitserlood 1989; see Tabossi 1996 for a relatively old review on cross-modal semantic priming) but the technique is certainly not as widespread as intra-modal priming. The main reason for this is probably technical: it is more time-consuming to prepare an experiment where different modalities are combined.

The logic behind cross-modal priming is quite sensible. It rests on the idea that priming effects obtained through the use of primes and targets in different modalities occur at the level of the abstract, modality-independent representation of a word. This representation bundles all specific representations of the word (e.g., orthographic, phonological), all of which are accessed through modality-specific access processes. For instance, the orthographic representation is accessed through a processing path that starts with a visual encoding of the stimulus, whereas the phonological representation is accessed through a processing route that starts with an auditory encoding. Cross-modal priming prevents the danger that the obtained effects only highlight the internal structure of the modality-specific representation (e.g., the orthographic one) rather than the structure of the abstract, modality-independent lexical representation.

5.3 Inducing interference

The attempt to create an interference effect in the processing chain deserves special mention, even though this technique is always combined with one of the two techniques mentioned above (single word presentation or priming). In contrast to the previous techniques, it is not based on the assumption that an important representational variable will leave its fingerprint on the RIs and errors or that a relationship that is encoded in the connectivity structure of the mental lexicon will affect these measures. Rather it sets out from the assumption that a processing conflict will arise when two independent processes access conflicting types of information, and that this will be detrimental for performing the experimental task: slower RIs and more errors will be observed compared to a control condition. Thus, inducing response interference can be informative on the processes that a particular stimulus type sets in motion and on the nature of the representations that are activated by these processes.

One of the best known examples of an interference effect in language processing is probably the well-known effect of Stroop interference (Stroop 1935). The experiment demonstrating the effect is very elegant one, i.e., it does not depend on a complex design, and the basic observation can be made by anyone on a home computer or even by printing the coloured stimuli on a piece of paper. The observation is that people find it quite hard to name the colour in which a particular colour name is printed when there is a mismatch between the to-be-named colour and the colour to which the name refers. For instance, people need time and/or make more errors when they have to say "red" when the word green is printed in red letters than when a word like chair is printed in red letters. This observation demonstrates very convincingly that experienced readers cannot ignore
words, i.e., that the process of word recognition is highly automatic and beyond readers' control. They cannot decide not to recognize a word once their eyes fixate on it, even in conditions where this ability would considerably simplify their task (e.g., the Stroop test).

A second illustration of a type of experiment that capitalizes on the concept of interference can be found in research on the bilingual mental lexicon (see the chapter by Dijkstra in this volume). One of the major questions in this research domain is whether bilinguals access their mental lexicon in a language-nonselective or a language-selective way. Are they able to suppress all lexical representations in a language that is irrelevant in the current context of language use? Cognates and interlingual homographs have been ideal word types for the study of this question in the domain of visual word recognition. Both these word types have the same orthographic form in a bilingual's two languages (say English and Dutch). Cognates also share their meaning (film), whereas interlingual homographs do not (room, which refers to cream in Dutch). If the language-nonselective view on lexical access is correct, interlingual homographs should cause a temporary conflict within the reader's processing system, at least when the task is to decide whether the stimulus is, for instance, an English word. Being confronted with a matching lexical representation in Dutch and English for a word like room, the ensuing response conflict will require the suppression of the task-irrelevant representation, which will result in longer RTs and/or an increased error rate.

A final illustration of the use of the interference concept comes from the domain of language production. A well-known paradigm in that area is the so-called picture-name interference technique, which has become popular since the study by Schriefers, Meyer, & Levelt (1990). Participants are shown a picture and are asked to name it as quickly as possible. However, on the picture itself a word is printed (at fixation position). For instance, a stimulus in this type of experiment could be the picture of a horse with the name mule written over it. Participants are told to ignore this word when naming the picture, as it is entirely irrelevant for their task. Obviously, the word has a well-defined function and, as researchers have learned from the Stroop effect, cannot be ignored. This makes it possible to choose words with respect to a particular variable on which the picture name and the printed word are related (semantically, phonologically,...) with the purpose of finding out whether this variable plays a role in the process of language production, and, if so, at what moment in the time course. The latter can be discovered by varying the onset of the picture and the word relative to each other (the word can also be presented auditorily).

For instance, when participants have to say “horse” but read the word mule at the same time, they may be confused by the semantic relationship between these two words, which will be reflected in slower naming times and/or more errors, compared to a control condition in which, for instance, the word road is written on the picture. By varying the time interval between the moments of picture and word onset and studying how this temporal relationship affects the absence, presence, and size of the interference effect, one can infer when semantic information is retrieved during the time-course of the speech production process (or, phonological information, when the distractor word is phonologically related to the picture name, e.g., the picture of a dog with the word doll printed on it). For a discussion of findings obtained with this technique, see Levelt, Roelofs, & Meyer (1999).

6. Studies on language perception

In this section I will review some of the most important insights into the perception of language, both with respect to the reading process (6.1) and to the process of speech perception (6.2). Note that this will be a highly restricted overview of the literature, as a separate chapter (or book) could be devoted to the research that has been done on each of the four language skills. For instance, I will restrict myself to research on lexical processing and not deal with the literature on syntactic parsing during sentence comprehension, even though many studies have dealt with this issue (and others) as well.

6.1 The process of visual word recognition

A lot of psycholinguists have been concerned with the process of visual word recognition. Obviously, the recognition of a written word is only a small part in the entire set of processes that are involved in the reading of a text. However, it is an important step, representing a core skill on which all other processes involved in reading a text are contingent: the retrieval of the word's meaning, the integration of that meaning into the preceding sentence fragment, which, in turn, is required to integrate that sentence meaning into the semantic representation of the whole text. As a result, people with severe problems in written word recognition usually have problems with text comprehension as well (Berninger, Abbott, Vermeulen, & Fulton 2006).

In what follows I will make a distinction between factors that are operational in the time window before a lexical representation is accessed and ultimately lead to such access, and factors that determine the accessibility of the lexical representation itself, i.e., factors that make lexical access easy or difficult.

6.1.1 Processes at the prelexical processing level

6.1.1.1 Prelexical morphological decomposition

In Section 5.1 we have already addressed the question whether morphologically complex words (inflected word forms, derived and compound words) are prelexically decomposed into their constituent morphemes, in other words, even before the lexical representation of the whole word is accessed. We will return to this issue here, because it is one of the most discussed prelexical processes in the literature. Needless to say, the hypothesized processes at this level operate automatically and fall beyond any form of conscious control on the part of the language user. Language users are also completely unaware of their existence.
Recent research seems to converge on the conclusion that morphologically complex words are automatically decomposed into their constituent morphemes at the prelexical level. This was already the conclusion of Taft and Forster's pioneering research (1975, 1976). In their 1975 paper, these authors demonstrated, for instance, that the rejection of a pseudo-word takes longer and gives rise to more errors when it occurs as a bound stem in a prefixed derivation (e.g., *juvenile* in *rejuvenate*) than when it is a matched control derived from a pseudo-prefix word (e.g., *per-toire* in *repertoire*). This suggested to them that even bound stem morphemes have a special representational status in the mental lexicon. In their 1976 paper, they showed that the frequency of the first constituent of a compound word determines participants' response speed in a lexical-decision task, suggesting that the whole compound is accessed on the basis of its first constituent, which requires prelexical decomposition.

Although the question regarding prelexical morphological decomposition has been a dominant research theme ever since, it has seen a strong revival in the past ten years, when researchers decided to tackle the question by using the masking priming technique. Two studies in the early ages of the new millennium seemed quite convincingly that there is a stage preceding lexical access in which the processor attempts to identify letter strings that match the orthographic pattern of morphemes. Using 43 ms masked morphologically complex primes and stem targets, Rastle, Davis, Marslen-Wilson, and Tyler (2000) obtained morphological priming effects that were independent of whether the prime was semantically transparent (*departure-depart*) or not (*apartment-apart*). This suggests that in the very early stages of visual processing potential morphemes are identified on the basis of orthographic patterns in the stimulus word. Note, incidentally, that a prelexical process has no way of “knowing” what the linguistic status of an orthographic segment is—it has not reached the lexical level yet—and can, hence, only be driven by pure form information.

In 2003 Logn, Segui, and Hallé reported strong evidence in favour of blind prelexical decomposition of a word’s ‘surface morphology.’ Using 46 ms masked morphologically complex primes and stem targets, they found significant facilitation effects on targets following semantically transparent primes (*gaufrite-gauffre*) or semantically opaque primes (*vinette-vigne*), effects which did not differ from each other, but also on targets following pseudo-derived primes, consisting of a pseudo-stem and a pseudo-suffix (*baguette-bague*). The evidence that the source of this effect was the surface morphological structure of the prime, i.e., the fact that the word was a concatenation of potential morphemes based on the segments’ orthographic pattern, came from a control condition. In this condition items consisted of a potential stem, followed by a letter sequence that did not match the orthography of a suffix (absence of a surface morphological structure, e.g., *avocet-avr*). An inhibition effect was found, indicating that only words that could be exhaustively parsed into potential morphemes (as is obviously the case for any true derivation) are decomposed.

In a recent masked priming study McCormick, Brysbaert, & Rastle (2009) wondered whether prelexical decomposition was limited to LF words or applied across the board. They observed evidence favouring prelexical morphological decomposition, both for HF and LF words, which made them conclude that the process is mandatory and indiscriminately applies to all incoming letter strings.

The discussion on whether the initial processing stages are really insensitive to semantic influences is still ongoing, as evidence has been presented, both for prefixed and suffixed derivations, both in experiments with French and Dutch words, that there is an effect of semantic transparency at very short prime durations (Diependaele, Sandra, & Grainger 2005; Diependaele, Sandra, & Grainger in press, Feldman & Basnight-Brown 2008).

### 6.1.1.2 Prelexical phonological recoding

A second candidate for a prelexical process has attracted the attention of many researchers as well: the possibility of automatic prelexical phonological recoding, i.e., the mapping of the orthographic representation of the written word onto its phonological representation. Does this process take place automatically during prelexical processing, such that lexical access can (or must) take place on the basis of a phonological representation? The findings obtained with several variants of the masking paradigm have led to a relatively strong consensus that automatic phonological recoding is indeed an automatic prelexical process. Note that in all experiments homophonous pseudo-words (i.e., a nonword that is homophonic with a word) were used, rather than homophonic words. If words were used, the effect could equally well originate within the lexicon as prior to lexical access, i.e., the outcome would be ambiguous.

The first paradigm, backward masking, comes in different versions. Perfetti and Bell (1991) used the following technique. First they presented a word in lowercase letters (*blue*) for 35 ms, followed by a nonword in uppercase letters for 30 ms, which was finally followed by a row of X signs (XXXXXXXX). The participants’ task was to identify the lowercase word. They found that this was easier when the uppercase nonword, which acted as a backward mask for the word, was a pseudo-homophone of that word than when it was a matched non-homophonic nonword (*BLOO* vs *BLOS*). This outcome suggested that the word had been phonologically recoded very rapidly and that this phonological representation was not destroyed by a subsequent homophonic nonword mask. In a different version of their backward masking paradigm Perfetti and coworkers (Perfetti, Bell, & Delaney 1988) used a pseudo-homophone effect. This time they only superimposed the briefly presented (20–50 ms) lowercase word by an uppercase nonword.

The second paradigm is the masked priming paradigm described earlier, in which a 500 ms row of hash marks (forward mask) is followed by a brief lowercase prime word, which is in turn overwritten by an uppercase target (acting as a backward mask). Using this technique with French items Grainger and Ferrand (1996) obtained evidence for fast phonological recoding: briefly presented masked primes caused faster responses when
they were pseudo-homophones of the target (###-lont-LONG) than non-homophonic controls sharing the same number of letters with the target (###-tune-LONG).

Finally, Brysbaert and coworkers combined the two paradigms. In their technique the nonword did not follow the word as a backward mask (as in Perfetti and Bell’s experiment), but preceded it as a prime. Hence, participants would see trials of the type blue-blu>-xxxxx, where both prime and target were presented very briefly (43 vs 29 ms, respectively), and would be asked to identify the target. Using this technique Brysbaert and colleagues reported two studies in which they showed pseudo-homophone priming effects. Moreover, in these studies, they demonstrated that these effects also occur across languages: an L2 word is more rapidly recognized when its prime is homophonic to it according to grapheme-phoneme correspondences in the participants’ L1 (e.g., soer-SOURD is an example of their Dutch-French pairs of L1 pseudo-homophones and L2 targets) and, even more surprisingly, an L1 word is recognized more rapidly following a pseudo-homophone in L2 (Brysbaert, Van Dyck, & Van de Poot 1999; Van Wijnenlaarle, & Brysbaert 2002). Thus these authors demonstrated mandatory prelexical phonological recoding in both the first and the second language.

In a recent review of the literature, Rastle and Brysbaert (2006) perform a meta-analysis on a set of published experiments, report new experiments, and simulate their results. They arrive at the conclusion that the prelexical effect of phonological recoding is small but reliable and supports the existence of a process of prelexical phonological recoding.

6.1.2 Factors determining the accessibility of a lexical representation

The accessibility of a lexical representation is determined by two sets of factors: those that determine the strength of its memory trace and those that affect the ultimate selection of the target representation.

6.1.2.1 Factors affecting the strength of a lexical representation

The strength of a lexical representation is determined by at least two factors: word frequency and age of acquisition. As the effect of word frequency has already been discussed earlier, I will only consider the age-of-acquisition effect here, although its relationship to the frequency factor has given rise to a lot of debate (see below).

The term ‘age of acquisition’ refers to the (estimated) age at which the word was acquired. Obviously, this factor is highly correlated with word frequency: a word that is acquired early in life often becomes a HF word (but not always; e.g., children acquire the word fairy early in life but in adult language use, this is a LF word). This correlation can, of course, be measured by estimating the word’s age of acquisition (e.g., by asking schoolchildren at what age a word should be known) and counting the number of times that this word appears in a representative corpus of texts that are adults confronted with (for instance, the CELEX database for Dutch, which is derived from a corpus counting 42 million word tokens, and Basyle, Piepenbrock, & Van Rijn 1993). The strong correlation between these two measures makes it almost inevitable that two word sets that are contrasted on frequency will also differ on their mean age of acquisition. Hence, what is the causal factor behind faster responses on HF words: their frequency or their age of acquisition?

Morrison and Ellis (1995) wrote a seminal paper in which they systematically disentangled these two factors. When contrasting two sets of words on frequency and matching them on age of acquisition they found no differences in response speed and errors in a naming task. In contrast, when contrasting two word sets on age of acquisition while matching them on frequency, they obtained reliably faster response times for the words that had been acquired early in life. This caused them to make the bold claim that all frequency effects in the literature were actually effects of age of acquisition.

However, several researchers made the counter-argument that was waiting in the wings: early acquired words are recognized faster for the simple reason that, in the course of someone’s life, they have been so frequently encountered that their lexical representations have become much stronger than the representations for words acquired later in life. This account was the exact opposite of Morrison and Ellis’ claim: the truly important factor is frequency, not the word frequency that can be found in a frequency count (which only estimates the frequency of a word in a selection of texts at a particular moment in time) but the cumulative frequency across a person’s life span. Given the theoretical importance of this debate — are there two distinct factors that separate the accessibility of a word’s representation or only a single one? — a lot of researchers set out to test the validity of the intuitively plausible cumulative frequency account.

The outcome of many experiments by even more researchers is that cumulative frequency cannot explain the observed effects. Ghyselinck, Lewis, and Brysbaert (2004) compared the effect of cumulative frequency and age-of-acquisition in a number of different tasks (naming, lexical decision, …) and arrived at two important observations and conclusions. One observation directly addressed the cumulative frequency hypothesis in order to predict the observed RIs with a mathematical formula, the weight of the age-of-acquisition factor had to be much stronger in the equation than the weight of the frequency factor. This rejects the hypothesis that age-of-acquisition effects can be reduced to frequency effects. The second observation was that, in the majority of tasks, independent effects of frequency and age of acquisition were observed. These effects did not interact with each other but were at the same time highly correlated: a large/small frequency effect was accompanied by a large/small effect of age of acquisition. This finding suggests that we are dealing with two separate factors that are, however, both tied to the same processing stage and affect the same learning mechanism. Accordingly, the authors claimed that the evidence strongly suggests “that AoA [Age of Acquisition] and frequency effects are both likely to be the result of the way in which information is stored and accessed in the brain.” [. . .] “making it difficult to maintain that they do not have a common basis” (Ghyselinck et al. 2003, pp. 50 and 62, respectively, my emphasis). As far as the nature of this common basis is concerned,
the idea is that early words and concepts are easy to store in an (almost) empty mental lexicon, whereas the task of maintaining an integrated memory structure by adding new words becomes ever more difficult as the lexicon expands, i.e., early acquired words are more strongly anchored in the mental lexicon.

The cumulative frequency hypothesis has also been rejected by other research teams, two of which will be mentioned here: Stadilhagen-Gonzalez, Bowers, and Damian (2004) used expert vocabularies to investigate the issue. They selected high-frequency words from scientific journals in the disciplines of psychology, chemistry, and geology and found that late-acquired words like cognition were responded to faster by experts (for whom these are HF words) than non-experts (for whom these are LF words), a pure frequency effect. More importantly, they also found two pieces of evidence ruling out the hypothesis of cumulative frequency and supporting a distinct effect of age-of-acquisition instead: (i) Late acquired HF words, i.e., specialist words, were not recognized faster than early acquired LF words like dragon, despite the fact that the late acquired HF words had a much higher cumulative frequency, (ii) early acquired HF words were recognized faster than late acquired HF words, despite being matched on their cumulative frequency. They concluded that "both Early and High F words may have 'stronger' lexical representations that are more easily accessed" (p 819). In other words, the suggestion is that both frequency and age-of-acquisition are independent factors, each of which determines the accessibility of a word's lexical representation.

De Deyne and Storms (2007) studied the phenomenon by using words that only recently entered the language, i.e., words like mango. Their reasoning was that an age-of-acquisition effect should be found when comparing two age groups: young adults (18-23) and older adults (52-56). The only difference between these groups is the age at which they acquired these words. For both groups the words were matched on frequency and cumulative frequency, as there is no reason to suspect why one group would encounter these words more often than the other (still, precautions were taken to control for this risk). The outcome of the experiments were clear: the difference in word recognition times in a lexical decision task were predicted by the difference in age-of-acquisition (for each word the average age-of-acquisition in the young group was subtracted from the average age-of-acquisition in the older group) but not by the difference in the participants' rated familiarity with the words. This supports the claim that age-of-acquisition is the crucial factor, not (subjective) frequency. In separate analyses on the data of the younger and older groups, RIs were predicted both by the participants' age-of-acquisition, their rated familiarity, and the words' frequencies in a frequency count. The effect of frequency was found after the effects of age-of-acquisition and the effect of familiarity had been statistically removed. The latter finding corroborates the claim made by Ghyselinck et al (2004) that age-of-acquisition and frequency are two independent factors that determine the accessibility of a lexical representation.

6.1.2.2 Factors affecting the selection of a lexical representation

I will discuss two factors that affect the ease with which a lexical representation gets selected from the large set of representations that is stored in the mental lexicon: orthographic neighbours and family size. Generally speaking, the purpose of word recognition is to select one word from a vast word pool in long-term memory. In a recognition system where only one representation gets activated by the incoming stimulus, this would be a relatively simple process. The sensory stimulus would be encoded into the format that is used by the representational system used in the mental lexicon and then directly mapped onto the appropriate representation. However, research has shown that this is not the way our word recognition system operates.

Orthographic neighbourhoods

One property of visual word recognition is that the activation of lexical representations does not follow an "all-or-none" principle, such that a particular lexical representation is either activated or not, but rather operates on the basis of similarity. The degree of similarity between a word that is being read and the orthographic pattern of a word that is stored in the mental lexicon determines the degree of activation of the latter word's lexical representation. Although similarity is a concept that is quite hard to define, a simple and straightforward operationalization for the practical purpose of item selection has been functioning pretty well for several decades already. In this operationalization a word is considered to be sufficiently similar to the input for being coactivated during word recognition when (i) it shares all letters but one with the target word and (ii) all of its shared letters occupy the same position in the word. Such a word is called an orthographic neighbour and can be obtained by replacing one letter in the presented word by another letter if that operation turns the target word into another word. For instance, given the target word book, some of its orthographic neighbours are cooK, hook, looK, teak, bonk, boom, boot. Some words have many such neighbours, like book, whereas others have very few or none, like echo. The concept was introduced by Max Coltheart a long time ago (Coltheart, Davelaar, Jonasson, & Besner 1977) and, as mentioned, its validity has been demonstrated in many experiments, such that it is widely accepted in the field. As orthographic similarity between word forms might affect the speed of lexical processing (for the simple reason that similarity would cause the processing system to remain uncertain about the identity of the target), a lot of attention has been devoted to effects of a word's orthographic neighbourhood.

An important distinction should be made between two concepts that are associated with the notion of orthographic neighbours: neighbourhood size, which refers to the number of neighbours of a word, and neighbourhood frequency, which refers to the frequency of the neighbours (with a focus on whether the neighbours' occurrence frequency is higher or lower than that of the word that must be recognized). Hence, there are two critical questions: Does the presence of more orthographic neighbours facilitate word recognition or
not? Does it matter whether one or several of these neighbours have a higher occurrence frequency than the target or not?

When studying the effects of orthographic neighbours one should be careful to methodologically separate the effects of neighbourhood size and frequency. Comparing two word sets that simultaneously differ with respect to their number of neighbours and the frequency of these neighbours would obviously make it impossible to make claims with respect to either factor. Hence, researchers should contrast one factor while keeping the other constant. Space limitations make it impossible to do justice to the literature on the effect of orthographic neighbours but in what follows I will try to sketch the outlines of the emerging picture (for a review up to the end of the nineties, see Perea & Rosa, 2000).

Neighbourhood size It has been found that neighbourhood size affects the speed with which a word is recognized but that the direction of the effect depends on the requirement made by the experimental task. Tasks in which the exact identification of a word is called for tend to reveal an inhibitory effect for words with high-density orthographic neighbourhoods, whereas tasks that make it possible to respond without identifying the word precisely tend to reflect a facilitatory effect when the word has many orthographic neighbours. For instance, the lexical decision task makes it possible to respond on the basis of a general index of ‘wordlikeliness’, when a letter string looks like a word, i.e., is similar to many orthographic patterns in the mental lexicon, participants can respond ‘yes’ before having fully identified the word. In such a task, the presence of many orthographically similar words will cause a lot of activation in the mental lexicon (Grainger & Jacobs, 1996), which participants can use as evidence that the target is likely to be a word. This is in line with a study by Andrews (1992), who found that low-frequency words are recognized faster in a lexical-decision task when they have a large orthogonal neighbourhood. It also fits in with the study reported by Pollatsek, Perea & Binder (1999), who equated the words in two sets on the frequency of their highest frequency orthographic neighbour while manipulating the words’ neighbourhood size. In a lexical decision task, faster responses were given to words with the higher number of neighbours.

Whereas lexical decisions can be made on the basis of a sense of familiarity or ‘wordlikeliness’, some response types cannot rely on such information because they require the unique identification of a word. Progressive masking is an example in case. The participants’ task in such an experiment is to identify the word that is on the screen. However, this is made difficult by showing the word only very briefly and immediately replacing it with a mask. The total duration of word and mask is a constant, but on each cycle the word is presented for a somewhat longer duration (say, 10 ms) and the mask is correspondingly presented for a shorter duration. The phenomenological experience is that the initially hidden word is gradually unmasked (hence the name ‘progressive masking’). This procedure continues until the participant says or writes down the word that she/he thinks to have seen. In such circumstances, where precise identification matters, a large number of neighbours delays responses relative to a matched word with fewer neighbours (Carreiras, Pereira & Grainger, 1997). According to activation-based theories of the mental lexicon, the more neighbours that are activated, the stronger the competition among these neighbours (inhibition) in the selection process of word identification, which complicates the process of unique identification and, hence, causes longer response times.

Another way to avoid that participants in reading experiments rely on cues that have a large degree of response validity in the experimental context (like orthographic similarity as an index of wordlikeliness) is to use a method in which participants must recognize words without making decisions on them. That can be achieved by presenting the critical words in sentences that have to be read for meaning and measuring the time that participants spend on reading the critical words, using the technique of eye monitoring. Studies in which participants have to read sentences for meaning (they have to reply to a question following some sentences) while their eye movements are being monitored seem to indicate exactly the opposite effect of neighbourhood density than a lexical decision task: more orthographic neighbours increase the duration of eye fixations and cause more regressions to the word, i.e., refixations after already having moved to the next word, because readers are apparently unsure whether they have recognized it correctly.

Again the study by Pollatsek, Perea, and Binder (1999) is important in this respect. When equating two sets of words on their highest-frequency neighbour while contrasting them on their neighbourhood size, they found that words with large neighbourhoods caused inhibition effects, i.e., longer reading times during eye tracking. In another experiment they equated the number of higher-frequency words in two word sets and manipulated the neighbourhood size between these sets. Words with more orthographic neighbours were skipped more often but this gave rise to later inhibition in the reading process (slower reading), suggesting that readers had been misled by the orthographic similarity of the critical word to many other words and selected the wrong word. In other words, the same index of familiarity, reflecting the presence of many orthographic neighbours, which is helpful in the lexical decision task, seems to cause problems in a natural reading task, which does not involve a decision component. This makes sense: a high degree of orthographic similarity with other words encourages a ‘yes’ response in the decision phase of a lexical decision task (“it is a word”) but seriously complicates the process of word selection that is required in normal reading (“it is that word”).

However, we should probably not be blinded by the task differences discussed above but by the commonality across tasks. The evidence indeed converges on the observation that neighbourhood size affects reading times and is, for that reason, a critical factor in the word recognition process. The observation that the direction of the effect is task-dependent is interesting, but tells us more about the processes that are mobilized by the task than about the importance of neighbourhood size for the process of word recognition.

Neighbourhood frequency What is the impact of the orthographic neighbours’ frequency? The first to study this were Segui and Grainger (1990), using the masked priming paradigm discussed earlier (Forster & Davis, 1984). They found that 60 ms primes...
that were higher-frequency neighbours of a LF target delayed lexical decision responses on that target whereas primes that were lower-frequency neighbours did not have an effect relative to a matched control word. This finding can easily be explained within an activation account of the mental lexicon, in which similar words are activated and compete for selection, which occurs when one candidate exceeds the activation threshold that is required for word recognition. A higher-frequency prime that is an orthographic neighbour of the target is easily activated and hence suppresses the process of activation build-up in the target's lexical representation, causing a delay in response times. A lower-frequency neighbour, on the other hand, cannot suppress the accumulation of activity in a higher-frequency target, which reflects itself in the absence of an effect.

Perera and Pollatsek (1998) came to the same conclusion as Segui and Grainger (1990), but used a different rationale and different experimental paradigms. They compared two word sets that were matched on all relevant factors but different in one respect: members from the first set had no higher-frequency neighbours whereas members from the second set had at least one such higher-frequency neighbour. They found that (unprimed) lexical decisions took longer for words of the latter set. Moreover, in an eye-tracking study they found more regressions for this same set of words, suggesting that participants were unsure about the identity of the word they had just read.

However, in two recent studies Stephen Lupker and colleagues have cast some doubt on the almost axiomatic idea that higher-frequency neighbours always delay the reading process and that, when experimental techniques produce a delay, these neighbours are the only ones to cause it. Sears, Sharp and Lupker (2006) used a lexical decision task to find out whether words with higher-frequency orthographic neighbours are indeed more difficult to recognize in a simple lexical decision task and came to the conclusion that "higher frequency neighbours have little, if any, effect on the identification of English words". In a very recent publication Nakayama, Sears and Lupker (2008) used the same masking technique as in the study by Segui and Grainger (1990) but manipulated both the frequency and the neighbourhood size of both the primes and the targets. They found the same effects as Segui and Grainger but with an important qualification: orthographic neighbours of the target that are more frequent than the target itself delay its recognition, but only when primes and targets have few neighbours. When they have many neighbours, lower-frequency primes were found to delay the recognition of higher-frequency targets as well. This study not only suggests that both the frequency and the density of the orthographic neighbours are important in word recognition (that is the conclusion when looking at the entire set of studies discussed above) but that they interact with each other in unexpected ways.

Family size
The effect of orthographic neighbours demonstrates that a written word activates more than just the lexical representation by which it is represented in the mental lexicon. Besides coactivation on the basis of orthographic similarity it has been found that another type of similarity also leads to the coactivation of lexical representations, and hence affects the speed of word recognition: morphological similarity. This effect is known as the family size effect and has been discovered by Harald Baayen and Rob Schreuder (Schreuder & Baayen 1997). The title of their paper, How complex simplex words can be, nicely captures the essence of the phenomenon. A simple monomorphemic word like, for instance, *man*, is related to a high number of morphologically complex words that contain it as their stem or as a constituent: derivations (*manly, manhood*) and compounds (*mankind, manpower, barmen, businessman, countryman, craftsman, snowman, strawman, *). This is the word's morphological family. What Schreuder and Baayen (1997) found -- and what has been replicated in numerous studies since the publication of their paper -- is that the number of these complex words (the morphological family size) affects the recognition time of the simplex word (e.g., the noun *man*). The larger the size of the family, the faster the target word is recognized. Note that inflected word forms are excluded from the morphological family, as these are merely form variants of the same word, i.e., belong to the domain of inflectional morphology, and not different words like derivations and compounds, i.e., which belong to the domain of lexical morphology.

Importantly, this effect cannot be reduced to a mere form frequency effect in disguise, resulting from a process of obligatory prelexical decomposition in the case of derivations and (possibly) compounds. If that were the case, the cumulative frequency of the morphological family members would account for the response times, whereas it does not. The family size effect is a type effect, reflecting the number of morphologically related words, not a token effect, reflecting their summed occurrence frequencies (Schreuder & Baayen 1997; De Jong, Schreuder, & Baayen 2000). Moreover, Moscoso del Prado Martín, Deutsch, Frost, Schreuder, De Jong and Baayen (2005) found that word frequency and family size were two independent factors in the prediction of reaction times to Hebrew words.

There are several reasons for believing that the effect of morphological family size reflects a high degree of connectivity among morphologically related words in the mental lexicon, and is hence a postlexical effect. The lexical representation of the incoming stimulus appears to propagate its activation through a morphologically organized network, such that the resulting high degree of activity leads to fast reaction times. Moreover, several strands of evidence suggest that the effect is morpho-semantic in nature and not purely form-based. Bertram, Baayen, and Schreuder (2000) demonstrated that removing the semantically opaque items from the morphological family increased the magnitude of the effect. Another argument against a form-based interpretation of the effect is that it also occurs for verbs that have undergone a vowel change and that are even homonyms of Dutch nouns, which should lead to the activation of the wrong family if the coactivation process were a pure form-based phenomenon. For instance, in a lexical decision experiment on verb forms like *vocht*, the past tense of *to fight* but also the homonym of the Dutch word for *moisture*, De Jong, Schreuder, and Baayen (2000) found that participants were sensitive to the morphological family size of the verb (*vechten, to fight*), despite the vowel mismatch.
and not to the family of the noun, which might have been activated on the basis of the word's form. Finally, the family size effect turns up in Hebrew, a language that makes use of a non-concatenative morphology. A particularly strong indication that the effect captures connectivity within the mental lexicon, at a morpho-semantically structured level, is Moscoso del Prado Martin et al.'s (2003) finding that the morphological family size of the Dutch translation equivalents for Hebrew words predicted the RLs to the Hebrew words and vice versa. Since these are typologically unrelated languages, this finding strongly suggests that the effect of morphological family size taps into a level of representation at which the meaning of words plays a role.

6.2 The spelling process

In contrast to what is the case for the study of reading, there has been very little attention for the study of writing. The studies that have been done within the tradition of psycholinguistics or experimental language psychology have generally been concerned with the process of spelling. In a review of this literature I recently sketched the main topics of investigation in this area and the obtained insights (Sandra 2007). The research on spelling has focused on two types of spellers: those who are still in the process of learning the skill and those who have already become very good spellers. Most studies have addressed issues that are related to the process of spelling development.

6.2.1 Spelling development

6.2.1.1 Stage models

In this research area, one important question has been whether children go through a natural series of stages, each corresponding to the use of a different type of knowledge, when grappling with the task of learning a written system that encodes spoken language (at least in alphabetic languages).

The idea that there might be stages in spelling development was not obvious from the start. Treiman (1992) points out that in the sixties of the previous century people set out from the assumption that learning to spell was a form of rote learning, i.e., basically memorizing letter sequences word by word. Charles Read, studying the invented spelling of children who have not learnt to read or write yet, was the first to argue that learning to spell is a creative activity, involving the induction of underlying correspondences between letter names and their alphabetic symbol. For instance, a child that develops its own spelling system before it goes to primary school might spell the word bay as ba, because b represents the first sound in its letter name and the sound [b] corresponds to the name of the letter a. His studies resulted in a book, Children's creative spelling (Read 1986), whose title emphasizes the importance of the learner's creativity.

After the publication of Read’s work several researchers started to realize that learning to spell involves complex mental processes and that the knowledge base behind these processes changes as the child grows older. Thus the process of learning to spell began to be conceptualized as a progression through a series of stages, until adequate spelling performance was reached. Gentry (1982), for instance, distinguishes five stages (we will not discuss them here), highlighting the fact that children naturally proceed from stages in which they think of spelling as the symbolic representation of the sounds they hear (the so-called semi-phonetic and phonetic stages) to stages in which they realize that there are orthographic conventions that take priority over simple sound-to-letter mappings. The general idea that children move from stages in which they encode sounds by letters/graphemes towards stages in which they rely on orthographic knowledge as well (principles and word-specific spelling patterns) has been embodied in the work of several researchers (Elwi 1997; Frith 1985; Seymour 1997).

Although the basic idea behind stage theories both sounds plausible and probably approximates the child’s general learning sequence, one should be careful when using the term ‘stages’. If stage models refer to a sequence of serially ordered steps in spelling development, which children must take in a fixed order, there are reasons to doubt these models on the grounds of their lack of flexibility. However, I hasten to add that, in my view, the researchers proposing stage models only attempted to mark the major developmental milestones, each being characterized by the fact that one type of spelling knowledge dominates the child’s spelling performance, but not necessarily to the exclusion of other types of knowledge. At any rate, a flexible view on spelling development is certainly required by experimental data. For instance, Martinet, Valdois and Fayol (2004) discovered that after only three months of learning to read and spell, children who had to write words to dictation were better at spelling HF words like French tête (head) than LF ones like trott (root) and more often spelled pseudo-words that sounded similar to words they could spell in correspondence with the orthographic pattern of these words (e.g., ROY, similar to sirop, syrup) compared to matched control pseudo-words (e.g., livoy). The observed frequency effect can only mean that children store fully specified orthographic sequences for words from the very beginning of their literacy education, alongside the phoneme-to-grapheme correspondences they are taught. The analogy effect also indicates that the spelling patterns of familiar words are stored in memory, so that they can be used as exemplars for spelling novel words on an analogical basis. The authors conclude that in the initial stages of learning to spell, both “lexical knowledge and general knowledge about sound-to-spelling correspondences are simultaneously acquired” (Martinet et al. 2004, B17).

6.2.1.2 Implicit learning of spelling principles

Another issue when studying the spelling development of children is whether they can acquire some spelling principles by sheer exposure, i.e., through implicit learning rather than explicit instruction, and, if so, which kinds of knowledge are acquired in such an implicit way. I will discuss only one example to demonstrate that children indeed learn more about spelling than what they are explicitly taught.
The orthographic encoding of morphology is a prime example. English relies on an orthography that encodes the morphological structure of words. For instance, one spells the plural nouns *deeds* and seats both with a *final* *s*, despite the pronunciation difference of their suffix, which shows that English spelling sometimes represents morphemic units instead of sound units. It turns out that children rapidly catch on to this morphographic spelling principle, before they are even taught about it. When using a spelling completion task, Freiman, Cassar and Zukowski (1994) observed that children had significantly less difficulty in spelling flat sounds that were the final sound of the stem in a morphologically complex word (*dirty*), than in a matched monomorphemic word (*duty*). A similar finding was reported by Freiman and Cassar (1996). First graders tend to omit the first of two successive consonant sounds in word-final position. However, this tendency was considerably smaller in words whose problematic sound was the final sound of the stem in a morphologically complex word than in a monomorphemic word. For instance, the letter *n* was omitted less often in words like *tunaed* than in words like *tunard*.

Recent research indicated that the causal factor behind children’s morphographic spelling was their morphological awareness, which does not necessarily imply a form of conscious awareness. Peter Bryant and his colleagues were the first to demonstrate this in a series of studies (Bryant, Nunes, & Bindman 2000; Nunes, Bryant, & Bindman 1997a,b, 2006). To measure children’s morphological awareness, they administered a word analogy test that could only be solved by discovering the nature of the morphological relationship between the two words in the example (e.g., *anger-angry, strength*-__). They discovered that children’s scores on this test predicted their success in spelling regular past tenses (e.g., *filled* instead of *filled*). Similar findings were reported in French by Sénéchal, Basque, and Leclaire (2006), who compared words whose spelling was phonologically transparent (*lac, lake*), morphologically motivated (*galop, silent final consonant, derived from *galloper*, *gallop*) or unmotivated and, hence, required memory retrieval (*tabac, tabacco, with silent final consonant*). Children performed better on words like *galop* than on words like *tabac*, even though both had a word-final letter that was not pronounced. However, words of the former type could be solved by relating the target word to a morphological relative in which the stem-final letter was pronounced, thus revealing its presence (*galop* => *galopper*). Again, a morphological awareness test showed that children who scored high on the words with a morphographic spelling (type *galop*) also had the highest level of morphological awareness. In this morphological awareness test, they had to initiate an example in which a word-final silent sound became audible by transforming the word into a morphological relative (e.g., *grit-grise, gray, blond*-__, *blond*).

The outcome of all these experiments certainly makes sense, but should not come as such a big surprise: only children who are able to consciously reflect on morphological relationships in an analogy test are able to come to grips with spelling principles that are based on morphological insight, which also requires an awareness of morphological relations between words. However, as with so many things that seem obvious once one knows them, it is not a trivial insight. This finding can moreover explain why some children keep struggling with the spelling of these words while others move on smoothly in their spelling progress.

### 6.2.2 Experienced spellers: What their spelling errors tell us

Research on the spelling of experienced spellers is hard to find in the literature. However, the findings regarding one particular problem nicely converge across languages, at least French and Dutch.

#### 6.2.2.1 The effect of relative homophone frequency

One of the most notorious spelling problems in Dutch is the spelling of a subset of regularly inflected verb forms. When thinking about it, this is strange. The spelling rules are descriptively quite simple (e.g., "add *d* to the stem when spelling the third person singular present tense", comparable to "add *-s* in English.") Moreover, there is a lot of emphasis on these rules during the entire education process (both in primary and secondary school). Finally, these errors are sometimes used to stigmatize people as being unconsiderate, sloppy, non-analytical, etc. Despite this, the errors persist, even in the written language (e.g., e-mails, but also final text versions) of highly educated people, like university students, journalists, professors, etc. So, what is going on with this aspect of language processing so easily details?

Some time ago we (Sandra, Friison, & Daems 1999) decided to identify the causal factor(s). First, it was not difficult to notice that the large majority of these errors involve homophones within the inflectional paradigm of a verb. For instance, the Dutch verb *worden* (become) is spelled as *word* in the first person singular present tense and as *wordt* ("add *-t*") in the third person. Although they have a different spelling pattern, they sound identical. The same applies to verbs like *gebeuren*, which have no stem-final *d* but are homophonous between their third person singular present tense (*gebeurt*) and their past participle (*gebeurd*). Using a speeded dictation task with 18-year-old students we found in two experiments (one focusing on the verb type *worden*, the other on the verb type *gebeuren*) that most errors were made on the lower-frequency homophone, irrespective of its grammatical function. Hence, we concluded that both inflected forms are stored in the speaker’s mental lexicon, that the spoken input (dictation) activates these two spelling alternatives, and that in the ensuing competition process the higher-frequency form has a higher chance of being selected under conditions of time pressure. The implication is that the lower-frequency form is more error-prone than its higher-frequency homophone, which is what we found.

These results demonstrate the interaction between two forms of memory systems: working memory and long-term memory, more particularly, the mental lexicon. Working memory is a limited-capacity system in which information can only be temporarily retained (e.g., a telephone number) and relatively simple calculations can be performed (e.g., arithmetic operations on relatively small numbers, grammatical operations like identifying the subject of a sentence). When this limited capacity system is overloaded, either
because the information that must be stored or processed exceeds the system's capacity, or because the time during which the operations must be performed is too short, the system will fail to achieve its goal. For instance, in one of our experimental conditions four words separated the inflected verb form from its grammatical subject. This was the condition where most intrusions of the wrong homophone occurred, quite likely because on many occasions the available time for identifying the sentential subject will have been too short for many participants. In the case of spelling a verb homophone a solution for this processing bottleneck is to fall back on long-term memory and select the higher-frequency spelling form. This is an unconscious process, but it minimizes the chance of making an error, as this form is the correct spelling pattern in the majority of cases (for similar results in a different task see Assink 1985).

Similar conclusions have been drawn from research on spelling errors in French (for a systematic comparison between the French and Dutch studies, see Sandra & Fayol 2003). Largy, Fayol, and Lemaire (1996) found that when a verb form and a noun are homophones, like the French verb form filtrer (from filtrer, to filter) and its homophonic noun filtré (fisler), writers more often misspell the verb form when the noun has a higher occurrence frequency than the verb. For instance, when writing a sentence like Le chimiste prend des liquides. Il les filtrer (The chemist takes the liquids. He filters them) there is a high error risk on the homophoneous verb form filtrer, which is sometimes misspelled as the noun plural filtrés. This finding of what could be called frequency dominance in homophone spelling is quite similar to our results in Dutch, although the French study focused on homophones crossing the boundaries between lexical categories whereas we targeted homophones within the inflectional paradigm of a single verb. Note that the above French example actually reveals the effect of two error sources: the tendency to spell the higher-frequency homophone and the tendency to spell a plural where a singular is expected, at least in certain contexts. Let's briefly discuss the latter error source.

6.2.2.2 The effect of words in the proximity

In previous work the same authors had found that limitations on working-memory capacity enhance the risk of an error type in which spellers make the inflected verb form agree with the nearest noun rather than with its preceding grammatical subject (Fayol, Largy, & Lemaire 1994). They called these errors proximity errors. For instance, French spellers may write Le chien des voisins arrive (the neighbors' dog arrives), spelling the plural verb form, instead of the correct Le chien des voisins arrive. This observation was made when participants not only had to listen to the sentence they had to write down but simultaneously had to count a number of clicks that were played together with the sentence. In other words, they spelled under conditions of working-memory overload. The fact that spellers rely on alternative sources of information when the capacities of their working-memory are exceeded ties in with our own finding (Sandra et al. 1999) on the effect of homophone frequency: this error source caused more errors when the verb form and its grammatical subject were separated by intervening words, i.e., when more working-memory resources were required.

The general conclusion from the research on spelling errors in experienced spellers is that a situation of cognitive overload causes spellers to rely on alternative information sources (relative frequency, the syntactic properties of a nearby word).

7. Spoken language processing

As I have announced in the introduction to this chapter my own research is situated in the domain of written language perception and production, which naturally implies that I am less well acquainted with all the issues that arise in the study of speech processing. Nonetheless, in what follows I will try to summarize the main issues that are discussed in the literature.

7.1 Speech perception

When listening to a speaker whose language we share (especially the native language) we seldom have the impression that we are going through a difficult process. It is as if the words are 'right there', one next to the other. What most of us fail to realize is that this is our phenomenological experience of what the speaker has said, i.e., the end-product of speech perception. Obviously, when all the processing work has been done, one will detect no problem. It is not difficult to demonstrate this without an experiment. If you do not happen to speak, for instance, Finnish, just listen to a Finnish speaker and try to find the words in the speech stream. You will not be able to. One might reply that this is quite obvious, as one cannot identify words that one does not know. So, let's make the task easier: identify where one word ends and another one begins, without knowing the identity of the words. You will not be able to either. You will fail because the speech signal radically differs from your perception of it as a listener in a familiar language: a grammatically ordered sequence of words. In contrast to this output representation, which gives rise to the phenomenological experience of one word sitting next to the other, the speech signal itself is a seamless, continuous flow of sound. This discrepancy is a major problem in the domain of speech perception: how is it possible that a continuous sound stream is experienced as a sequence of neatly separated words? Trying to unravel the processes that link this type of input to this type of output is one of the main challenges for researchers on speech perception. Our daily experience that linking these two is not a demanding task at all (whereas it is a puzzle to understand how it can be accomplished) underscores an essential aspect of practically all language skills: speech perception is an automatic process. We cannot prevent it from happening when someone starts speaking to us and we focus on the speaker's voice, i.e., there is no way to bring it under conscious control.
There are many important issues in this research. I will focus on two of them. How do listeners succeed in discovering the phonemes in the speech stream? How do they succeed in segmenting speech into words? Because we perform these operations effortlessly several thousands times a day, the questions sound strange. However, precisely because we perform these tasks so fluently, the seemingly obvious turns into an intricate scientific puzzle when taking a closer look at the acoustic signal we start from.

7.1.1 Finding the speech sounds

As mentioned above, speech is a continuous sound stream. What has been said about the absence of boundaries between words equally applies to the speech sounds in words. These are not arranged as a neat concatenation of discrete sounds. Quite the contrary, due to the phenomenon of coarticulation, properties of neighboring sounds affect each other. When uttering a word, the speaker’s articulators move in a continuous manner through the oral cavity, describing the optimal trajectory that is needed to articulate the sounds contained in the word. As a result, the articulatory position for the pronunciation of a particular sound will adapt itself as much as possible to the articulatory positions of the surrounding sounds. For instance, the articulatory position of a back vowel will slightly differ when it is pronounced after a back consonant (cook) than after a front consonant (took) — the articulatory ‘distance’ is smaller in the former case — resulting in two acoustically different vowels. For the same reason, a front consonant will be articulated with a slightly different form of the oral cavity when a back vowel (took) follows than when a front vowel follows (thick). In other words, an optimal articulatory trajectory is the result of a mutual adaption of the speech sounds’ articulatory positions to each other, which involves both anticipation and perseveration. The result is that the same phoneme leaves a different fingerprint on the acoustic signal depending on its neighboring sounds, which implies that it cannot be recognized by looking for an invariant realization. The obvious problem, then, is how to recover the identity of phonemes amidst this large degree of phonetic variability. As Kraljic and Samuel (2005) state: “For the last half century, the defining issue in speech perception research has been the ‘lack of invariance’ problem” (p.167), referring to seminal research papers by the Liberman group at the Haskins Laboratories (Liberman, Cooper, Shankweiler, & Studdert-Kennedy 1967; Liberman, Harris, Hoffman, & Griffith 1957). Understandably, “A large body of early speech research was dedicated to the search for invariant acoustic properties of phonemes. This enterprise may be reckoned unsuccessful with respect to its ultimate goal” (Mitterer & Cutler 2006, p.772).

7.1.1.2 Categorical perception

When the identity of phonemes that make up words is continuously masked by the effects of coarticulation, how then do we extract these phonemes from the signal? One plausible suggestion is that a process of normalization takes place between the phonetic input signal and the phonological representation that will be used to activate lexical representations.

The assumption behind this idea is, of course, that there are relatively stable and fixed phoneme representations, which are in turn used to activate representations in the mental lexicon. A process that seems the ideal candidate for normalizing the phonetic input is the process of categorical speech perception. Categorical perception of speech sounds refers to our ability to cut up an acoustic continuum, defined by variation on a speech-related parameter, into two discrete categories.

For instance, the syllables /ba/ and /pa/ differ along the parameter of voice onset time (VOT), i.e., the time at which the vocal cords start to vibrate. This is due to an articulatory difference between the consonants: whereas a /b/ is a voiced sound, pronounced with vocal cord vibration, a /p/ is voiceless. Hence, for the syllable /ba/ vocal cord vibration starts on average some 60 ms earlier than for the syllable /pa/, where it starts only at the onset of the vowel. With a speech synthesizer one can create a number of /ba/ and /pa/ allophones in-between these prototypical instances, all separated from each other by a constant VOT-interval. A typical finding is that two syllable tokens with VOTs of, say, 0 ms and 20 ms are both perceived as the same syllable /ba/, exemplars with VOTs of 40 ms and 60 ms as identical instances of /pa/, and instances with VOTs of 20 ms and 40 ms as instances of /ba/ and /pa/, respectively. Note that each pair of syllables has a 20 ms VOT difference. These results indicate that the mid-position between the prototypical ‘human’ /ba/ and /pa/, i.e., a token with a 30 ms VOT, is treated by our perceptual system as a sharp boundary between two categories, despite the underlying physical continuum: one containing variants (allophones) of /ba/ and one containing variants of /pa/. It has been suggested that this capacity for categorical perception makes it possible to tolerate a lot of variance in the speech signal. Categorical perception would ignore this variability and discover the phoneme category behind the phonetic token.

7.1.1.2 Perceptual learning of category boundaries

However, since the turn of the millennium, serious doubts have been cast on the basic premise behind the concept of categorical perception, i.e., the notion that there are stable phoneme representations in the speech processing system. Several studies have indicated that these representations are much more flexible than thought before and dynamically adapt to speaker-specific input. Most of this work has been done by Anne Cutler and James McQueen, both at the Max Planck Institut für Psycholinguistik in Nijmegen, and Dennis Norris from Cambridge University. The main conclusion emanating from several of their papers is that the representations onto which sounds are mapped may be phonetic in nature (rather than true phonemes in the linguist’s sense of the word) and that they can quite flexibly adapt to the articulatory properties of the speaker.

In an auditory lexical decision task Norris, McQueen and Cutler (2003) presented a set of 20 words whose final fricative had been replaced by an ambiguous sound somewhere in between the /ʃ/ and /s/ (henceforth: /ʃ/). In an auditory lexical decision experiment one group of participants heard words whose final /ʃ/ had been replaced by this ambiguous
sound (wildeit, i.e., wilen, chicory), whereas they also heard words with clear instant of final /s/ sounds (naaldas, pine forest). In another participant group the /s/ sounds were made ambiguous but the /f/ sounds were clear (naaldas and wilfe). At the end of the experiment, the participants had to classify the final sound in syllable tokens that were situated at the [ej]–[es] continuum as either an instance of /f/ or /s/, whereas the tokens were no clear instances of either sound. It turned out that participants who had previously heard the ambiguous sound in words in final /f/ classified more syllable tokens as belonging to the /f/ category, whereas those who had heard the ambiguous sound in words ending in /s/ classified more tokens as instances of /s/. Importantly, this shift of the category boundary on the phonetic continuum required that the ambiguous sound was presented in a lexical context. When it was presented at the end of nonwords, the effect did not occur. The authors’ conclusion is that hearing an ambiguous sound in a lexical context makes it possible to learn which familiar sound it corresponds to, whereas this is impossible in the case of nonwords. They argue that their effect represents a case of perceptual learning, which occurs very rapidly – only 20 words with ambiguous endings were presented in a lexical decision task containing 200 items, divided equally between words and nonwords. In other words, the phonological categories that speech sounds are mapped onto before lexical access can take place are not rigid but can rapidly adapt to speaker-specific properties. This can explain our ability to quickly adapt to the peculiarities of somebody’s accent.

In subsequent studies the authors reported corroborating evidence for their findings and their significance for models of speech perception. McQueen, Norris, and Cutler (2006) showed that this mechanism of perceptual learning is automatic. It also occurred when people did not have to make decisions on the experimental words and nonwords preceding the categorization task, but only had to count them. The bottom-line is again that adaptation to peculiar properties of a sound is a natural and automatic process.

In an important study McQueen, Cutler, and Norris (2006) demonstrated that this process not only affects category decisions but also lexical decisions on words that had not been presented in the training phase. Their rationale was that a shift of the category boundary should create a bias in the perception of words belonging to a minimal pair, when the discriminating phoneme was replaced by the ambiguous sound: participants should perceive more often the member containing the phoneme that had been replaced by the ambiguous sound in the training phase. In the experiment they made use of Dutch words and first administered the same auditory lexical decision task, including words with ambiguous consonants, as in their previous experiments. In the test phase the participants had to perform a cross-modal priming lexical decision task: they heard a word through a headphone but had to make a lexical decision on the letter string that simultaneously appeared on a computer screen. They found that ambiguous auditory items like doof, which can be both doos (box) and doof (leaf) in Dutch, facilitated responses to the visual word dos only for participants who had initially heard words in which the ambiguous sound replaced a word-final /s/. In contrast, the same auditory items only facilitated visual lexical decisions on the other member of the minimal pair (doof) for participants who had initially heard words where the lexical context dictated an /f/ interpretation of the ambiguous sound. Importantly, the words in the test phase had not been encountered in the training phase of the experiment, which is strong evidence that the process of perceptual learning does not only affect the specific words in which the ambiguous sound occurred (item-specific learning) but has an effect that generalizes across the lexicon. This, in turn, is an argument that the effect does not occur at the lexical representations themselves but rather at the prelexical representations of the sounds they contain. The interface between the sensory signal and the lexical representations subserving spoken word recognition quickly adapts to speaker-specific properties, which allows listeners to generalize these properties to all words uttered by this speaker.

In another paper Cutler, Eisner, McQueen & Norris (2006) tested another prediction made by the concept of speaker-specific tuning of prelexical phonological categories: such a perceptual learning effect should be relatively stable, so that listeners can use it again when encountering the same speaker. In line with this prediction they found that listeners who had shifted their boundary between two sound categories on the basis of an ambiguous sound in the context of words maintained this shift twelve hours past exposure.

Kraft and Samuel (2005) also demonstrated the speaker-specificity of this type of perceptual learning. In their experiments an auditory lexical decision task, in which they used ambiguous sounds that were situated in-between an /s/ and a /f/, was followed by a 25 minutes’ distraction task and two sound classification tasks: one in which the stimuli were spoken by the same voice as the items in the first task and one in which another voice was heard. They made three important observations. First, participants who had been lexically guided in the first task to identify the ambiguous sound as /f/ not only made more /f/-responses on the /s/-/f/ speech continuum than those who had been trained to interpret the sound as /s/, the magnitude of this effect was larger 25 minutes after the learning phase than immediately following it. Second, this effect of perceptual learning could be ameliorated, but only when two conditions were met: (i) when the same voice pronounced the /s/ and /f/ sounds correctly in the intervening phase between lexical decision and sound categorization (unlearning), in a task requiring participants to select pictures on the basis of spoken instructions, and (ii) when the same voice also pronounced the items in the sound categorization task at the end. When a different voice used the correct sounds in the intervening task, no unlearning took place: the boundary shift remained in the sound categorization task, at least if the items were pronounced by the same voice as in the lexical decision task (where perceptual learning had taken place). Third, when the items in the sound classification task were spoken by a different voice than the one in the lexical decision task (male vs female), the perceptual learning effect disappeared, provided that the difference in basic acoustic properties between the two voices was sufficiently large.

Taken together, the authors argue on the basis of these results that the perceptual learning effect is relatively stable and is speaker-specific. On the one hand, the effect of
perceptual learning did not disappear when a different voice pronounced the sounds correctly during an intervening task but the same voice pronounced the items in the lexical decision and sound categorization tasks. On the other hand, it did disappear when different voices pronounced the items in the lexical decision and sound categorization tasks. Apparently, listeners use subtle acoustic cues to identify a voice and use these to activate speaker-specific settings for the boundary between two sound categories when they have to categorize sounds.

Quite recently, Cutler, McQueen, Butterfield and Norris (2008) reported that even sublexical information like phonotactic constraints can cause a boundary shift between categories. Presenting nonwords whose phonotactics only allowed one phonemic interpretation of an ambiguous sound made the category boundary shift in the direction of that sound. For instance, when listeners were presented with an ambiguous sound in-between /f/ and /s/ in contexts like *rul* they learned that the sound could only be an /f/ and not an /s/, according to English phonotactic rules, and shifted the category boundary to cover more /f/ allophones. *rul* was observed for participants who were exposed to nonwords like *rul*, which could only be interpreted as *rul* on phonotactic grounds.

To conclude: coarticulation makes it difficult to cut up the acoustic signal of a word into a sequence of discrete sounds, but there is reason to believe that sound 'segments' in a word, even though they are not clearly delimited in the speech signal itself, are mapped onto prelexical phonetic/phonemic categories by relying on the process of categorical perception. Cutler, McQueen, Norris and co-workers published a number of studies in which they demonstrated that these categories are quite flexible and that listeners automatically adjust the boundaries between two acoustically similar phonemes on the basis of limited perceptual learning, which is triggered by degraded sound input. This adjustment can be driven by lexical or phonotactic knowledge but must be guided by information sources that enable the identification of the ambiguous sound. For that reason, the effect of perceptual learning is absent when this sound occurs in nonwords.

7.1.1.3 Feedback from lexical representations to phoneme representations?

The notion that lexical information can have an effect on phoneme representations is an issue that turns up in another crucial debate in speech perception research: is spoken word recognition entirely a bottom-up process, in which later processing stages can have no effect on previous ones, or is it an interactive process, making it possible that higher-level representations can provide feedback to lower-level representations, thus facilitating their identification? The crucial issue in the present context is whether there is feedback from the lexical to the phoneme level.

When replacing the sound in a word by an equally long portion of, for instance, white noise, Warren (1970) found that listeners did not hear there was something wrong with the word and could not identify which sound had been replaced by something else when asked to do so. It was clear that 'something' restored the missing phoneme in listeners' perception. Using signal detection techniques Samuel (1981, 1987) argued that this effect of phoneme restoration was caused by the lexical context on the basis of lexical-to-phoneme feedback. As a result, the effect was not the outcome of educated guessing on the basis of the available sounds, but was truly perceptual in nature, a perceptual illusion. However, the issue is far from settled.

Samuel (1996) admitted that some studies did and others did not find a perceptual effect of phoneme restoration. Using a sensitive technique he could demonstrate that the effect is real, though fragile. One year later, Samuel (1997) reported a series of experiments in which he strongly argued in favour of the interactive position, i.e., feedback from the lexical to the phonological level of representation. In this paper he used the adaptation technique. Technically, a participant in such an experiment has to perform a phoneme categorization task (e.g., on the /g/-/k/ continuum), is then exposed repeatedly during the so-called adaptation phase to a stimulus (the adaptor) one of whose sounds matches an extreme on the sound continuum (e.g., gift), and is then again confronted with the same items as in the first sound classification task. The effect is a boundary shift between the two sound categories. For instance, if the adaptor is voiced fewer test items will be identified as voiced than in the first classification task.

After demonstrating that using long words as adaptors (involving the /b/-/d/ continuum) caused an adaptation effect he showed that a (reduced but quite significant) adaptation effect also emerged when the target sound was replaced by white noise but not, crucially, when it was replaced by silence. Hence, stimuli that create an auditory word illusion behave like real words in the adaptation paradigm. In line with this effect, he found adaptation effects with nonwords that were highly similar to words (hence, caused lexical activation) but not with nonwords that were dissimilar to words. Finally, he found that the adaptation effect did not differ between words (gift and kiss) and nonwords (kift and giss) as adaptors. This, he argued, demonstrated that the adaptation effect from word adaptors causing perceptual illusions had its focus at the level of phoneme representations and was, hence, mediated by the lexical level through a feedback mechanism. Accordingly, the conclusion is that the effect of phoneme restoration is a perceptual effect: participants have the same perceptual experience, whether the target phoneme is present or replaced by, for instance, white noise.

For another series of experiments leading to the same conclusion see Samuel (2001).

A radically different point of view is adopted by James McQueen and colleagues. In two recent papers (McQueen, Norris, & Cutler 2006; McQueen, Jesse, & Norris 2009) they strongly argue against feedback from the lexical to the phonemic level. McQueen et al. (2009) argue against a phenomenon that has been taken as solid evidence in favour of lexical feedback to the phonemic level in online speech perception. The so-called lexical compensation for the fricative-stop coarticulation. Let's first explain the concept of fricative-stop coarticulation, which dates back to the experimental work by Mann and Repp (Mann & Repp 1981; Repp & Mann 1981, 1982). These researchers found that an unambiguous fricative like /s/ or /f/ determines the interpretation of a following ambiguous sound.
The importance of the rhythmical heuristic

A major strategy, which has received a lot of attention in the literature, is the so-called rhythmical strategy. When acquiring their language, children seem to adapt to the rhythmical properties of their spoken language input rather than adopting a universal segmentation strategy. The first to demonstrate sensitivity to these rhythmical properties of the native language were Jacques Mehler and his co-workers in France (Mehler, Dommergues, Frauenfelder, & Segui 1981). They showed that participants detected a sound sequence like *pan or *pal that was embedded in a target word faster when that sequence matched the first syllable in the presented word, irrespective of the phoneme length of the sequence. For instance, in French words like *palace (palace) the sound sequence *pa was detected faster than the sequence *pal; in contrast, in words like *palmier (palm tree) the sound sequence *pal was detected faster than the sequence *pa. The authors’ explanation was that French is a syllable-timed language, i.e., its syllable boundaries are very clear, and that such a language invites listeners to cut up the speech stream in a series of syllables, which are then used to recognize words.

Research in other languages confirmed this language-specific segmentation strategy. From a rhythmic perspective, English radically differs from French. French has clearly defined syllables whereas English is plagued by the problem of ambiasyllabicity (how do you syllabify mislabeled as mis-tile or mist-tile?)? However, English has another reliable rhythmic property: the majority of its words begin with a strong syllable (i.e., non-reduced vowel). Cutler and Carter (1987) used two computerized dictionaries to find out that the number of lexical English words (i.e., omitting function words) beginning with a stressed syllable was three times as large as the number of words beginning with an unstressed syllable. Additionally, words with a strong initial syllable have a higher occurrence frequency in the language. Together, these two phenomena are responsible for the fact that about 85% of the word tokens one hears in English speech have a strong initial syllable. Cutler and Carter actually counted the number of word tokens with this metrical pattern in a 190,000 word token corpus of spontaneous conversation and indeed found that 90% of the tokens had a word-initial strong syllable.

This lexical–statistic property of English allows for an ideal word segmentation heuristic: each strong syllable is a potential word boundary. Cutler and Norris (1988) reported evidence that native speakers of English indeed rely on this heuristic. They embedded an existing one-syllable word (e.g., mint) at the beginning of a nonword and compared a condition in which the second syllable was also stressed (e.g., mint-tayve) with a condition in which this syllable was unstressed (e.g., mint-tish). The participants had to spot the embedded word. They were faster when the second syllable was unstressed. According to the authors, two stressed syllables induce a segmentation within the nonword because each stressed syllable is a potential word beginning. This makes it necessary to integrate the first phoneme from the second syllable (t from tayve) with the first syllable (min) to spot the target (mint), which is not required for items with a weak final syllable, as these are not segmented. This finding supports a stress-based segmentation strategy for English.
This result was corroborated by Cutler and Butterfield (1992), who created word sequences that were in conflict with the typical distribution of strong and weak syllables in English, in an attempt to show that such sequences would induce missegmentations on the part of the participants. For instance, in a sequence of three words with the atypical weak-strong pattern (e.g., conduct ascends uphill) a stress-based segmentation strategy would predict that the strong word-final syllables trigger a segmentation attempt within the words and, hence, signal the beginning of a possible word at the wrong place, leading to misperceptions of the word sequence. This is indeed what happened: often participants reported having heard sentences like the doctor sends her bill, a sentence that contains three words with a strong initial syllable, as one would expect if English listeners treat each stressed syllable as a possible word beginning.

Several other studies in different languages were in line with the general idea that listeners 'tune in' to the language-specific rhythmic properties. For instance, Otake, Hatano, Cutler, and Mehler (1993) found that the Japanese mora, which determines the rhythmic pattern in the Japanese language, was used as a cue for segmenting a sentence into candidate words. With respect to Dutch, Vroomen, van Zon, & de Gelder (1996) also reported evidence for stress-based segmentation. They reasoned that it was not only important to compare languages with different rhythmic structures but also languages with identical structures (such as English and Dutch). The concept of rhythm-induced segmentation predicts that such languages should show similar properties. And indeed, this turned out to be the case. Dutch, which resembles English in its preference for words with strong syllables in initial position, behaved as English had done in the earlier studies by Cutler and colleagues (although the authors suggested a second information source for segmentation, see below).

Taken together the evidence forms a coherent picture. However, what would happen when speakers of one language (say, English) have to perform a monitoring task in a different language (say, French)? Would they automatically catch on to the rhythmic structure of the language or would they adhere to the heuristic they have learnt for their native language? A theoretically important distinction is, moreover, whether the answer to this question will differ for monolingual and bilingual speakers. The almost obvious hypothesis is that monolinguals will be unable to switch to the other-language pattern (which they are unaware of) whereas bilinguals will swiftly adapt to the rhythmic structure of each language. However, the answer to the question was not so simple, which explains why the experimental results did not only find their way into the psycholinguistic literature but also in the highly renowned journal Nature.

Cutler, Mehler, Norris & Segui (1988) came up with a brilliant idea: to tease apart the role of the language and the role of the listener in sentence segmentation one could confront monolinguals, who have adopted a language-specific segmentation strategy, with a language that requires a different type of segmentation. Will such listeners stick to their familiar strategy or adopt a strategy that is tailored to the new language? So, they presented the knowledge of French, and the English materials from an earlier study to a group of French students with a limited knowledge of English. The task was the same kind of monitoring task used in previous studies (e.g., press the button if you hear a word that begins with pal). The researchers found that monolingual speakers do not capitalize on the rhythmic structure of the new language. They cling to their own one, such that the English students monitoring French materials showed the same pattern of results as English students monitoring English materials. Also French speakers used their syllable-based segmentation strategy on English, even though it did not fit the structure of that language. They, too, showed the same pattern of results as the group of French listeners who were confronted with their native language. In other words: the native language imposes a segmentation strategy on listeners, which they rigidly apply to another language, even when its rhythmic structure does not lend itself to this strategy (see Cutler, Mehler, Norris, & Segui, 2002, for a similar set of results).

The same authors (Cutler, Mehler, Norris, & Segui 1989) published a letter in Nature, in which they raised the same question: Is the segmentation strategy determined by the listener or by the rhythmic properties of the linguistic input? – but used a different population this time: bilinguals who mastered French and English almost equally well and were accepted as native speakers in both speech communities. Still participants were forced to identify their favourite language by replying to the question ‘if you had to lose one of your languages to save your life, which would you keep?’ (p. 229). The answer to this question was used to classify them as English-dominant or French-dominant. One might expect that balanced bilinguals like these have become tuned to the rhythmic structure of the two languages. However, this is not what the results showed.

English-dominant speakers used their metrical heuristic for both languages. Despite their extremely high skill in French they apparently could not use its syllable-based segmentation strategy. In contrast, the French speakers used the syllable-based strategy for French and the metrical strategy for English. The authors conclude that, despite being a very good bilingual, one language remains the dominant one: English-dominant bilinguals could not apply a syllable-based strategy to French. In order to explain the dissociation between English-dominant and French-dominant bilinguals, the researchers appealed to the concept of markedness. They argued that the syllable-based rhythmic structure of French is a marked (exceptional) case, since many languages share the metrical structure of English. In their view, someone who is good at using the marked structure (French-dominant speakers) can easily switch to the use of the unmarked structure but someone who is good at using the unmarked structure (English-dominant speakers) is not able to adapt to the marked, syllable-based structure of his other, near-native language. They claim: "Our present study suggests that at this level of language processing there are limits to bilingualism: a bilingual speaker has one and only one basic language" (p. 229).

The prominent role of rhythmic structure in cutting out the words from a spoken sentence seems to be deeply rooted in very early language development. In the early eighties of the pre-
attractive to very young infants younger than two months (the average age in each group was below 90 days). Over a headphone they played either a CVC (i.e., phonotactically possible syllable, e.g., pat) or a CCC (i.e., phonotactically impossible but pronounceable because the medial was a fricative, e.g., tʃp) to the infants. They used the habituation procedure by monitoring the sucking rate of the infant during continuous presentation of the same syllable, treating the sucking rate as an index of the infant’s interest. What interested them was how the sucking rate would be affected when switching to a different auditory stimulus, made by swapping the initial and final consonants (e.g., pat became tap and tʃp became pʃt). The percentage increase in the sucking rate following this switch was larger for true, CVC syllables than for CCC sequences. They concluded that, in French, the syllable is a basic unit in speech perception. 

Even though the Bertocci and Mehler (1981) study highlighted the importance of the syllable, which ties in with the later discovered important role of the syllable in the segmentation of French, Jusczyk, Cutler and Redaraz (1993) targeted young children’s sensitivity to the rhythmic structure of their language more directly. They were curious whether young children were already sensitive to the stress-based structure of English, more particularly, the dominance of the strong-weak (SW) pattern in word-initial position. Children aged 9 months heard either a list of SW or a list of WS English words through a loudspeaker. Examples of such lists are: plant, faller, donor, comet, neighbour, butter, final, stalkyard, gentle, sinus, rhcessu (SW) and comply, befall, condone, comfort, parmaude, abart, define, restore, resent, assign, caprice (WS). A flickering light was situated at the loudspeaker position when the list was being played. The rationale of this so-called head-turn procedure is that, as long as the child keeps watching the red light she/he is apparently interested in the words coming out of the loudspeaker, whereas turning its head away is a sign of a loss of interest. The authors found that 9-month-olds were more interested in SW lists than in WS lists, which is in line with adults’ preference for a metrical segmentation procedure in English. In contrast, 6-month-olds did not show this preference, ruling out an interpretation that the difference was due to the SW patterns being acoustically more attractive than to a preference for the dominant stress pattern in the language. Finally, when the same word lists were again played to 9-month-olds but the words had gone through a low-pass filter, such that only suprasegmental information like stress remained, the children again preferred the SW stimuli. This evidence clearly demonstrates a very early sensitivity to the basic rhythmic structure of the language, a type of knowledge that will later turn out to be quite useful when they have to segment sentences in words.

7.1.2.2 The contribution of lexical competition

Despite the fact that relying on the language’s dominant rhythmic structure is a big help in finding word boundaries in a spoken sentence, it cannot be sufficient to find all words. Recall that in English, for instance, about 10% of the word tokens that speakers daily encounter do not match the dominant stress pattern, i.e., are WS words. Hence, other knowledge sources must be involved as well. One of these sources is the principle of lexical competition. Since thousands of words in the vocabularies of languages like English are constructed from a small set of some 30 phonemes, it is statistically unavoidable that many words will contain identical phoneme sequences, even phoneme sequences that occur as words in the lexicon but do not function as a word (morpheme) in the longer word in which they are embedded. For instance, the phonological sequences for the words scan, can, and candle all appear in a spoken word like scandal McQueen and Cutler (1992) found an average of 2.6 embeddings per word, when only counting embeddings whose boundary coincided with that of the carrier word (e.g., scan was counted as an embedding in scandal, but can and candle were not). Considering all English polysyllabic words McQueen, Cutler, Bates and Norris (1995) found that 84% contained embeddings Cutler, McQueen, Bayen and Drexler (1994) found that 71% of the words in a real-speech corpus of English contained embedded words whose syllable boundaries were aligned and 92% when these boundaries were ignored.

When the spoken input is so ambiguous with respect to the identity of the words, how can listeners ever identify the right words in the speech stream? The solution that all current models of spoken word recognition employ is the notion of competition in a lexicon in which multiple candidates are activated by the spoken input (McQueen & Cutler 2001). The idea is simple: activated word representations inhibit the activation level in the representational set of other activated candidates with a strength that is proportional to their frequency. In the case of scandal there will be parallel activation of all words whose phoneme structure is present in the signal: scan, can, candle and scandal itself. The competition among these word candidates will be proportional to their frequencies but the lexicon will also (have to) take into account whether the phonological sequence of the ‘winning’ candidate aligns completely with the input. Thus the competition process will result in the recognition of the word scandal because all other word candidates cover only part of the input and no combination of two candidates aligns with the input (e.g., scan+candle does not match scandal).

7.1.2.3 The possible word constraint

Research by Norris, McQueen, Cutler and Butterfield (1997) has shown that the competition process is constrained by an additional principle: the Possible Word Constraint (FWC). This constraint was formulated to account for the observation that, for instance, people find it harder to recognize the word apple in a spoken input like fapple than in one like vafapple The FWC states that listeners will not segment a word such that a vowelless residuum is left between an embedded word and the nearest word boundary. Indeed, in the example above, f could not be a word, which makes it impossible that apple is an internal constituent, whereas vaf might in principle activate a word or be sufficiently similar
to stored words, such that *apple* must temporarily be treated as a possible word candidate. Cutler, McQueen, Jansoniussen, and Bayerl (2002) made a lexical-statistic analysis of the English and Dutch lexicons, as assembled in the CELEX database, which contain all words in the language, drawn from a representative corpus of texts together with their occurrence frequency (among many types of orthographic, phonological, and syntactic information). The English database comprises 70,000 words and is based on a corpus size of 17.9 million tokens, whereas the Dutch database has a total of 280,000 words based on 42.4 million tokens. In line with previous studies the authors found that only a very small minority of words did not contain an embedded other word (English: less than 29%, Dutch: less than 1%). The crucial part of the study was, of course, whether taking the PWC into account to constrain the process of lexical competition would be beneficial. The answer for both languages was a clear yes. In English, the ratio of embeddings leaving a vowelless residu (PWC violations) to those leaving a possible word was 1.58 to 1. Including the occurrence frequency of these words in the calculations brought the ratio to 2.69 PWC violations to 1, which amounts to a saving of 73% (i.e., 2.69/(2.69+1)).

A similar picture emerged in Dutch: 1.28 PWC violations to 1 non-violation in a type-based count, 1.54 violations to 1 in a token-based count, which amounts to a saving of 61% (i.e., 1.54/(1.54+1)). Not that the figures in Dutch are smaller because this language makes abundant use of compounding, which diminishes the proportion of vowelless residues.

The authors conclude that taking the PWC into account would remove the majority of spuriously embedded words like *ample* in *ample sample*.

7.1.2.4 Reliance on statistical regularities

Another information source that is likely to guide the parsing of a sentence into words is the language user’s knowledge about the co-occurrence probabilities of phonemes, or the probability that phoneme x will be followed by phoneme y (transitional probability, TP). In the extreme case, TP will have the value zero, which means that these two phonemes cannot co-occur in a word and that, hence, a word boundary can be postulated (or a morpheme boundary, as in *handbook*, but we will not go into this issue now).

However, not all phoneme pairs have a TP of 0. Some are more likely than others and unlikely co-occurrences could be used to postulate a possible word boundary in-between the phonemes. As it happens, it has been shown that even 8-month-old babies already rely on such statistical information to chop out the words in a continuous speech stream. Saffran, Aslin, and Newport (1996), in a study reported in *Science*, reported two ingenious experiments with 8-month-old infants. They reasoned that on average the TP between two phonemes will be lower when they straddle a word boundary than when they co-occur within a word. In principle, these lower TP values could be a useful heuristic for postulating word boundaries in the continuous speech signal. Saffran et al. wondered whether 8-month-old infants would already be sensitive to TP information.

They made up four nonsense words, each consisting of three syllables (*tupiro*, *golabu*, *bidaku*, and *padoli*). The babies were exposed to a continuous speech stream of 2 minutes in which these four words were repeated in random order. A speech synthesizer with a female monotone voice generated the speech stream at a rate of 90 words per minute, such that each word appeared 45 times over the course of the 2 minutes’ training phase. The generated signal contained “no acoustic information about word boundaries, resulting in a continuous stream of coarticulated consonant-vowel syllables, with no pauses, stress differences, or any other acoustic or prosodic cues to word boundaries” (p. 1927). In the test phase the children were exposed to two of the words that had been embedded in the speech streams (*tupiro*, *golabu*) and to two nonwords, consisting of syllables that had never co-occurred in the training phase and were taken from at least two different words (*lapiku*, *tilado*). Accordingly, the TP value between the syllables in the words was 1 whereas it was 0 in the nonwords. Using the headturn procedure (flickering red light) the researchers found that the infants showed significantly more interest in the novel items than in the familiar ones (a difference of about 1 sec).

The second experiment was even more challenging for the infants. They were again familiarized with four three-syllable words (*pabika*, *tibudo*, *golatu*, and *diaropi*) in the learning phase. Each word was presented 45 times and it followed each of the other three words 15 times, such that the TP value between each word-final syllable of a word and each word-initial syllable of the other three words was 0.33. In the test phase the babies heard two familiar words (*pabika*, *tibudo*) and two nonwords, each of which was made by recombining the syllables of the untested familiar words, more particularly, by taking the final syllable of one of these words and stringing it together with the first two syllables of the other one (*tudaro*, *pigola*). Notice that the TP distributions in the words and nonwords made the discrimination task considerably more difficult than the previous one: the two TPs in the words were 1, those in the nonwords were 0.33 and 1. Yet the babies again showed significantly more interest in the nonwords than in the familiar words (again, a difference of about 1 sec).

Saffran et al.’s observations demonstrate the power of young children’s inductive learning through the estimation of statistical regularities in the data, even within an interval as short as two minutes. They can use this knowledge to discriminate familiar syllable sequences from unfamiliar ones, suggesting that also adult listeners may rely on statistical knowledge about the phonological material that co-occurs in the sentences of a language for segmenting the spoken language stream into words. As an aside, albeit not a trivial one, the authors finish their paper by situating their findings in the context of the language innateness debate: “some aspects of early development may turn out to be best characterized as resulting from *innately biased statistical learning mechanisms rather than innate knowledge*” (my emphasis). Their own, further research along the same lines more and more suggests that they may have an important
point (see the introductory contribution to this volume for more discussion on their work: Perspectives on language and cognition).

7.1.2.5 Reliance on subtle acoustic cues

Finally, there is evidence that subtle phonetic differences can assist listeners in their segmentation task. Spinelli, McQueen, and Cutler (2003) studied the effect of liaison in French, more particularly in contexts where the liaison gave rise to lexical ambiguity, as in *le der nier oignon* (the last onion), where the process of resyllabification makes the first word's final consonant form a combination with the next word, making it homophonous to the word *rogon* (kidney), as in *un demi rogon* (half a kidney). In cross-modal priming experiments participants first heard a French sentence, containing a phoneme sequence that was lexically ambiguous due to the possible liaison between a word-final consonant and a subsequent word-initial vowel (*C'est le dernier oignon/rogon*; it is the last onion/kidney). At the offset of the spoken sentence the target word *oignon* or *rogon* appeared on a computer monitor and a lexical decision had to be made. Facilitation effects were found on the vowel-initial words when they had been intended by the speaker (e.g., *oignon in le dernier oignon*), despite the fact that resyllabification resulted in a homophonous consonant-initial competition (rogon). Facilitation effects on non-intended vowel-initial words (e.g., *oignon in le dernier rogon*) were not absent but weak. How could listeners tell the difference between the intended and non-intended words in a situation of homophony? The authors measured the length of the consonant that was responsible for the ambiguity and found that its duration was shorter when the vowel-initial word was intended (between 10%-18% difference). These findings suggest that the acoustic signal contains subtle acoustic cues (like durational differences) that listeners can use to make the segmentations that match the speakers intentions. The authors refer to earlier findings by Gow and Gordon (1995), who found that a word like *lips* was activated in two-word contexts like *two lips* but not in one-word contexts containing the phonetic sequence of the word (*tulips*). These authors, too, observed that the duration of the /t/ was longer in word-initial position, which is in line with Spinelli et al.'s finding that the word-initial /t/ was longer when the consonant-initial word was intended. In short: subtle acoustic differences can help listeners in their segmentation attempts as well.

7.2 Speech production

Needless to say, speech production is the pivotal language ability. Without our ability to speak, there would simply be no need to explain speech perception, as there would be no speech. Similarly, there would be no need to study reading and writing (spelling), as the written code for words is secondary to what is a species-specific skill, due to our human genetic set-up (which is clearly not equivalent to the claim that there is an innate set of grammatical principles, as is the case in Chomskyan linguistics). Still, despite the crucial nature of speech in human verbal communication, studies of this process represented the minority of psycholinguistic research for a long time, although this is no longer the case. Speech production research can be divided into two eras: one in which researchers attempted to infer the design underlying this skill from speech errors and one in which experimentation has taken over to study this process.

7.2.1 Speech error research

One of the biggest problems to study speech production processes has to do with the fact that the stimulus that initiates the process is an internal one, more particularly, a message that somebody wants to convey to somebody else. The only observable aspect of speech is its end product, whereas all relevant mental processes have been performed: the spoken sentence. This strongly contrasts with research on reading and speech perception, where the experimenter has full control over the stimuli that are presented for processing and where careful manipulations of factors that are deemed important lead to the creation of material sets that can easily be related to, for instance, RT and error measures. The problems with manipulating an internal stimulus is without any doubt an important reason why early researchers of the speech production process focussed on the observable output. At first sight, it would seem that spoken sentences are uninteresting from the perspective of somebody who is interested in the mental architecture that makes the generation of these sentences possible. This architecture cannot be read off the spoken sentences. Or can it? The early investigators of speech production must be credited for realizing that, indeed, this architecture leaves different kinds of fingerprints on some spoken sentences, and that good ‘detective’ work could lead them to the identification of the hidden process that was responsible for each type of fingerprints. The fingerprints in case are speech errors, those instances when the process of speech production fails and results in an error.

The logic of collecting naturally occurring speech errors and systematically studying them to arrive at a theory of speech production has a straightforward logic. Any system has a structure (be it a car, a thermostat, a text processor, or a speech production system) that seriously constrains the possible operations that the device can perform. Due to this very same structure it also sets limits on the types of errors that it can make. For instance, a thermostat’s components have been designed in such a way that its function is to keep the room temperature at the programmed temperature. When we come home and observe that the temperature is much too high, we will infer that the thermostat has a defect. Theoretically, the cause is probably easily identified. At the level of its internal structure the thermostat must have a device that constantly measures the room temperature and compares it to the value set by the user. When the temperature is too low, this device gives a signal to the central heating that it must switch on. When the temperature is too high, the device sends an off signal, such that the central heating stops producing heat. Now, several things can be broken when the thermostat does not function: the device itself or its communication with the central heating (the central heating itself can be excluded because it is too warm, etc.).
so it works fine). When the device itself is broken, the defect can be either a failure in the component that measures the room temperature, a failure in the comparison with the preset temperature, or (trivially) empty batteries. When this device works well, the defect is in the communication with the central heating system, which may be situated either at the beginning or the end of this transmission: the thermostat may no longer transmit the signal or the central heating may receive the signal but no longer respond to it. Now, I know nothing about thermostats or central heating systems but it is clear that this kind of reasoning may help diagnose the problem and that any of these possible error-causing components must be part of the structure of the thermostat.

The purpose of this digression is to show that when a device malfunctions (makes an error) a systematic analysis may help you uncover its internal structure and the way it operates. You will not be able to sketch the design behind the device at a microscopic level (e.g., the microcircuits in the thermostat) but you will be able to discover the large-scale internal architecture that it must have in order to function the way it does. Moreover, the nature of its errors will suggest components in its design that are likely to malfunction. The bottom-line is this: the structure of a device allows certain errors and makes others impossible (the design of a car makes it possible that it will miss a curve in the road but makes it impossible that it will fly; an airplane can fly but cannot drive backwards). Through a systematic study of the errors that occur one can arrive at the identification of what has gone wrong inside its internal architecture and thus reconstruct this architecture.

This is the rationale behind the study of speech errors by the early practitioners (Piirto, 1971, 1973; Garrett, 1975; Nooteboom, 1967; Shattuck-Hufnagel, 1979). Assemble a corpus of speech errors, analyze them systematically by trying to identify what may have gone wrong, make a taxonomy of different types of malfunctions, and, finally, integrate all thus identified speech production components that have been derived from the errors into a model of human speech production. This is an example of clean reasoning in a time when current-day experimental procedures were unavailable to study the speech production process more directly.

I will not dwell too long on speech error research and soon turn to the modern techniques. However, it would be unfair to pass over the influential literature on speech errors that dominated the research on speech production for at least two decades and still leads to the publication of papers. The pioneer on the use of speech errors to discover the design of the human speech process was Victoria (Vicky) Fromkin (1973), who wrote a book whose title is self-explanatory: Speech errors as linguistic evidence. Still in what follows I will focus on Merrill Garrett's contribution to this literature.

The model that Garrett (1975) designed already contains many ingredients of the current-day models, even though it is a box-and-arrows model that was typical of that time. At the message level the speaker decides what to say about whom or what. This content selection sets a whole series of processes in motion, which will ultimately lead to the articular movements that transmit the message to the listener. Crucially, in Garrett's model there are two grammatical levels: the functional level and the positional level. At the former the abstract grammatical information on the appropriate lexical items is retrieved and the syntactic relations among these items are determined (e.g., subject, objects). Among the abstract grammatical information of a word belongs its word class (noun, verb, ...), but also its semantic information. Semantic substitution errors like He rode his bicycle tomorrow (yesterday) originate at this level.

At the functional level the phonological form has not been retrieved yet. That only happens at the positional level. At this level, a series of processes takes place. The phonological form of the lexical items is retrieved, as can be derived from phonological errors like his immoral soul (immortal). Additionally, the grammatical information retrieved with the lexical items at the functional level is now used to plug the phonological forms of the words into their appropriate slots in the syntactic frame of the sentence, slots that are also labelled for syntactic category. It is at this stage that word exchange errors occur, like I must let the house out of the cat instead of I must let the cat out of the house. As the example shows, words can exchange positions when the mechanism that allocates word forms to positions in the syntactic frame is confronted with two words sharing the same syntactic category and two sentential slots marked for this same category (which is why most of these exchange errors involve two words of the same syntactic category, i.e., nouns generally exchange with nouns, verbs with verbs, etc.). Note that the occurrence of word exchange errors reveals the existence of such a process of word-to-sentence-position assignment. The distance over which such exchange errors occur has been used to estimate the number of words speakers plan ahead when uttering a sentence.

Such exchange errors reveal a remarkable phenomenon: when two words exchange places in the sentence the morphosyntactic information associated with these words does not migrate together with the words. For instance, in the speech error 'I'd hear that if I knew it' (correct: I'd know that if I heard it) the words hear and know have exchanged places but the grammatical encoding of past tense has not: the form heard did not move, only its infinitive. Its past tense marker remained at its proper place and was applied to the migrated word know; that moved into the slot where the verb hear should have been. The take-home message is that there is a processing level at which the phonological form of the lexical words is already available and linked with its position in the sentence frame, but the morphosyntactic information that is linked to this same positional slot is still represented in an abstract format (e.g., past tense, plural suffix). Only in a later processing stage, this abstract information will be translated into a contextually appropriate phonological form, i.e., adapted to the phonological form inserted into the slot. A bound morpheme (or, rather, its underlying morphosyntactic specification) that remains stuck at its original sentence position whereas the word that it should be attached to migrates to a different sentential slot and is replaced by a different word, is known as a morpheme stranding error.
Besides retrieving the phonological form of words and putting the words in their correct linear order, the positional level finally contains a process that puts the phonemes within a word in their correct order. The latter process occasionally details as well, producing a so-called phonological error. Such errors are misplacements of phonemes, which implies that the phonological structure of a word is not an undivided whole, but rather a sequence of nearly ordered phonemes, which can also be rearranged. Consequently, in a sentence context they can migrate away from the correct position within their word. There are many different types of phonological errors: anticipation errors, like leading list instead of reading list, where the first phoneme of the first word has been replaced because the speaker anticipates the first sound of the following word. There are perseveration errors, like watching rabbits instead of waking rabbits, where a phoneme is repeated in a later slot, replacing the correct phoneme. Phonemes can also exchange places, as in some swimmers sunk instead of some swimmers sunk. Phoneme exchanges in word-initial position that cross word boundaries, as in left hemisphere for left hemisphere or God is my loving shepherd, are known as a separate variety. They are called Spoonerisms, after the reverend Spooner, who is said to have made this type of speech error quite frequently.

Once the order of the words within their grammatical frame and the phonemes within their word frame have been specified the sentence is ready to be sent on to the phonetic level, where the phonetic realization of the abstract phonological structure takes place. At this level, errors that have been made when arranging the phonemes can cause phonetic accommodations, just like erroneous word insertion errors at the positional level caused allomorphic accommodation or morpheme standing errors. For instance, the error an unkey's mandle (intended: a monkey's mantle) shows that a phoneme movement error at the previous processing makes the indefinite article appear before a vowel-initial word. A phonetic process turns the intended form of the article, a, into the form an, thus adapting it to the erroneous phonetic environment.

Finally, this phonetic sentence representation is encoded into an articulatory representation of the sentence, i.e., a specification of the motor gestures that are required by the articulators in the vocal tract to actually utter the sentence.

7.2.2 Experiments on speech production

Despite the unmistakable value of speech errors for models of speech production, their easy accessibility in naturalistic conditions (any conversation on any day), researchers in the eighties of the previous century have started to devise techniques that make it possible to tap into the processing of speech production in real time, i.e., while it is going on. Indeed, there are a few disadvantages to the study of speech errors that can be compensated for by the use of online techniques. First, they are offline data. They are the product of the speech production process and, hence, only indirectly shed light on the underlying processes that make speech production possible. Second, one might get the impression from the speech error literature that speech production is a highly error-prone process and that errors abound. The contrary is true. Levelt (1992) refers to a study by Garnham, Shilcock, Brown, Mill, and Cutler (1982) in which they analyzed a corpus of spoken text consisting of 200,000 word tokens and counted 86 lexical selection errors and 105 other slips of the tongue, which means that in the process of speech production, we make, on average, only one error per 1,000 spoken words. This testifies to the high reliability of this process and at the same time underscores the fact that speech production models that are based on speech error data depend on only 1% of the data in natural speech. Needless to say, a scientific model that is built on what is the exception turns the risk of being wrong or at least incomplete, even though the underlying rationale behind speech error analysis is plausible and hard to refute.

For these reasons, it became increasingly clear to researchers that the error-based analyses and models had to be complemented by data obtained in paradigms where correct speech was studied while taking place in real time. At the end of the eighties a large, comprehensive book appeared, in which a speech processing model was presented, discussed in great detail, and supported by a large number of empirical facts. This standard book, Speaking (1989) by Levelt, which still functions as a benchmark in the discipline, marks a definite turn in the approach to the study of speech production. A new era had begun.

7.2.2.1 The picture-word interference paradigm

A technique that soon became one of the most popular online research techniques for tapping into online speech production was the picture-word interference paradigm. In this paradigm, participants are shown a picture and asked to name it as fast as possible (speech production). However, a word is either written on the picture or presented auditorily, and participants are asked to ignore this word. The rationale is based on our knowledge that it is impossible for experienced readers or listeners to ignore a word. With respect to reading, the Stroop effect (Stroop 1935) has demonstrated this by showing that we cannot ignore a word when naming the colour of its letters: it is hard to say “red” to the word green printed in red letters. By manipulating the nature of the word that is presented with the picture, researchers have a probe to study the processing sequence during picture naming, i.e., word production.

In an influential study Schriever, Meyer, and Levelt (1990) used this technique to demonstrate that the process of lexical access is a two-stage process: a first stage in which semantic information becomes available at a time when there is no access to phonological information yet and a second stage in which phonological information becomes available after the semantic information has been retrieved and the processor has no longer access to this previous representational level. When, for instance, the picture of a dog had to be named, presenting a semantically related distractor word like cat auditorily 150 ms before
picture onset resulted in delayed naming times (interference) relative to a control condition
(where e.g., the name \textit{lp} was presented), whereas a phonologically related distractor like
\textit{fog} had no effect. In contrast, when presenting the two distractors and their controls
150 ms after picture onset, the phonologically related distractor delayed picture naming
latencies whereas no effect was measured of the semantic distractor. So, when the distrac-
tor’s meaning and the meaning of the visualized object reach the semantic system about
simultaneously, as is the case when the word is presented shortly before picture onset,
semantic interference results. When the phonological representations of both the distrac-
tor word and of the picture name reach the level of phonological representation simultane-
ously, as is the case when the word is presented shortly after picture onset, phonological
interference results. These results clearly indicate that the process of lexical access is a two-
step procedure: one during which so-called lemmas are retrieved, when only semantic and
grammatical word information become available, and one during which so-called lexemes are
retrieved, when only phonological word information becomes available.

Levelt, Schriefers, Vorberg, Meyer, Pechmann, & Hovinga (1991) raised a somewhat
different question with respect to the stages of semantic and phonological access during
the process of lexical activation. The earlier observations that semantic and phono-
logical interference occurs, suggest that word meanings and the phonological forms of
words are both organized in semantic and phonological networks, respectively. When
a speaker must name a picture, the meaning of the picture’s name will cause spreading
activation through the semantic network. A question that arises is whether the speaker
immediately selects the semantic representation corresponding to the picture before
any activation can propagate to the phonological representational level or not. Is it pos-
sible that semantic relatives are coactivated without concomitant activation of their
phonological representations?

In their experiments the authors showed a picture and presented participants with
an auditory word at different intervals after picture onset. These auditory words belonged
to one of four types: (i) a name that was semantically related to the picture (e.g., \textit{goat} for a
picture of a sheep; to test whether the semantic representational level was still accessible at
this stage), (ii) a name that was phonologically related to the name of this semantic relative
(e.g., \textit{goal} related to \textit{goat}; to test whether coactivated semantic representations activated
their phonological representations, which in turn coactivated their phonological neigh-
bours), (iii) the name of the picture itself (\textit{sheep}), and (iv) an unrelated name (\textit{door}).

The authors tested two types of relationship at the ‘semantic’ level: associative related-
ness, which would cause coactivation due to the frequent co-occurrence of the words
(\textit{sheep} coactivating \textit{wool}) and true semantic relatedness, in which case the two words
belonged to the same semantic category (e.g., \textit{sheep} coactivating \textit{goat}). If coactivation at
the semantic level was transmitted to the phonological level before the ultimate selec-
tion of the meaning of the picture name, the presentation of an auditory stimulus that
is related to the name of either the associate (\textit{wood} for \textit{wool}) or the semantic relative
(\textit{goal} for \textit{goat}) should considerably delay naming times, as the result of interference. This
turned out not to be the case, which made the authors conclude that coactivation within
the semantic network does not ‘leak’ into the layer of phonological representations. This
suggests that the selection of a word meaning is already finished at the representational
level of the lemmas, where its phonological form is not yet available, before the activa-
tion is fed forward to the next representational layer, where the words’ phonological
forms are stored.

It should be clear from the foregoing that the online techniques that are used in
current-day experiments allow for insights at a more fine-grained level than the analysis
of speech errors. At the same time, researchers adopting these techniques do not reject the
validity of speech error data. On the contrary, one of their goals is to design processing
models that cannot only account for the RT data but also for the speech error data.

7.2.2 Implicit priming

Instead of asking people to name pictures as a tool for studying the process of word pro-
duction one can also ask them to produce a word that they have previously learnt as an
associate of a probe word. For instance, when you learn the associative pair \textit{bottle-lamp},
you will be able to say \textit{lamp} upon hearing the prompt \textit{bottle}. Obviously, using associative
pairs to study the process of speech production does not serve the purpose of studying
the associations themselves. Rather, by manipulating properties of the associations that
participants have to learn one can achieve insight into the pieces of information that are
relevant for word production. When all associative pairs have something in common and
this shared property makes it easier to produce the associate, the researcher will be able to
infer that this property plays a role in the course of word production.

This was the basic idea that inspired the initial work of Aniye Meyer and led to the
introduction of a new technique in the study of speech production. Levelt, Roelofs, and
Meyer (1999) refer to it as the technique of implicit priming and point out that this tech-
inique has the advantage of imposing fewer constraints on the experimental materials.
Indeed, the word that one wants the participants to produce no longer has to be pictur-
able. Meyer addressed several theoretical issues regarding the processes underpinning
speech production.

She (1990) wondered whether the phonological encoding of a word’s syllables,
one of the components in Levelt’s model of speech production (Levelt 1989; see also
Levelt et al 1999) proceeds in a serial left-to-right order or in parallel. She presented
people with lists of three paired associates, which they had to memorize until they
could swiftly produce the second word when prompted with the first one. The so-called
homogeneous condition contained three pairs of semantic associates in which each
second word started with the same syllable (e.g., \textit{single-loner, place-local, fruit-latus};}
signal-beacon, priest-beadle, glass-beaker; captain-major, cards-maker, tree-maple). In
the so-called heterogeneous condition participants studied the same materials but these
were rearranged such that all first syllables in a triple differed (single-loner, signal-beacon,
captain-major; place-local, priest-beadle, cards-maker; fruit-lotus, glass-beaker, tree-maple).
Meyer found that response times were significantly faster in the homogeneous condi-
tion than in the heterogeneous one. However, in bisyllabic words, she found only facilita-
tion when the common syllable across response words was the first one, not when it
was the second one (murder, pound, boulder). Moreover, she found that in trisyllabic
words, a shared second syllable did facilitate response speed, but only when the first
syllable was shared as well. She concluded that participants who know that the first
(or first two) syllables of the response are always the same in the set of studied words
can start preparing the process of phonological word encoding and complete it when
they receive the prompt. However, they cannot initiate this process when they do not
know the identity of the first syllable. Hence, she showed that words are phonologically
encoded in a strictly left-to-right order.

Meyer (1991) used the same technique to focus on the syllable constituents, the onset
and the rhyme, the onset being the consonant(s) preceding the vowel and the rhyme being
the vowel and all subsequent consonants (if any). For monosyllabic words she found an
advantage in the homogeneous condition when the three responses shared their onset but
not when they shared their rhyme. This outcome also obtained for bisyllabic words. More-
over, for the latter word type she found larger facilitation in the homogeneous condition
the more phonemes of the first syllable were shared by the response words. There was facilita-
tion with shared onsets, more so when both the onset and the vowel were shared, and
the largest facilitation when the whole syllable was shared. Meyer concluded that the pho-
nological encoding of the syllable is also an incremental process, the onset being encoded
before the rhyme.

With the same technique Roedofs and Meyer (1998) showed that the fully phonologi-
cally encoded word is a representation in which the syllabified phoneme sequence is associ-
ated with a metrical frame that specifies the stress pattern across the word's syllables.
This conclusion was based on two observations: (i) the homogeneous condition facilitated
responses to the probe words when all response words shared the word-initial segments
but also had the same metrical frame, i.e., the same number of syllables with the same
stress pattern, independently of whether they had the same number of vowels and con-
sonants; (ii) the homogenous condition did not facilitate responses when the responses
shared their metrical pattern but did not have the same word-initial segments. In order
to have a headstart and being able to start preparing the response one needs both segmental
and suprasegmental information.

The implicit priming technique has thus shown that the process of phonological
encoding in speech production is a highly incremental process. This incrementality was
also demonstrated by Wheelton and Levelt (1995) in experiments where participants had
to monitor their internal speech production.

7.2.2.3 Producing multiple words

Whereas much reaction time research on speech production has focused on single word
production, studies from the mid-nineties of the previous century started to focus on the
processes that are involved in the production of multiple words.

In Meyer's study (1996) participants were shown two pictures next to each other on a
computer monitor. In one type of experiment they had to name the pictures in a phrase of
the type "the N₁ and the N₂" (e.g., de boom en de vlag – the tree and the flag), in another
they had to make a short sentence (e.g., de boom staat naast de vlag – the flag is next
to the tree). The purpose of the study was to find out at which processing level (semantic or
phonological) the production process had arrived for each word at the moment of response
initiation. To that end, two experiments (one for each of the above mentioned response
types) made use of semantic distractors for either the first picture (straat for boom – bush
for tree) or the second picture (wapen for vlag – weapon for flag), which were presented at
picture onset or 150 ms before it. In two other experiments the auditory distractors were
phonologically related the first picture (boor for boom – drill for tree) or the second picture
(vlas for vlag – flax for flag).

There was no difference between the two response types (conjunction vs short sen-
tences), for neither distractor type. Semantic distractors for both nouns delayed responses,
whereas phonological distractors facilitated responses only for the first picture name.
Meyer concluded that at the time participants began to speak they had retrieved the lemma
for both pictures, which contains the semantic and grammatical information, but only the
phonological form for the first picture. Thus she demonstrated that one production process
was still underway while the other one was finished.

Meyer, Slederink, and Levelt (1998) also studied the naming of pairs of pictures. They
wanted to know whether participants' viewing times for pictured objects could shed light
on the planning stages involved in speech production. On each trial participants saw a pair
of objects whose names were either both HF or LF names. Both their naming latencies
and their viewing times on the left object were measured (objects were named from left
to right). They found a significant frequency effect on both naming latencies and viewing
times. However, when participants did not have to name the objects but only to discrimi-
nate them from non-existing objects, the frequency effect on viewing times disappeared.
The authors conclude that if, as is often assumed, frequency determines the accessibility
of the word's phonological form, the finding that frequency affected viewing times on the
left object only in a naming task indicates that the participants' eyes do not leave this
object before they have phonologically encoded it. Meyer's (1996) study mentioned in the
previous paragraph further suggests that before speech initiation participants will already
process the second object till they have retrieved its semantic properties (lemma level) but not its phonological form.

7.2.2.4 Models of speech production

The big divide between two types of speech production models is the debate about their degree of interactivity. Should an adequate model of speech production be considered as a sequential, feedforward model, in which later processing stages have no impact on previous ones? Or should it be considered as a highly interactive model in which later stages feed back their information to previous stages, thus helping these earlier processes when they are ‘in trouble’ by feeding them the information assembled at later processing stages?

Ten years ago Levelt (1999) wrote a paper in which he stated that the best-known interactive model, the one developed by Gary Dell (1986), and the best-known feedforward model, the one developed by himself and collaborators, stem from two different research traditions – speech error research and chronometric speech production studies, respectively. He emphasized the commonalities between the models rather than stressing their differences and made the optimistic claim that “these research traditions have begun to emerge in recent years, leading to highly constructive experimentation. Currently, they are like two similar knives honing each other. A single pair of scissors is in the making” (p. 223).

I will not enter into this debate here, because it could form the topic of a separate chapter on speech production. However, it is worth emphasizing that Dell’s interactive model indeed originated in the tradition of speech errors and contains properties that are especially intended to account for a particular frequent type of error: so-called mixed errors, which evidence that neighboring representations have been activated at both the semantic and phonological level of word representation. For instance, a speech error like rat when the intended word is cat, suggests that both a related semantic and phonological representation have become active. The shared phonemes /a:/ and /t/ with the target word cat and the feedback of the activity in phonological representations to the semantic level (and back again) accounts for the emergence of a mixed speech error.

The model developed by Levelt (1989) and laid out in his book Speaking: From Intention to Articulation, has its roots in chronometric experiments and, as mentioned before, consciously so. The percentage of speech errors is so low (1%–2%) that the starting-point of this model is non-erroneous language use. Using an analogy one could say that the best way to study how a device works (e.g., a radio, a car) is by studying its normal function, not by studying its defective behaviour. Note that this does not discount my earlier statement that the structure of a device determines its function and that by studying faulty function one can derive restrictions on the way it is internally structured. And indeed, Levelt and colleagues readily admit that their model, too, (which has been implemented in the computer model WEaver by Ardi Roelofs 1997) must be able to account for the patterns observed in speech error corpora.

Some researchers still adhere to highly interactive processing models whereas others try to limit interactivity as much as possible. Still others make an attempt to unite the two, claiming that both discrete and interactive processing are probably involved but that it remains to be seen which stages interact and which behave as modules (see, for instance, Rapp & Goldrick 2000).

8. Conclusion

In this chapter I have made an attempt to achieve two goals. My first goal was to characterize psycholinguistics as a separate discipline, with its own set of goals, methodologies, and techniques. My second goal was to sketch some of the major topics in past and (especially) current psycholinguistic research on all four language skills: reading and writing, speech perception and speech production. Thus the two dimensions behind our language skills have been covered: the mental processes and representations involved in oral and written language (modality) and the mental processes and representations involved during language perception and production (language user’s activity). As mentioned before, this review does not pretend to be complete. For instance, it largely focusses on lexical processing. Nonetheless, by attempting to achieve the above goals and making the discussion as concrete as possible by describing actual experiments, I hope readers have achieved a good view on what psycholinguistics is about.

References


