

# Psychology of Language

WILLEM J. M. LEVELT

|     |                                             |     |
|-----|---------------------------------------------|-----|
| 9.1 | WHAT IS LANGUAGE?                           | 152 |
| 9.2 | HISTORY                                     | 153 |
| 9.3 | PRODUCING AND UNDERSTANDING SPEECH          | 153 |
| 9.4 | ACQUIRING LANGUAGE                          | 160 |
| 9.5 | READING AND WRITING                         | 162 |
| 9.6 | CONCLUDING REMARKS – LANGUAGE AND THE BRAIN | 163 |

The psychology of language is the study of how we produce and understand language, how we read and write, and how we acquire these skills. These will also be the main topics of this chapter. Their discussion will be preceded by some remarks on language evolution and by two short sections on language as a generative system and on the history of the discipline. Language is the species-specific communication system of *homo sapiens*. As a product of evolution, it provided selective advantage to our ancestral hunter-gatherer clans that roamed the African savannahs. Much has been speculated about what exactly provided the evolutionary cutting edge of language. Dunbar (1996) stressed its role in social bonding. Primates regulate their bonding largely through grooming, and the time spent on grooming – up to 20% of the waking day – is directly related to group size. Early *homo sapiens* should have groomed some 40% of the day in order to maintain social cohesion in clans of typically one to two hundred members. Language is obviously an attractive alternative. Conversation can be conducted among several participants, who can be outside each other's tactile reach. More importantly, conversation has always been an

indispensable means of exchanging information about one's intentions, beliefs, fears, hopes, joys, that is about mental states. Such a device must have attained substantial significance for the one primate that has a 'theory of mind'. Sole among the great apes, our species has developed the ability to attribute mental states (such as intentions and beliefs) to our conspecifics in order to explain and predict their behavior (Tomasello & Call, 1997), and we understand our own behavior in similar terms. This ability allows us to monitor the state of the social network in which we participate and to act accordingly. Language is a marvelous device for exchanging information about mental states. How often do we use expressions such as 'Peter wants to leave', 'Sue hopes to come'? But mental states tend to be recursive and that is easily grasped by saying such things as 'Mary thinks that Peter wants to leave', or 'I don't believe that Mary thinks Peter wants to leave'.

The evolution of human culture is unexplainable without assigning a central role to language. The ability to converse makes it possible to share information of almost any kind. We use language to exchange useful experiences, to

transmit traditional skills to our children, to plan joint actions of various kinds, etc.

Though the selective advantages of having a language are obvious, we will never be able to explain the present structure of natural language from the history of selective pressures that shaped it. That history has been irrevocably lost. Quite different from what Wundt (1900) supposed to be the case, all existing languages are highly complex systems; there are no peoples with simple languages. The long co-evolution of culture and genetic endowment has universally done its work. What kind of complex system is a natural language?

### 9.1 WHAT IS LANGUAGE?

The major function of language is to share information about whatever is relevant or dear to us. This 'aboutness' is one of the core features of language. Language is referential: 'the dog' (usually) refers to a particular dog; 'a dog' refers to one that has not yet been introduced in the conversation. 'I' refers to the present speaker. By saying 'yesterday' you pick out the day before today. Most words have meanings that can be used to refer to persons, objects, states of mind, or to even quite abstract states of affairs, such as 'democracy'. Language is, in addition, predicative. It is used to say something about these referents: 'the dog is limping', 'I am thirsty', 'yesterday was my birthday'. Predication is language's core business. When you say 'Tom drove his new car from Paris to Rome', you are predicating something about an agent (Tom), namely that he is driving some theme (his new car) from some source localization (Paris) to some goal localization (Rome). Agent, theme, source, and goal are the 'arguments' or 'thematic roles' of this proposition. There are other thematic roles that can be expressed in language, such as 'experiencer' ('John' in 'John loves flowers') and recipient ('John' in 'Mary gave John some flowers'). The predicate plus the thematic roles assigned form the sentence's 'argument structure'.

Closely connected to predication is modification. When you say 'his new car' you are modifying or further specifying 'car' by 'his' and 'new'. You can easily turn predication into modification to make reference more specific: 'the dog that is limping' or 'the limping dog'.

The versatility of language is largely due to its generativity. We can talk about just everything and to do so, we produce ever new utterances. That would not be possible with a fixed set of expressions, such as 'how are you today' and 'beg your pardon'. Rather, language allows

us to combine and recombine a fixed set of elements to produce ever new composite expressions. For instance, every language has a small, fixed set of consonants and vowels ('phonemes'). They can be combined in particular ways for building new words: 'tran' is a possible word in English, but 'rtan' is not. Every language has a set of morphemes (some 20,000 for the English of a normal native speaker). Most of them are simple, meaningful words, such as 'dog', 'follow', 'green', 'you'. But others are meaningful elements that cannot stand alone as a word: 'un-' (as in 'undo'), '-s' (as in 'dogs'), '-ed' (as in 'walked'), etc. A large part of a language's generativity resides in combining and recombining morphemes, to create more and more complex words; there is no end to what we can do: 'vaccin' ⊢ 'vaccinate' ⊢ 'prevaccinate' ⊢ 'prevaccination' ⊢ 'anti-prevaccination' ⊢ etc. This is called morphological or lexical productivity. There is, finally, syntactic generativity. In speaking, we combine words to create ever new phrases and sentences. There is no obvious limit here either. For every sentence we can create a more complicated one that contains it: 'Peter wants to leave' ⊢ 'Mary thinks Peter wants to leave', 'I don't believe that Mary thinks Peter wants to leave' ⊢ etc. Some languages, such as Turkish, do most of their work with lexical generativity. Others, such as English or Chinese, capitalize on syntactic generativity.

All languages have a generative system for dealing with the semantics of predication and modification, all have generative phonology and morphology for coining new words, and all have generative syntax for creating ever new sentences. Linguists have also noticed surprisingly universal properties for each of these generative systems. There is no semantic system without agents, patients, or recipients. There is no phonological system without phonemes and syllables. There is no syntactic system without verbs. However, as the archive of analyzed languages expands, linguists discover that many of their hypothesized universals turn out to be strong tendencies at best. Languages have evolved to solve similar problems in quite different, often idiosyncratic ways. Most European languages, for instance, have terms for 'left' and 'right'. Their semantics of space makes a major divide between what is left and what is right of the speaker; this is called 'deictic perspective' (Levelt, 1996). But other languages partition space in quite different ways. Speakers of the Guugu Yimithirr in Australia, for instance, use an 'absolute perspective', roughly equivalent to our north-south. They would happily say the equivalent of 'I have a fly on my north cheek' and change it to 'south cheek' when reorienting

by 180 degrees. Syntactically, all European languages have phrase structure; they chunk words together into meaningful, contiguous phrases, such as 'the big elephant', and phrases can become part of larger phrases, such as 'on the big elephant'. But many Australian languages have no obvious phrase structure. Words that belong together semantically or syntactically are spread out all over the sentence, with other words intervening. Also Latin belonged to this class of free word-order languages. Finally, it is not even the case that all of the world's 5–10 thousand languages are spoken languages. There also exist natural sign languages. Most of them developed in deaf communities. These languages are as complex and versatile as spoken languages.

## 9.2 HISTORY

Language and its use has always been a topic of great fascination. Its systematic study goes at least back to the sixth century B.C.E., when the great Indian linguist, Panini, devised the first systematic theory of the sound structure of language. This sophisticated system of phonology was orally transmitted to our present day. Explicitly studying the psychology of language is, however, a much more recent enterprise. The term 'psychology of language' ('Psychologie der Sprache') was coined mid-nineteenth century by the German linguists Steinthal and Lazarus. Evolution theory was in the air. Linguists already had a good understanding of how the Indo-European languages had evolved over the last two to three millennia. But they were entirely in the dark (as we still are) about the original natural causes that gave rise to language. Their guess was that language would naturally arise in the human mind. The origin of language would ultimately be explainable from psychological principles. 'Fortunate advances in linguistics presuppose a developed psychology',<sup>1</sup> Steinthal wrote in 1855, but alas, such a psychology did not exist. The 'psychology of language' was invented *pour besoin de la cause*. It was Wilhelm Wundt who took up the challenge. He went all out to lay the psychological foundations for a theory of language origins. Although that theory has long been abandoned, the spin-off of his two-volume *Die Sprache* (1900) has been substantial. With the greatest psychologist of his time dedicated to the psychology of language, it had become a respectable discipline, which easily adopted the work of many others, such as the work by Galton, Ebbinghaus, Marbe, and Watt on verbal memory

and word associations, Binet's efforts to measure vocabulary size, Meringer and Mayer's (1896) classical study of spontaneous speech errors, Broca and Wernicke discovering the left brain's involvement in the production and comprehension of speech, Huey's experimental studies of reading, Clara and William Stern's (1907) first thoroughly data-based study of children's language acquisition and so on. The psychology of language had become established.

The European tradition, continued by Bühler, Claparede, Vigotsky, Piaget, and many others has always been a mentalistic one. Language, after all, is a mental device. In North America the psychology of language increasingly eschewed mentalistic theories or explanations. Dominant behaviorism considered language behavior to be a system of conditioned reflexes, an opinion forcefully defended in the ultimate monument of that tradition: Skinner's *Verbal Behavior* (1957). North America needed the so-called 'cognitive revolution', vigorously stirred by Noam Chomsky, George Miller, Jerome Bruner, and others, to return to a mentalistic psychology. The mood of change had its terminological effects. The new generation of cognitive psychologists coined 'psycholinguistics' for what had previously been called 'psychology of language'.

The American cognitive revolution exerted a major influence on late-twentieth century psychology of language. Language came to be considered as a biological endowment of the human mind, the generative system unfolding quickly during the first years of life, stimulated by linguistic interaction with caretakers and peers. The mature, generative system came to be studied as part and parcel of a complex information processing system, which performs the high-speed feats of speaking and language comprehension. Entirely new approaches were developed to study the implementation of the component linguistic processes in the human brain.

## 9.3 PRODUCING AND UNDERSTANDING SPEECH

The systematic experimental analysis of speech production and comprehension has produced growing insight in what is called the 'functional architecture' of language, the network of psychological processing components involved in the generation and comprehension of language. Figure 9.1 is a schematic diagram of this processing network. I will call it a 'blueprint' to stress that it is as much a way to summarize the plethora of research findings, as a guide to

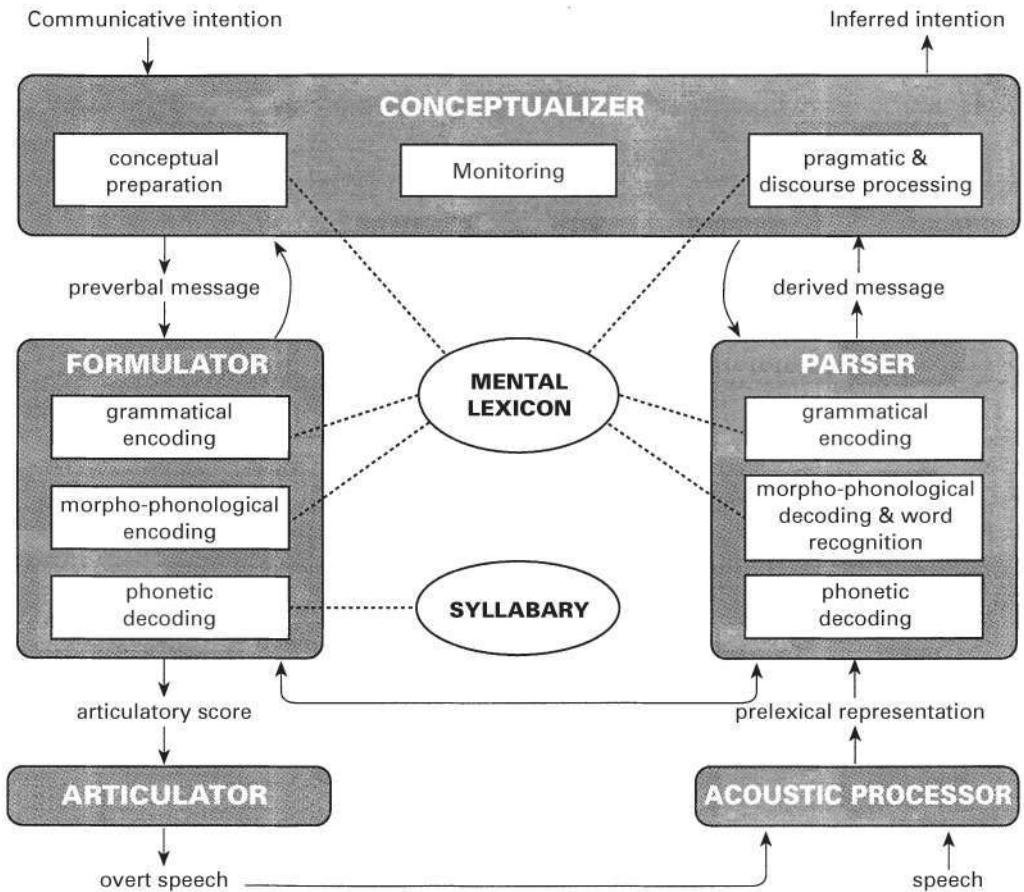


Figure 9.1 *Blueprint of the language user. Processing components involved in the production and comprehension of speech*

research: each component, its way of operating, its distinguishableness from or interaction with other components, its cerebral implementation, is or should be the topic of our research efforts.

In conversation, speaking and listening are closely bound. A participant is both speaker and listener and the verbal interaction is often a joint enterprise among participants to accomplish some goal of mutual interest. What a speaker is going to say is at any moment dependent on the state of the interaction, the attentional states of the interlocutors and their common ground (H. H. Clark, 1996).

### Speaking

The left side of the diagram shows the component processes involved in the process of speaking. Its input is called the 'communicative intention'. Almost every verbal move in conversation

is made to affect an interlocutor in some way. The intended effect is the speaker's communicative intention. The first step for a speaker is to consider how the intended effect can be brought about. Imagine the speaker wants his interlocutor to lend him her bike. How to accomplish this? The military approach is to make a command: 'Lend me your bike!'. But this approach may deter a listener as being too direct; it can be more effective to ask whether the other party is able or willing to lend her bike: 'Can/will you lend me your bike?' A very careful approach would be to ask 'Would you know how I can get there – it is really too far to walk?' Such dances of politeness in conversation are universal in human cultures, as Brown and Levinson (1987) have pointed out. A speaker will estimate which move is most appropriate in the present circumstances. The move you opt for is the 'message' you are going to express. Messages come in three main varieties or 'moods'. You can declare something,

for instance assert that the dog is limping; you can be imperative, for instance by commanding your interlocutor to lend you a bike, or your message can be an interrogative one, for instance when you ask what time it is.<sup>2</sup>

Whatever the content of a declarative, imperative, or interrogative message you opt to express, it must have an argument structure. In addition, the predicate and the arguments in the message must ultimately be expressible in words. Therefore, a message consists of 'lexical concepts', concepts such as DOG, LIMP, LEND for which there are words in the speaker's language.<sup>3</sup> The arguments in your message refer to entities such as the agent, the recipient, the location of your predication. How to cast them in terms of lexical concepts? The referring function of language is at work here. You will try to make an effective reference for your interlocutor. If you can truthfully refer to the same person as 'the plumber', 'my friend', or 'the giant', then which concept will you choose to activate for your interlocutor? If the interlocutor happens to know the tall man, but not that he is a plumber or your friend, you had better conceptualize the referent as 'giant'. This procedure of selecting an effectively referring concept is called 'perspective taking' (Levelt, 1996). Even two-year-old children show surprising flexibility in perspective taking (E.V. Clark, 1997).

A speaker's intention may be more complex than to express a single message. When you are asked to describe your apartment, you must decide what to say first, what next, etc. That is called the speaker's 'linearization problem' (Levelt, 1989). Most speakers will make an imaginary tour, beginning at the front door and going from room to room in some connected fashion. Linearization is always a problem when complex information is to be expressed. Try, for instance, to explain the game of chess to somebody. Linearization is also at stake when we tell a story, when we talk about personal events we were involved in (and this is most speakers' dearest occupation). Linearization, developing the plot, is one important aspect of our narrative skill, but there are many more, such as foregrounding and backgrounding of information, introducing referents in the story, etc.

Perspective taking and linearization are both ways to guide the interlocutor's attention. Still another way is to focus or defocus arguments for the listener. When you say 'I have a dog', you are introducing a new entity for your interlocutor, namely your dog. You focus it by using the full noun 'dog' and by giving it pitch accent. Your next sentence can be 'it limps'. You now defocus the same argument by using a pronoun

('it') to refer to it. That signals to the listener that it is the same argument you just introduced. But now you focus the limping for your listener, the newly introduced predicate.

Any message must ultimately get formulated. This involves a trio of operations: grammatical, phonological, and phonetic encoding. Grammatical encoding begins by retrieving words from the 'mental lexicon'. That is the repository of words we have built up in the course of our lives. Normal, educated persons will have some 50-100 thousand words available (Miller, 1991). Retrieving words proceeds in two steps. We first access what can be called 'the syntactic word' or the 'lemma', defined as the information about the word's syntactic properties, such as that it is a noun, or that it is a transitive verb. Quickly thereafter (van Turenhout, Hagoort, & Brown, 1998) we access the word's 'phonological code', which is used in phonological encoding (see below).

Assume the speaker begins to formulate the message PEOPLE UNDERSTAND. It consists of a predicate (UNDERSTAND) plus an argument in the role of experiencer (PEOPLE). These are two lexical concepts, and they activate the corresponding lemmas in the speaker's mental lexicon. There is always some competition with other meaning-related lemmas (such as *comprehend* when *understand* is the target).<sup>4</sup> The amount of competition determines the speed at which the lemma can be retrieved from the mental lexicon (Roelofs, 1992). Here the retrieved lemmas are a verb (*understand*) and a noun (*people*). The next step in grammatical encoding is to couple the retrieved lemmas syntactically. The lemma *understand* has a syntactic slot for a subject and that subject should express the experiencer argument. The noun *people* is of the correct syntactic category to fill that slot and it also fits the experiencer role. The syntactic coupling (or 'unification') succeeds and the result is a small syntactic tree *people understand*, with the lemmas in this particular order. Such output of grammatical encoding is called a 'surface structure'. We can, of course, encode far more complex syntactic structures and we normally do this incrementally: we begin with the initial parts of the sentence and then 'grow' it to the end. That makes it possible for a speaker to begin uttering a sentence even before it is fully constructed syntactically. The most detailed theory of how speakers accomplish this feat is Kempen's (submitted).

As grammatical encoding is proceeding, phonological encoding does it as closely as possible. Shortly after the lemmas *people* and *understand* are selected, their phonological codes

become available as well. A word's phonological code is, by and large, its string of phonemes. For the word 'people' this is the string /p/, /i/, /p/, /e/, /l/.<sup>5</sup> For 'understand' it is the string /u/, /n/, /d/, /e/, /t/, /s/, /t/, /æ/, /n/, /d/. All phonemes of a word become simultaneously available (see Levelt, Roelofs, & Meyer, 1999 for details of the experimental evidence on these and other aspects of phonological encoding). These phonological codes plus the syntactic tree are used by the speaker to construct the syllables and the prosody of the utterance.

How are syllables constructed? The evidence is that they are incrementally built up, starting at the beginning of the word. For 'understand', for instance, the speaker first concatenates /u/ and /n/, which completes a legal English syllable /un/. Then /d/, /e/, and /t/ are concatenated into a next syllable /der/, and then follows the composition of the last syllable /stænd/. Notice what happens if the speaker is rather producing the utterance 'people understand it'. The syllabification would now become /un/-/der/-/stæn/-/dit/; there is no syllable /stænd/ here. A word's syllabification is variable and context dependent. That is because syllabification straddles lexical boundaries; the syllable /dit/ belongs to both the words 'understand' and 'it'. This variability of syllabification makes it unlikely that syllables are stored in the speaker's lexicon. They are, rather, generated 'on the fly' as phonological encoding proceeds.

Generating the prosody of an utterance involves the metrical grouping of words into phrases and the assignment of intonation. The listener can use these cues to decode the syntax of the utterance and to detect what is focused or defocused by the speaker. In English, Dutch, German, and many other languages (but not in Turkish) speakers generate an intonation contour with pitch accent on the head word of the last phrase in the sentence. For the simple sentence 'people understand' this word is 'understand', which gets pitch accent on its stressed syllable /stænd/. This is followed by a so-called 'boundary tone'. If the mood of the speaker's message is declarative, pitch will normally drop on the last syllable. In the example this happens to be the same syllable as the pitch accented one (/stænd/). But it will be the syllable /dit/ in the utterance 'people understand it'. The boundary tone is very informative about the mood of an utterance. If the mood is interrogative, as in 'people understand it?', the boundary tone tends to go up.

The final step in the speaker's formulating process is phonetic encoding. The whole purpose of phonological encoding is to build a pronounceable structure. But how do the articulators know how to execute the pronunciation?

Let us return to the syllables. Syllables such as /un/ and /der/ and /stæn/ and /dit/ have been produced so often by a normal speaker that they are overlearned articulatory motor patterns. It is most likely (though not yet proven) that they are stored in the forebrain as whole syllabic gestures, and retrieved as soon as they turn up in grammatical encoding. These gestural patterns are further adapted to the current metrical and intonational plan. In addition, we adapt the loudness and intonational excursion of our speech to the environmental conditions, such as prevailing noise and distance from the interlocutor. The outcome of phonetic encoding and hence of the speaker's formulation process (see Figure 9.1) is an 'articulatory score', which can be executed by the articulatory system.

Speech articulation is the most complex motor behavior we can produce. Some hundred different muscles are involved in the generation of some 10–15 consonants and vowels per second. This high speed is only possible because these speech sounds are coarticulated (A. Liberman, 1996). When you say 'cool' you are already rounding your lips for the vowel /u/ when you pronounce the initial consonant /k/; you will not lipround the /k/ in /kill/. As a consequence, consonants and vowels do not appear as distinct entities in the speech signal (like letters in printed words). The listener has the formidable task of reconstructing them from the continuously flowing signal. I will not review the articulation process here, but see Levelt (1989) and Kent, Adams, and Turner (1996).

## Monitoring and Self-Repair

The speech we hear most is the speech we produce ourselves. We can not only attend to our own overt speech, but also to our 'inner speech' which keeps babbling during our waking hours and which we misattribute as the speech of others when we dream or when we are in an acute schizophrenic state. It is not exactly known what inner speech is. Jackendoff (1987) and Wheeldon and Levelt (1995) provide theoretical and empirical arguments respectively for the supposition that it is a phonological code, roughly like the output of grammatical encoding. Whatever it is, we can attend to it and parse it just as we parse what is said to us by others. That gives us the ability to monitor our own speech production and correct impending trouble even before articulation has begun. This was probably done by the speaker who said

and further to the ye- uh to the green dot.

Here the word 'yellow' got interrupted within its first syllable. What does the speaker do when

communicatively disruptive trouble shows up in internal or overt speech? Levelt (1983) provided evidence that speech gets halted immediately upon detection of such trouble. But detection can be relatively slow, for instance when the speaker is just initiating a new clause and has little attention available for self-monitoring. In that case, halting may follow several words after the trouble spot, as in:

And from *green* left to pink – er from blue left to pink.

When we speak we monitor for two classes of trouble. The first one is appropriateness, as in:

I am trying to *lease*, or rather sublease my apartment.

Here the speaker noticed that what she said is potentially ambiguous or underspecified for her interlocutor and she repaired it by becoming a bit more specific. The second class of trouble is all-out error. That happened in the yellow/green and green/blue errors above. We may also detect and correct syntactic error, as in

What things are this kid – is this kid going to say correctly?

or any other type of formal error.

When we interrupt our speech, we often signal to the listener that there is trouble and the kind of trouble. If it is appropriateness that is at issue, we drop in an editing term such as 'or rather' or 'I mean'. If it is all-out error, we may say 'no', 'sorry', or just 'uhm'. The factual repair is produced quite systematically. Almost always the speaker takes up and completes the interrupted syntax (Levelt, 1983).

### Speech Comprehension

When we listen to someone else, our main business is to find out what the speaker intends to convey (Ho"rmann, 1976). Just as speech production, this involves a multilevel processing system. A first step is an initial acoustic analysis of the incoming signal. One thing most listeners are good at is 'streaming', following a speaker's voice amidst all sorts of interfering noise from the environment. We can even attend to a single voice when several people are talking at the same time – the so-called cocktail party effect. We probably especially attend to those aspects of the speech signal that give contrastive information on vowels (telling 'put' from 'pot' from 'pat' from 'pet') and on consonants (telling 'sea' from 'fee' from 'bee' from 'me'). Maybe we already detect whole syllables, at least stressed syllables. It is not well known what exactly is extracted from the noisy speech signal during

this initial stage, but it does have a name: the 'prelexical representation'.

The real parsing of speech requires phonetic decoding in the first place. As mentioned, the units of speech such as phonemes, syllables, words, prosodic phrases, are not freestanding units in the speech stream. One of the paramount problems in automatic speech recognition is the segmentation of the speech signal, in particular segmenting it into words. Human listeners do this mostly easily and automatically. How do they do it? Let us first consider the English listener. In numerous experiments, Cutler and her colleagues have shown that a main word segmentation strategy is to postulate a word boundary just before a heavy syllable (see Cutler & Clifton, 1999 for a review). A heavy syllable in English is one with a full, non-reduced vowel, a syllable that usually receives primary or secondary word stress. Using this strategy, a listener will correctly spot the beginning of words like 'boy', 'beacon', 'article', etc. The listener will mis-parse for words in the speech stream such as 'alert' or 'connect'. How often will the listener, using this method, segment correctly when listening to normal fluent talk? In about 90% of the cases, Cutler and Carter (1987) computed. The segmentation strategy works well for stress-timed languages, such as English, Dutch, or German, but it is not a universal one. In a stress-timed language there is a rhythmic alternation of strong and weak syllables and listeners capitalize on that rhythmic property of their native language for the purpose of segmentation. The main discovery of the cross-linguistic research project of Cutler and her colleagues is that listeners can always exploit the rhythmic properties of their language to derive word boundary information. But these rhythmic properties vary substantially across languages. There is good experimental evidence now that Japanese listeners are particularly sensitive to the characteristic moraic rhythm of their language (a word like *Honda* consists of three timing units or morae, *Ho-n-da*), whereas French, Spanish, and Catalan listeners have been shown to utilize the dominant syllable rhythm of their languages.

Can listeners exploit the speech signal to extract information on phrasal constituency? Some typical experiments in the psychology of language show that they can, to some extent at least. For instance, Levelt, Zwanenberg, and Ouweneel (1970) had four female native speakers of French read texts that contained an ambiguous sentence. Here is an example with its two versions (with their unambiguous English translations):

A. // *veut vendre cet objet* / *vole a son ami*

(He wants to sell that object / stolen from his friend)

B. // *veut vendre cet objet vole* / *a son ami*

(He wants to sell that stolen object / to his friend)

The text preceding the target sentence made it completely unambiguous and rarely did a speaker notice any ambiguity. The two versions (A and B) were spliced out of their texts: the 'context' versions. After having read the texts, the speakers were informed about the potential ambiguity of the target sentences. Now they were asked to read them as unambiguously as they could in either version (A and B). This provided the 'isolated' versions (A and B) of the sentences. The context and the isolated versions of the sentences were presented to groups of native French listeners. They were explicitly informed about the ambiguity of the sentences they were going to hear, and invited to judge which interpretation had been intended by the speaker. They were correct in 75% of the isolated cases (significantly different from the 50% chance level) and there was no significant bias towards the one or the other interpretation. This means that speakers can explicitly provide disambiguating prosodic information if asked to do so. For the versions spoken in context the correct score dropped to 60%. This still differed significantly from chance, but also from the 75% above. Clearly, speakers do not bother much to provide disambiguating prosodic information if the semantic context does the work already. An acoustic analysis of the context versions of the example sentence above showed that the cues they provide are of two kinds. In the A version, but not in the B version, there is a slight pause between *objet* and *vole*, indicating a phrasal break. In addition, that break is further marked by a marked pitch movement: in the A version *objet* ends at a high tone, *vole* begins at a low tone, whereas there is intonational continuity in the B version.

Cutler, Dahan and van Donselaar (1997) concluded their comprehensive review of research in listeners' exploitation of prosodic cues to syntax by saying that listeners can pick up cues that mark a break. A prosodic break is quite likely to mark a syntactic break, as is the case in the above examples. However, speakers often do not mark syntactic boundaries prosodically, as was the case in most of the text-embedded sentences above. Hence, listeners must have additional ways of parsing a sentence syntactically.

The initial segmentation of word-like and phrasal units is the listener's stepping stone to further morpho-phonological decoding of the speech signal. Central here is the process of

word recognition. The core result of word recognition research since Morton's (1969) seminal publication is that an incoming speech signal causes multiple word activation in the listener's mind. Words compatible with a given stretch of speech are simultaneously activated by it. Such a set of co-activated word candidates is technically called a 'cohort' (Marslen-Wilson & Welsh, 1978). Models of word recognition give varying accounts of how a cohort gets resolved. How does the listener reduce it to a single, most likely solution, the recognized word? Here is an example of cohort reduction as originally proposed by Marslen-Wilson and Welsh: The listener receives as input the spoken word signal *trespass*. The first stretch of speech, *tr* activates words in the listener's mental lexicon that begin with 'tr-', such as 'trap', 'tremble', 'treasure', 'treat' and, of course, 'trespass'. All of them are compatible with this word-initial speech segment. This is called the 'word-initial cohort'. When as much as *tre* has come in, incompatible words, such as 'trap' and 'treat' are deactivated, whereas compatible words such as 'tremble' and 'treasure' are further activated. When the input signal has come as far as *tres*, only a few candidate words remain, among them 'treasure' and 'trespass'. Uniqueness is reached when the listener has taken in as much as *tresp*; 'trespass' is the only word in English that has 'tresp' as word-initial part. Reaching this so-called 'uniqueness point', the listeners can recognize the word and experiments show that they often do. Notice that the word can be recognized before it has fully sounded. That is often the case (dependent on where the uniqueness point is located) and that can help the listener to 'predict' the upcoming word boundary.

But the situation is more complicated than this. Most words have other words embedded in them. In *trombone* there is the word 'bone', and indeed 'bone' gets activated when you listen to 'trombone'. When you hear *start* you will coactivate 'star', 'tar', 'are', 'tart' (see Frauenfelder & Floccia, 1998 for a comprehensive review of research in word recognition). To make things even worse, there are often embedded words across word boundaries. Given the uncertainties of initial word segmentation, these can also play a role. For instance, the speech signal *first acre* will not only activate 'first' and 'acre', but also 'stay', 'steak', and 'take' (Cutler & Clifton, 1999). Modern theories of word recognition (see Frauenfelder and Floccia's review) explain how the ensuing within-cohort competition is efficiently resolved with optimal speed. Indeed, word recognition is very fast and often results before the word's end has sounded.

Many words that we encounter while listening to running speech are morphologically complex

(though languages differ substantially in this respect, see Section 9.1 above). They can be inflected forms, such as the past tense: 'walk-ed' and pluralization: 'tree-s'. They can also be derivationally complex words, such as nominalizations: 'walk-er' and verbalizations: 'vaccinate'. There is increasing evidence that listeners attend to words as wholes, but also to their constituent morphemes. Listeners follow a 'double' or 'parallel route' in parsing morphologically complex words (Baayen, Dijkstra, & Schreuder, 1997). For instance, when a Dutch listener hears a plural form (such as 'tree-s' in English), the speed of recognition is determined by the frequency of occurrence in language use of that plural word.<sup>6</sup> But if the same listener hears a singular form (such as 'tree' in English), the speed of recognition is determined by the sum of singular and plural word frequencies. In other words, it is determined by the frequency of the stem (the frequency of 'tree', whether occurring with or without plural inflection). For an excellent review of morphological decoding, see Schriefers (1998). It may be useful for a listener to parse morphology, in particular inflection. In many languages grammatical decoding would be all but impossible without the listener's attending to inflectional markers. In German, for instance, listeners would be at a loss to distinguish between 'Der Peter rief den Hans an' (Peter called up Hans) and 'Den Peter rief der Hans an' (It was Peter that Hans called up) if they did not notice the nominative versus accusative case marking on the determiners, 'der' versus 'den' (the prosody of the two sentences can be the same).

As soon as words and their inflections are recognized, their syntax and semantics become available for the grammatical decoding of the utterance. This is, by and large, decoding the argument structure. Like word recognition, grammatical decoding is an incremental operation. It begins as soon as the first word is recognized and develops with every further recognized word or morpheme. Also, grammatical decoding is 'omnivorous': it ingests everything, uses any kind of information it encounters. The relevant information can be syntactic, but in many cases the listener uses semantic or pragmatic cues just as fast as syntactic ones. This immediate and massive use of whatever useful cues we can pick up is necessary, because almost every utterance we listen to is multiply ambiguous. The utterance 'my pupil' is ambiguous, because 'pupil' can mean 'part of the eye' or 'student'. A listener will not notice the ambiguity when hearing 'my eye's pupil' in which the 'part of the eye' interpretation is strongly invited. Only very precise reaction time measurements show that the alternative 'student'

meaning is temporarily activated, but then quickly suppressed. As listeners, we are rarely aware of the fact that almost any word we encounter has multiple meanings. In order to derive the argument structure of an utterance, information of the kind 'Who did what to whom', we are much dependent on recognizing the verb or verbs in an utterance. Encountering the word 'see', we know immediately that it must induce the argument structure 'someone sees something' and we will try to assign other words or phrases in the sentence to these 'someone' and 'something' roles. But this is not without problems either, because we will often meet local syntactic ambiguities as we proceed. The 'something' argument of 'see', for instance, can be an entity such as a person or an object ('I see the signpost'), but it can also be an event ('I see the weather change'). This can easily lead to parsing trouble, as in reading the following newspaper headline (from Altmann, 1997):

*Crowds rushing to see Pope trample six to death*

Certainly, the newspaper did not intend to convey that the Pope trampled six to death, but it is a possible reading of this sentence. That this ambiguity goes unnoticed is due to a combination of parsing strategies we have as readers or listeners. One is that we prefer the simplest possible syntactic structure compatible with the data. When reaching 'see', we prefer to assign the 'something' role to a simple noun phrase (like 'Pope') rather than to a more complex complement clause (like 'Pope trample six to death'). Also, we tend to take early decisions. We can close the argument assignment of 'see' as soon as we get to 'Pope' and we simply take the risk that we must revise it later on. These and other syntactic parsing strategies have been the topic of intensive research, see Frazier and Clifton (1996).

Clearly, other considerations will help us as well to cope with local ambiguities. We use all sorts of circumstantial evidence to zoom in on one solution rather than another. Our knowledge of the world tells us that wherever a Pope goes, crowds assemble to see him. There is increasing evidence that different knowledge sources, such as syntax, semantics and knowledge of the world, affect our parsing in parallel and almost immediately (Kempen, 1998; Tanenhaus & Trueswell, 1995).

Grammatical decoding, in particular the assignment of argument structure, strongly interacts with what is called 'pragmatic and discourse processing' in the diagram of Figure 9.1. In order to derive what the speaker intended to convey, the listener must find out what referents the speaker is referring to by means of the various arguments. 'The Pope' – which Pope?,

'six' – six what?, etc. Here the listener draws heavily on general knowledge, on knowledge of the conversational situation and on what was said before in the current discourse. An example of the latter is this. When somebody tells me 'My car has broken down. The battery is flat', the speaker is making a definite reference by means of the phrase 'the battery'. Which battery is intended? There was no previous talk about a battery. Still, I will immediately infer that it must be the battery of the broken-down car, mentioned in the previous sentence. I can make this inference because I know that cars have batteries. I also tacitly assume that my interlocutor is 'cooperative' – there would be little gain in conversation if we tried to trick our listeners into drawing the wrong inference<sup>7</sup> (H. H. Clark, 1992). In quick interactions, referring can be quite indirect, but still solvable. Nunberg (1979) presents the example of one waiter in a restaurant telling the other: 'The hamburger wants the bill'. Here 'the hamburger' apparently refers to the customer who had a hamburger, not the hamburger itself. The ambiguity is easily resolved given shared knowledge of the situation. Still, the ease with which we resolve such ambiguous reference is a miracle. Inferring the referents of arguments in discourse is a major stumbling block for artificial language comprehension technology. That also holds for inferring the referents of pronouns. In their review of discourse comprehension, Noordman and Vonk (1998) discuss various cues that listeners use in guessing the referents of pronouns. When you hear 'Harry won the money from Albert, because he . . .', your first hunch will be that 'he' refers back to Harry, the topic of the sentence. But when you hear 'Harry trusted Albert, because he . . .', you will guess that 'he' refers back to Albert. How come? The inference is initiated by 'because'; the speaker is going to explain why Harry trusted Albert. The typical explanation here is some quality of the person trusted. This is only one of many subtle determinants of pronoun interpretation.

It will often be the case that a listener, in spite of tremendous skills of interpretation, does not succeed in deriving the speaker's intention. That is not dramatic – we are, in fact, accustomed to it and we possess a whole arsenal of interactive strategies to solve ambiguities of reference in conversation. Here is one example from H. H. Clark (1996, p. 172), the major study of these skills. Brenda and Alva are discussing paintings along the wall:

*Brenda: that green is not bad, is it, that landscape?*

*Alva: What the bright one?*

*Brenda: Yes.*

*Alva: \*It's\*<sup>8</sup>*

*Brenda: \*Well it's\* not very bright, no I meant the \*second one along\*.*

*Alva: \*Oh, that one over\* there.*

Here the correct reference of the original 'that green' is progressively cleared up in extensive interaction between the interlocutors. In producing and understanding speech we negotiate meaning.

#### 9.4 ACQUIRING LANGUAGE

It has always surprised psycholinguists that the complex skills of language use develop so early in the life of a child. By the age of 12 to 14 months infants master basic phonetic operations, such as distinguishing the consonants and the vowels of their native language and producing the language's most frequent syllables. At the age of 18–20 months most toddlers have a vocabulary of about 50 words, but then comes the 'word spurt': soon new words are added at a rate of some 8–10 per day, one per waking hour. With the expansion of the vocabulary, the child begins to make the relevant phonological distinctions and with each new word it learns how the word can figure in a particular type of argument structure. Grammatical encoding and decoding begin to shape up by the end of the second year of life. Surely, the five-year-old is not yet an accomplished native speaker, but the basic system is up and running. Evolution clearly designed us to be skilled linguistic communicators early in life, but it is unknown how this is programmed in our genome.<sup>9</sup> Moreover, that consideration does not provide us with enough restrictions to explain this early pattern of development. Only detailed observational and experimental analysis can help us unravel the forces that drive early language development.

The various components depicted in Figure 9.1 have their own developmental pattern. The maturation of a component is in some cases quite autonomous, in other cases quite interactive. Surprisingly autonomous is the early maturation of phonetic encoding and decoding. A newborn baby is already able to distinguish native from non-native prosody. The womb is a low-pass filter and the fetus is able to pick up the low-frequency rhythm of the mother's tongue. Two-month-old infants can distinguish [ba] from [pa], [bæ] from [dæ] from [gæ], [wa] from [ja], [ma] from [na], [da] from [di] from [du], etc. In other words, they can pick up phonetic distinctions, such as voicing, place of articulation, nasality, that may turn out to be relevant for the acquisition of their native language. But when they approach the end of the first year of

life, infants become increasingly insensitive to phonetic distinctions that are not relevant to their native language. Whereas six-to-eight months old Japanese babies readily distinguish between [ra] and [la], they lose this sensitivity by the age of ten to twelve months. Infants become 'phonologically deaf' to most non-native sound distinctions (see Jusczyk, 1997, for a review of speech perception during the first year of life).

A similar development is apparent in production. A normal infant begins to babble at the age of about seven months. Babbles are repetitive and alternating syllabic patterns, such as [ba-ba], [gi-gi] or [di-ti]. These first utterances are entirely meaningless. Rather, infants begin building up a syllabary by attending to the auditory effects of their own spontaneous articulations. Only when the basic system is functioning (around the age of 12 months), do they begin to prefer producing native syllables over non-native ones. And only then, one or a few of these babbles begin to denote a person, animal, or action. In other words, meaning is not the driving force behind the child's initial phonetic development; these skills develop autonomously during the first year of life in a sufficiently rich speech environment.

Building up a mental lexicon is a highly interactive enterprise. By the end of the first year, the infant already comprehends a few words. This probably drives the first attempts at production; syllabic patterns that approximate the auditory effects of known words are selected to make the appropriate reference. The simple coupling of syllabic patterns to referents soon gives way to a tripartite development. The first aspect is phonological development. As more and more protowords are added to the lexicon, more and more syllabic patterns must be kept apart to make the relevant meaning distinctions. Somewhere between 1;6 and 2;6 children solve this problem by 'phonologizing' their protolexicon. Initially, their protowords are whole articulatory gestures, but slowly they start attending to word beginnings, to word ends and to vowels as independently variable segments (C. C. Levelt, 1994). This provides them with a powerful 'bookkeeping' system for distinguishing words in any of these positions: [pin] from [tin], [tin] from [tan], [pin] from [pit], etc. Also, during the same period, they develop basic skills in coupling two or more words prosodically. The first multiword utterances superficially sound like single words uttered in succession. And indeed, the constituent words often have long pauses between them. But more precise measurements show that non-final words in such utterances are shorter than final ones and that intonation drops from non-final to final words;

early multiword utterances are generated as wholes (Branigan, 1979).

A second aspect is the initial development of word meaning. During the first year of life, children have acquired basic knowledge about persons, animals, objects, actions, events and the first words are attempts to denote some of them. They tend to pick out whole objects (dog, chair) and whole actions (go, put), not parts of them. Although children assume that different words have different meanings, they quite early know that you can use different words to refer to the same entity, as one child (age 1;7) did when indicating his bowl of cereal first by 'food', then by 'cereal' (E. V. Clark, 1997). Perspective taking is an early skill.

The third and most dramatic aspect is the acquisition of word argument structure. A child's first two-word utterances can perform various functions, such as to express location ('there book'), possession ('baby shoe'), event structure ('hit ball'), etc. There is beginning argument structure for the expression of declarative ('big boat'), interrogative ('where ball?') and imperative ('more milk') moods. The need to express more complex argument structure is the first trigger for the emergence of syntax. A child acquiring an inflecting language, such as English, French, or German, begins to mark argument structure by inflection towards the end of the second year. Words for actions begin to become inflected for tense (progressive, past, present) as in 'Christy forgot milk' (Bowerman, 1990; child's age 1;11). This marks them syntactically as verbs. For each newly acquired verb, the child learns how to map its semantic arguments onto syntactic roles. In the example, the verb 'forget' puts the agent argument ('Christy') in syntactic subject position and the theme argument ('milk') in first object position.

The way in which verbs map their semantic arguments onto syntactic roles varies considerably. Take the verb 'give'. One month later, the same child produced 'I give mommy a bottle'. Here, the agent ('I') is again in subject position, but the theme argument ('a bottle') does not end up in first, but in second object position. It is the recipient argument ('mommy') that becomes the first object in the sentence. In spite of these differences between verbs, children initially make surprisingly few errors in their syntactic rendering of semantic arguments when they acquire new verbs. But errors do appear much later on, when the child has already mastered a substantial number of verbs. By then the child has sufficient experience with verbs to discover more general patterns in the syntactic realization of argument structure. A child will then occasionally overgeneralize such patterns.

Take so-called 'mental verbs', such as in 'something pleases / excites / bores / surprises / scares somebody'. All these verbs put the stimulus argument ('something') into subject position and the experiencer argument ('somebody') in object position. But there are other mental verbs, such as 'like' or 'hate' that do it the other way round: 'Somebody likes/hates something'. When a child at a later age acquires a less frequent verb of the latter type, such as 'enjoy', she may erroneously put it in the wrong class and make the error 'I saw a picture that enjoyed me' (child aged 6;6 – example from Bowerman's (1990) analysis of these developmental patterns). By the age of twelve, children still occasionally err on Latinate verbs, such as 'donate', patterning them after non-Latinate near-synonyms (such as 'give') to produce errors like 'he donated the church some money'.

As these examples show, important aspects of acquiring the syntax of the native language are lexically driven (E. V. Clark, 1995). The child first learns each verb's typical syntactic frame and only later generalizes these frames to particular general kinds, for instance how to make a complement construction ('I'll help you to find the butter' – see Bloom, Tackeff, & Lahey, 1984). But there is much more to be acquired in syntax, such as the construction of questions, the use of pronouns, the appropriate use of negation, etc. Linguists try to discover patterns of syntactic acquisition that hold cross-linguistically, in an effort to uncover universals of our linguistic endowment (see Weissenborn, Goodluck, & Roeper, 1992).

To become a native language user, the child must acquire more than phonology, meaning, and syntax alone. In order to act through language, the child must become versed in a variety of conversational skills: the ways of turn taking and turn giving, the ways in which to phrase intentions politely and indirectly, the appropriate addressing forms, etc. Each of these begins to develop quite early in life, but full competence is not reached before puberty. This is especially apparent for the skill of narration. The child must learn to guide the listeners attention by making use of various linguistic devices. For instance, in an English-language narration one can use tense marking to foreground an event against a background for the listener. The child (3;9) who describes a scene with a boy and a frog in it as 'The frog got out, when he's sleeping' focuses the listener's attention on the frog's action by using past tense and marks the boy's sleeping as background by using progressive tense. The child narrator will also try to keep the listener's attention focused on some agonist, once introduced, by using pronominal

reference (as the three-year-old above did by using 'he' when referring to the boy that had been introduced earlier). And the narrator will package the information in chunks or 'paragraphs' that the listener can oversee. Young children do not know how to do that. They often introduce each bit of information independently with 'And then . . .'. As these skills develop, narration becomes more and more cohesive. See Berman and Slobin (1994) for an extensive study of how narrative skills are acquired by children in different language communities. It shows that our narrative skills are not in full swing before we reach adolescence.

In literate cultures, finally, the child will have to acquire the culture's writing system. It usually takes years of training for a child to become a skilled reader and writer.

## 9.5 READING AND WRITING

Evolution did not design us to become users of written language. The widespread use of writing systems is so recent in human history that our genome has not been affected by any selective advantage that a writing system might provide. The cultural evolution of writing systems involves a discovery and an invention. The discovery is that continuous speech is based on underlying discrete units, in particular words, syllables, phonemes. This discovery was made, long ago, in oral cultures. Panini, for instance, developed a detailed, orally transmitted theory of the phonemic structure of language. The invention was to map any of these unit types onto visual symbols. This greatest of all cultural inventions has been an exceedingly difficult process. It succeeded twice, or at most three times in the course of our cultural history, first to the Sumerians a good five millennia ago, then to the Chinese some four millennia ago (though they may have had access to already existing writing systems) and finally to the Olmec and Maya in Central America around 200 B.C.E. It is even much more recent – in fact not much more than a century ago – that mostly Western cultures began to impose literacy as a general educational requirement.

The challenge to the psychologist of language is to explain the apparent fact that most of us are indeed able to acquire a writing system and to become fluent readers. Clearly, the skill is parasitic on two pre-existing skills, language and visual pattern recognition. For a major part reading is language comprehension as discussed in Section 9.3 above, only the input is visual instead of auditory. Following (visual versus auditory) word recognition, the two processes

largely coincide. The remaining differences concern the absence of prosodic cues to phrasal parsing in reading (though partly compensated for by the presence of commas and dots), and the more dominant use of low-frequency words and low-frequency syntactic patterns in written texts. A major difference, also, is that reading is not an interactive process, where meaning is 'negotiated' between interlocutors; the reader is alone in deciphering the author's intention.

As hunter/gatherers we evolved refined pattern recognition. We became fast saccadic scanners, quickly detecting and recognizing small visual patterns that are of potential relevance to us, such as shapes of leaves, fruits, silhouettes, footprints. That ability is marshaled when we scan a text. A skilled reader scans some five to six words per second, fixating about 80% of the content words in a text. This is about twice the rate of normal spoken language understanding. The fixation of about 200 ms usually suffices to recognize the word.

The process of visual word recognition has become a major topic in reading research. When the script is alphabetic, the graphemic units to be recognized are letters or letter combinations that represent phonemes. These overlearned units are activated in the visual system by the characteristic contour patterns in the visual input. The ordered pattern of graphemic units activated by a fixated word has direct, parallel access to the orthographic word representation in the mental lexicon; this often suffices to recognize the word, in particular when the word is high-frequency. In addition, the graphemic units also activate 'their' phonemes and the reader can 'assemble' the word from the string of phonemes. This is, in fact, the only way for a reader to handle a new word or non-word, such as 'flork'; it is also what a child does in the first stage of learning to read. This phonemic assembling route remains active in the fluent reader; it gives access to the phonological word representation in the mental lexicon. It is often the faster route for the recognition of less frequent words, words we have less visual experience with. Quite probably the direct visual and indirect phonemic processes are mutually reinforcing in skilled reading (see Perfetti, 1999 for a review of reading comprehension).

The one most critical step in acquiring a script is to become aware of the linguistic units that are to be encoded by the visual symbols, or in other words to repeat the original cultural discovery. This is relatively easy if the unit is a word. Children develop word awareness without much effort. But none of the existing writing systems, including Chinese script, is a pure word-to-symbol matching system; there is always phonology involved. It is harder for

children to become aware of syllables as spoken language units. That is what a Korean child must acquire in order to learn Korean syllabic script. But most difficult for children is to become aware of phonemes. There is no spontaneous awareness of phonemes in illiterate cultures, it is absent in illiterates living in literate cultures and many children never acquire reliable phonemic awareness, in spite of extensive training. It is not surprising that phonemic units do not stand out in spoken language. The articulatory gestures that realize successive phonemes substantially overlap in time, different from words and syllables. As a consequence, the acoustic word pattern does not contain discrete temporal units that correspond to phonemes. One can be a normal, fluent, and even skilled language user without having more than a rudimentary ability to become aware of the phonemic structure of a word. Persistent lack of phonemic awareness is a major cause of dyslexia (I. Y. Liberman & Shankweiler, 1991). It is, however, a short-sighted misnomer to call dyslexia a language disorder;<sup>10</sup> it is not.

## 9.6 CONCLUDING REMARKS – LANGUAGE AND THE BRAIN

It has always been a challenge for the psychology of language to dissect the implementation of language skills in the brain. Traditionally it was responded to by combining the study of language disorders with post-mortem brain anatomical research. The advent of modern brain imaging technology, such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI), magnetic encephalography (MEG), and the registration of event-related electrical brain potentials (ERP) has dramatically changed the possibilities to meet the challenge, because language processes can now be detected and localized in the intact living brain. For reviews of these important developments see Brown and Hagoort (1999), Stemmer and Whitacker (1998), the chapter by Friederici in Friederici (1998), and the section on language in Gazzaniga (2000).

ERP and MEG, with their millisecond time resolution, provide on-line measures of the brain's dealings with linguistic tasks. For instance, when a subject reads or listens to a sentence, every new content word releases a negative brain potential, peaking around 400 ms after word onset. The better a word fits the semantic context, the smaller the N400 response. This N400 is probably generated in the anterior temporal lobes. Other ERP components are particularly sensitive to syntax rather

than to semantics and they are generated in other cortical regions. Such findings support the view that syntactic and semantic operations, as much as they interact in sentence understanding, are subserved by separate, specialized systems in the brain (Friederici, 1998).

The precise localization of specialized regions, however, requires the measurement of metabolic activity in the brain by means of PET or fMRI. Such measurements have already undermined the classical notion that the vicinity of Broca's area in the left inferior frontal lobe has an exclusive role in speech production. Rather, this region is as much involved with speech comprehension, in particular with rapid phonetic and syntactic processing.

The psychology of language is now firmly functioning in the larger context of the cognitive neurosciences.

#### ACKNOWLEDGEMENTS

I am grateful to Eve Clark and to Anne Cutler for their help in the preparation of this chapter.

#### NOTES

1. 'Glu'ckliche Fortschritte in der Sprachwissenschaft setzen eine entwickelte Psychologie voraus.'

2. Speakers express other kinds of messages as well. We often use expressives such as 'heavens' (or worse) and what H. H. Clark (1996) has called 'collaterals', such as 'uhm'.

3. I will denote lexical concepts by capitals.

4. I will denote lemmas by italics.

5. Phonemes are represented by symbols from the International Phonetic Alphabet (IPA) between slashes '/'. I will only slash the beginning and end of the string or delete slashes entirely when there can be no ambiguity.

6. Word frequency counts are intensively used in psycholinguistic research, because word frequency is one of the strongest determinants of word recognition latency and of word production latency. The best frequency count for English, German, and Dutch is the one published on CDROM by CELEX and distributed by the Linguistic Data Consortium.

7. We do do this occasionally, but that can only be successful just because an interlocutor doesn't expect it to happen.

8. Stretches in between '\*' marks indicate speech overlap between the speakers.

9. The only information available is the involvement of chromosome 7 in three different developmental disorders that have a strong linguistic dimension: autism, developmental dysphasia, and William's syndrome.

10. If a society suddenly requires all of its children to acquire bike riding (as is factually the case in Holland), some otherwise perfectly normal children will turn out to find this very hard. By the same reasoning they apparently have a disorder of locomotion, a pathology in urgent need of medical intervention.

#### RESOURCE REFERENCES

- Brown, C., & Hagoort, P. (Eds.). (1999). *Neurocognition of language*. Oxford: Oxford University Press.
- Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.
- Clark, H. H. (1997). *Using language*. Cambridge: Cambridge University Press.
- Friederici, A. (Ed.). (1998). *Language comprehension. A biological perspective*. Berlin: Springer.
- Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Pinker, S. (1994). *The language instinct*. London: Allan Lane.
- Slobin, D. I. (Ed.). (1985). *The crosslinguistic study of language acquisition*. 2 Vols. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Stemmer, B., & Whitacker, H. A. (Eds.). (1998). *Handbook of neurolinguistics*. San Diego, CA: Academic Press.

#### ADDITIONAL LITERATURE CITED

- Altmann, G. T. M. (1997). *The ascent of Babel*. Oxford: Oxford University Press.
- Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language*, 37, 94–117.
- Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bloom, L., Tackeff, J., & Lahey, M. (1984). Learning to in complement constructions. *Journal of Child Language*, 11, 391–406.
- Bowerman, M. (1990). Mapping thematic roles onto semantic functions: are children helped by innate linking rules? *Linguistics*, 28, 1253–1289.
- Branigan, G. (1979). Some reasons why successive single word utterances are not. *Journal of Child Language*, 6, 411–421.

- Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Clark, E. V. (1995). Language acquisition: The lexicon and syntax. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 303–337). San Diego, CA: Academic Press.
- Clark, E. V. (1997). Conceptual perspective and lexical choice in language acquisition. *Cognition*, 64, 1–37.
- Clark, H. H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133–142.
- Cutler, A., & Clifton, C. (1999). Comprehending spoken language: Blueprint of the listener. In C. Brown & P. Hagoort (Eds.), *Neurocognition of language*. Oxford: Oxford University Press.
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40, 141–201.
- Dunbar, R. (1996). *Grooming, gossip, and the evolution of language*. London: Faber and Faber.
- Frauenfelder, U., & Floccia, C. (1998). The recognition of spoken word. In A. Friederici (Ed.), *Language comprehension. A biological perspective* (pp. 1–40). Berlin: Springer.
- Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.
- Gazzaniga, M. (Ed.). (2000). *The cognitive neurosciences* (2nd ed.). Cambridge, MA: MIT Press.
- Ho'rmann, H. (1976). *Meinen und Verstehen*. Frankfurt: Suhrkamp.
- Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.
- Kempen, G. A. M. (1998). *Sentence parsing*. In A. Friederici (Ed.), *Language comprehension. A biological perspective* (pp. 213–228). Berlin: Springer.
- Kempen, G. (submitted). Human grammatical coding. Manuscript. Leiden University.
- Kent, R. D., Adams, S. G., & Turner, G. S. (1996). Models of speech production. In N. J. Las (Ed.), *Principles of experimental phonetics* (pp. 3–45). St. Louis: Mosby.
- Levelt, C. C. (1994). *The acquisition of place*. Leiden: Holland Institute of Generative Linguistics Publications.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.
- Levelt, W. J. M. (1996). Perspective taking and ellipsis in spatial descriptions. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 77–108). Cambridge, MA: MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–38.
- Levelt, W. J. M., Zwanenburg, W., & Ouweneel, G. R. E. (1970). Ambiguous surface structure and phonetic form in French. *Foundations of Language*, 6, 260–273.
- Lieberman, A. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.
- Lieberman, I. Y., & Shankweiler, D. (1991). Phonology and beginning reading: A tutorial. In L. Rieben & C. A. Perfetti (Eds.), *Learning to read: Basic research and its implications* (pp. 3–17). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10, 29–36.
- Meringer, R., & Mayer, K. (1896). *Versprechen und Verlesen*. Stuttgart: Goshensche Verlag. (Re-issued, with introductory essay by A. Cutler and D. A. Fay (1978). Amsterdam: John Benjamins.)
- Miller, G. A. (1991). *The science of words*. New York: Freeman.
- Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, 76, 165–178.
- Noordman, L., & Vonk, W. (1998). Discourse comprehension. In A. Friederici (Ed.), *Language comprehension. A biological perspective* (pp. 229–262). Berlin: Springer.
- Nunberg, G. D. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy*, 3, 143–184.
- Perfetti, C. A. (1999). Comprehending written language: A blueprint of the reader. In C. Brown & P. Hagoort (Eds.), *Neurocognition of language*. Oxford: Oxford University Press.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42, 107–142.
- Schriefers, H. (1998). Morphology and word recognition. In A. Friederici (Ed.), *Language comprehension. A biological perspective* (pp. 101–132). Berlin: Springer.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Steinthal, H. (1855). *Grammatik, Logik und Psychologie*. Berlin.
- Stern, C., & Stern, W. (1907). *Die Kindersprache*. Leipzig: Barth.
- Tanenhaus, M. R., & Trueswell, J. L. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language and communication*. San Diego, CA: Academic Press.
- Tomasello, M., & Call, J. (1997). *Primate cognition*. Oxford: Oxford University Press.
- Van Turenhout, M., Hagoort, P., & Brown, C. (1998). Brain activity during speaking: From syntax to

- phonology in 40 milliseconds. *Science*, 280, 572–574.
- Weissenborn, J., Goodluck, H., & Roeper, T. (1992). *Theoretical issues in language acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wheeldon, L., & Levelt, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory and Language*, 34, 311–334.
- Wundt, W. (1900). *Die Sprache*. 2 Vols. Leipzig: Kroñner.

*Communications concerning this chapter should be addressed to:* Professor William J. M. Levelt, Max-Planck Institute, Wundtlaan 1, NL-6525 XD Nijmegen, The Netherlands