4.3

PSYCHOLINGUISTICS

Willem J. M. Levelt
Max-Planck-Institut für Psycholinguistik, Nijmegen, The Netherlands

Psycholinguistics is the study of the mental processes and skills underlying the production and comprehension of language, and of the acquisition of these skills. This chapter will deal with the former aspect only; for the acquisition of language see the suggested "Further reading" at the end of this chapter.

Although the term "psycholinguistics" was brought into vogue during the 1950s, the psychological study of language use is as old as psychology itself. As early as 1879, for instance, Francis Galton published the first study of word associations (Galton, 1879). And the year 1900 saw the appearance of Wilhelm Wundt's monumental two-volume work *Die Sprache*. It endeavoured to explain the phylogeny of language in the human mind as an increasingly complex and conscious means of expression in a society, and to describe how language is created time and again in the individual act of speaking. Although Wundt deemed it impossible to study language use experimentally, his contemporaries introduced the experimental study of reading (Huey), of verbal memory and word association (Ebbinghaus, Marbe, Watt), and of sentence production (Bühler, Seltz). They began measuring vocabulary size (Binet), and started collecting and analysing speech errors (Meringer and Mayer). The study of neurologically induced language impairments acquired particular momentum after Paul Broca and Carl Wernicke discovered the main speech and language supporting areas in the brain's left hemisphere. In the absence of live brain tomography, aphasiologists began developing neurolinguistic tests for the purpose of localizing brain dysfunctions.

All of these themes persist in modern psycholinguistics. But developments since the 1950s have provided it with two of its most characteristic features, which concern linguistic *processing* and *representation*. With respect to processing, psycholinguistics has followed mainstream psychology in that it considers the language user as a *complex information processing* system. With respect to representation, psycholinguists stress the gigantic amount of *linguistic knowledge* the language user brings to bear in producing and understanding language. Although the structure of this knowledge is the subject matter of linguistics, it is no less a psychological entity than is language processing itself (Chomsky, 1968). Psycholinguistics studies how linguistic knowledge is exploited in language use, how representations for the form and meaning of words, sentences, and texts are constructed or manipulated by the language user, and how the child acquires such linguistic representations.

I shall first introduce the canonical setting for language use: conversation. Next I shall consider the mental lexicon, the heart of our linguistic knowledge. I shall then move to the processes of speaking and speech understanding respectively. Finally I shall turn to other modes of language use, in particular written language and sign language.

CONVERSATION

Our linguistic skills are primarily tuned to the proper conduct of conversation. The innate ability to converse has provided our species with a capacity to share moods, attitudes, and information of almost any kind, to assemble knowledge and skills, to plan coordinated action, to educate its offspring, in short, to create and transmit culture. And all this at a scale that is absolutely unmatched in the animal kingdom. In addition, we converse with ourselves, a kind of autostimulation that makes us more aware of our inclinations, of what we think or intend (Dennett, 1991). Fry (1977) correctly characterized our species as *homo loquens*.

In conversation the interlocutors are involved in negotiating meaning. When we talk, we usually have some kind of communicative intention, and the conversation is felicitous when that intention is recognized by our partner(s) in conversation (Grice, 1968; Sperber & Wilson, 1986). This may take several turns of mutual clarification. Here is an example from Clark and

Wilkes-Gibbs (1986), where subjects had to refer to complex tangram figures:

A: Uh, person putting a shoe on.
B: Putting a shoe on?
A: Uh huh. Facing left. Looks like he's sitting down.
B: Okay.

Here the communicative intention was to establish reference, and that is often a constituting component of a larger communicative goal. Such goals can be to commit the interlocutor or oneself to some course of action, as in requesting and promising, or to inform the interlocutor on some state of affairs, as in asserting, for example. The appropriate linguistic acts for achieving such goals are called *speech acts* (Austin, 1962).

Although what is said is the means of making the communicative intention recognizable, the relation between the two can be highly indirect. Conversations involve intricate mechanisms of politeness control (Brown & Levinson, 1987). What is *conveyed* is often quite different from what is *said*. In most circumstances, for instance, we don't request by commanding, like in "Open the window". Rather we do it indirectly by checking whether the interlocutor is able or willing to open the window, like in "Can you open the window for me?" It would, then, be inappropriate for the interlocutor to answer "Yes" without further action. In that case, the response is only to the question (whether he or she is able to open the window), but not to the request.

How does the listener know that there is a request in addition to the question? There is, of course, an enormous amount of shared situational knowledge that will do the work. Grice (1975) has argued that conversations are governed by principles of rationality; Sperber and Wilson (1986) call it the *principle of relevance*. The interlocutor, for instance, is so obviously able to open the window that the speaker's intention cannot have been to check that ability. But Clark (1979) found that linguistic factors play a role as well. If the question is phrased idiomatically, involving *can* and *please*, subjects interpret it as a request. But the less idiomatic it is (like in "Are you able to . . ."), the more subjects react to the question instead of to the request.

Another important aspect of conversation is *turn-taking*. There are rules for the allocation of turns in conversation that ensure everybody's right to talk, that prevent the simultaneous talk of different parties, and that regulate the proper engaging in and disengaging from conversation (Sacks, Schegloff, & Jefferson, 1974). These rules are mostly followed, and sometimes intentionally violated (as in interrupting the speaker). Turn-taking is subtly controlled by linguistic (especially prosodic) and non-verbal (gaze and body movement) cues (Beattie, 1983).

THE MENTAL LEXICON

Producing or understanding spoken language always involves the use of words. The mental lexicon is our repository of words, their meanings, their syntax, and their sound forms. A language's vocabulary is, in principle, unlimited in size. Take, for instance, the numerals in English. They alone form an infinite set of words. But it is unlikely that a word such as *twenty-three-thousand-two-hundred-and-seventy-nine* is an entry in our mental lexicon. Rather, such a word is constructed by rule when needed. We have the ability to produce new words that are not stored in our mental lexicon.

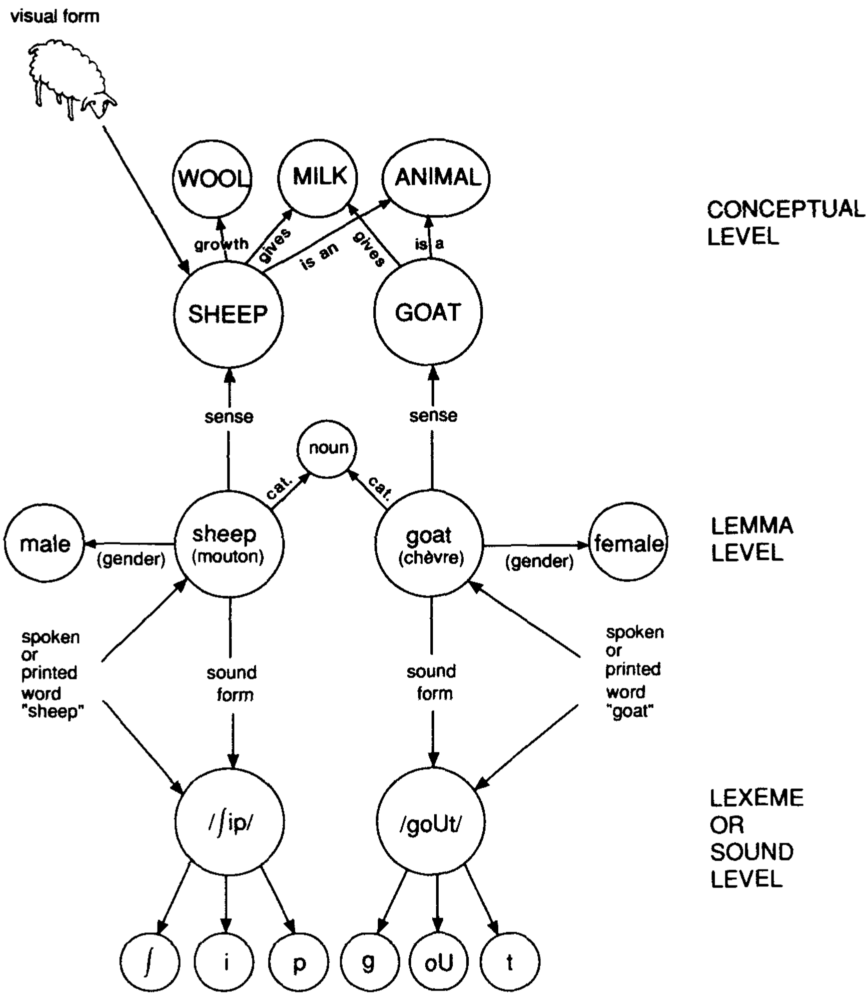

Figure 1 Fragment of a lexical network. Each word is represented at the conceptual, the syntactic and the sound form level
Source: Bock and Levelt, 1993

How many words *are* stored? Miller (1991) estimates that the average high school graduate knows about 60,000 words (under one definition of "word").

One way of representing this enormous body of knowledge is by way of network models. Figure 1 shows a fragment of such a network. Each word is represented by three nodes, one at the conceptual level, one at the syntactic (grammatical) or lemma level, and one at the sound form (phonological) or lexeme level. The lemma is the syntactic representation and the lexeme is the phonological representation. A word's semantic properties are given by its connections to other nodes at the conceptual level (for instance, that a sheep is an animal, gives milk, etc.). A word's syntactic properties are represented by its lemma node's relations to other syntactic nodes (for instance, "sheep" is a noun; French "mouton" has male gender, etc.). The sound form properties, finally, such as a word's phonological segments, are represented in the way a word's lexeme node relates to other sound form nodes ("sheep" for instance contains three ordered phonological segments, /ʃ/, /i/, and /p/, as shown in Figure 1).

Different authors have proposed different network models (e.g., Collins & Loftus, 1975; Dell, 1986; Roelofs, 1992), and for different purposes. It is unlikely that such networks can adequately represent all complexities of our semantic, syntactic, and phonological knowledge about words. But they can be useful in predicting speed of word access in comprehension and production, as well as in explaining various kinds of errors that we make in speech production and various disorders of accessing words in aphasic speech.

Especially important for theories of language use are the ways that verbs are represented in the mental lexicon. As a semantic entity, a verb assigns semantic roles to its arguments. The verb *walk*, for instance, requires an animate argument that specifies the role of agent, as in *John walked.* The verb *greet* governs two arguments, one for the agent and one for the recipient of the action, as in *Peter greeted the driver.* As a syntactic entity, a verb assigns syntactic functions to the sentence constituents it governs. In the above sentence, *Peter* is the subject and *the driver* the object. A verb's argument-function mapping is not random. Most verbs, for instance, map a recipient argument on to a syntactic object function, but not all. The verb *receive* doesn't. In *Mary received the book, Mary* is both recipient and sentence subject. Also, verbs often allow for multiple mappings. In *the driver was greeted by Peter*, the recipient, not the agent appears in subject position.

For each verb, the mental lexicon contains its possible mapping frames. These play an important role in the speaker's syntactic planning and in the listener's syntactic and semantic parsing.

SPEAKING

Speaking is our most complex cognitive-motor skill. It involves the conception of an intention, the selection of information whose expression will make that intention recognizable, the selection of appropriate words, the construction of a syntactic framework, the retrieval of the words' sound forms, and the computation of an articulatory plan for each word and for the utterance as a whole. It also involves the execution of this plan by more than 100 muscles controlling the flow of air through the vocal tract. Finally, it involves a process of self-monitoring by which speech trouble can be prevented or repaired. The following is a bird eye's view over these processes.

Conceptual preparation

The question where communicative intentions come from is a psychodynamic question rather than a psycholinguistic one. Speaking is a form of social action, and it is in the context of action that intentions, goals, and subgoals develop. It is not impossible, though, that the intention *what* to say occasionally arises from spontaneous activity in the speech formulating system itself. It can create rather incoherent "internal speech", which we can self-perceive. This, in turn, may provide us with tatters of notions that we then consider for expression (cf. Dennett, 1991).

Conveying an intention may involve several steps or "speech acts". The speaker will have to decide what to express first, what next, and so on. This is called the speaker's *linearization* problem (Levelt, 1989). It is especially apparent in the expression of multidimensional information, as in describing one's apartment (Linde & Labov, 1975). The conceptual preparation of speech, and in particular linearization, require the speaker's continuing attention. The principles of linearization are such that attentional load is minimized.

Each speech act, be it a request to do *X*, an assertion that *Y*, etc., involves the expression of some conceptual structure, technically called a "message" (Garrett, 1975). That message is to be given linguistic shape; it has to become "formulated".

Grammatical encoding

A first step in formulating is to retrieve the appropriate words from the mental lexicon and to embed them in the developing syntactic structure. In normal conversation we produce some two words per second. At this rate we manage to access the appropriate words in our huge mental lexicon. Occasional errors of lexical selection (such as "Don't burn your toes" where *fingers* was intended) show that the lexicon has a semantic organization.

The standard explanation for such errors is that activation spreads through a semantically organized network, as in Figure 1. In such a network, each node has an activation level between 0 and 1. When the lexical concept node SHEEP is active, then activation spreads to semantically related concept nodes, such as GOAT. Both nodes spread activation "down" to their lemma nodes. Which one of the lemmas will then be selected for further processing? Normally it will be the most activated one, in this case the lemma for "sheep". But the occurrence of an occasional error shows that there is a small probability that a less activated lemma gets selected. According to one theory (Roelofs, 1992) the probability that a particular lemma becomes selected within a time interval *t* is the ratio of its activation to the sum of the activation of all other lemma nodes. For instance, if "sheep" and "goat" are the only two active lemmas during interval *t* after presentation of the picture, and they have activation levels of 0.7 and 0.1 respectively, the probability that the target word "sheep" will be selected during that interval is 7/8, whereas the erroneous word "goat" will be selected with the probability 1/8. Hence, if there is more than one lemma active in the system, there is always a small probability that a non-intended word becomes selected (and it is likely to be semantically related to the target).

Spreading activation theories of lexical selection are typically tested in picture-naming experiments, where naming latencies are measured. For a review of issues in lexical selection, see Levelt (1992a).

As soon as a lemma is retrieved, its syntactic properties become available. Among them are the lemma's grammatical class (preposition, noun, verb, etc.). Each lemma requires its own specific syntactic environment or "frame". Syntactic planning is like solving a set of simultaneous equations. Each lemma's frame has to fit its neighbour's frames, and since Garrett (1975) there are theories about how this is realized (see Levelt, 1989, for a review). Actually, the equations are not quite "simultaneous"; the lemmas for an utterance are typically not concurrently retrieved. Lemmas for salient concepts, such as animate objects, tend to be retrieved faster than for non-salient concepts (Bock & Warren, 1985), and that affects their position in the developing syntactic structure. For a review of grammatical encoding, see Bock and Levelt (1994).

Phonological encoding

A selected lemma (but only a selected one: see Levelt et al., 1991) spreads its activation to its lexeme node (cf. Figure 1). At this level two kinds of phonological information become available. The first one is the word's segments, which are "spelled out" one after another. The second one is the word's metrical structure. For "sheep" it is the information that it is a one-syllable word. For "father" it is the information that it is a two-syllabic trochaic word. The metrical frames of successive words are often combined, creating so-called phonological word frames. In *Peter gave him it*, the last three words form one phonological word *gavimit*. In a process of *segment-to-frame association* spelled-out segments are inserted one by one into the corresponding phonological word frames. It is during this ordered insertion that phonological syllables are created, one after another (such as *ga-vi-mit*; see

Levelt, 1992b). How this string of phonological syllables determines the precise articulatory gestures to be made by the speech organs is still a matter of much debate (see especially Browman & Goldstein, 1991).

The notion that segments and frames are independently retrieved arose in the analysis of phonological speech errors (Dell, 1986; Shattuck-Hufnagel, 1979). Spoonerisms such as *with this wing I thee red*, or *fool the pill* (instead of *fill the pool*) show that segments can become associated to the right place in the wrong frame.

Phonological encoding also involves the planning of larger units than phonological words. There is, in particular, the planning of intonational phrases. These are units that carry a particular intonational contour. Such contours can be rising, falling or combinations thereof. They often express a speaker's attitude towards what is said: doubt, certainty, or towards the interlocutor: reassuringness, inviting reaction. See Levelt (1989) for a review of phonological encoding.

The output of phonological encoding is an articulatory programme. Phenomenologically, it appears to the speaker as internal speech. This internal speech need not be articulated. It can be kept in an articulatory buffer, ready to be retrieved for articulatory execution (Sternberg, Wright, Knoll, & Monsell, 1980).

Articulation

The articulatory apparatus consists of three major structures. The respiratory system controls the steady outflow of air from the lungs. The breathing cycle during speech is quite different from normal breathing, with very rapid inhalation and very slow exhalation. The laryngeal system has the vocal cords as its central part. It is the main source of acoustic energy. The vocal tract, finally, contains the cavities of pharynx, mouth, and nose. They are the resonators that filter the acoustic energy in frequency bands or *formants*. Vowels are characterized by their formant structure. The vocal tract can be constricted at different places, and these constrictions can be made or released in different manners. In this way a wide range of consonantal and other speech sounds can be made.

The control of this utterly complex motor system has been the subject of much research. Present theories converge on the notion of *model-referenced control* (Arbib, 1981; see also Figure 2). The motor system is given an "articulatory task" (as part of the articulatory programme), such as "close the lips". There are usually many degrees of freedom in executing such a task. For instance, lip closing can be realized by moving the lips, by moving the jaw, or by doing both to various degrees. The internal model computes the least energy-consuming way of reaching the goal, given the actual state of the articulators (there is continuous proprioceptive feedback to the internal model). The output is a set of efferent control signals to the relevant

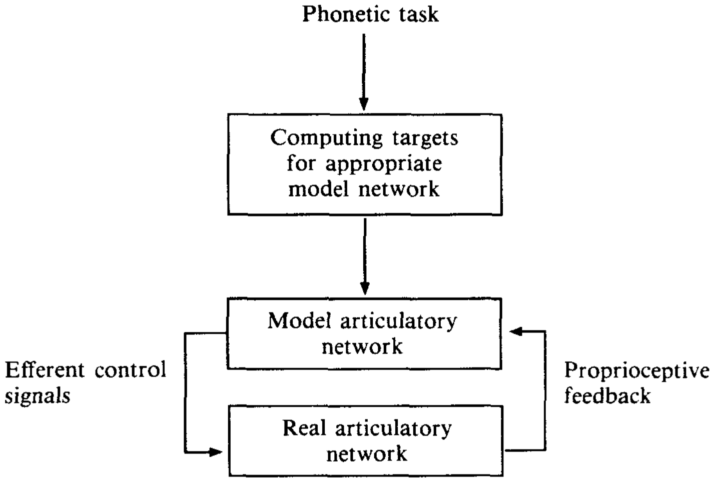

Figure 2 Model-referenced control in articulation
Source: Levelt, 1989

muscles. Saltzman and Kelso (1987) gave a precise mathematical rendering of this theory. See Levelt (1989) for a review of theories of articulation. The output of articulation is overt speech.

Self-monitoring

We can listen to our own overt speech and detect trouble, just as we can listen to the speech of others and detect errors or infelicitous delivery. This involves our normal speech understanding system. We can also detect trouble in our internal speech. When the trouble is disruptive enough for the ongoing conversation, a speaker may decide to interrupt the flow of speech and to make a self-repair.

Not all self-produced trouble (such as errors of selection) is detected by the speaker. Self-monitoring requires attention; we mostly attend to *what* we say, far less to *how* we do it. Detection of trouble is better towards the end of clauses, where less attention for content is required (Levelt, 1989). There are two main classes of trouble that induce repairing. The first one is an all-out error (as in *and above that a horizon-, no a vertical line*); the error can be lexical, syntactic, or phonological. The second one is that something is not really appropriate (as in *to the right is blue – is a blue point*). The speaker then repairs in order to make the utterance more precise, less ambiguous. Upon detecting either kind of trouble, the speaker can self-interrupt. And this ignores linguistic structure; a speaker can stop in the midst of a phrase, a word, or a syllable. But then, the speaker often marks the kind of trouble by some *editing expression*: "no", "sorry", "I mean", for errors; "rather", "that is", for something inappropriate.

Restarting, that is, making the repair proper, is linguistically quite principled. The speaker grafts the repair on to the syntax of the interrupted utterance, which has been kept in abeyance. As a consequence, repairing is like linguistic coordination. One seldom finds a repair such as *is she driving – she walking downtown?* And indeed, the corresponding coordination *is she driving or she walking downtown?* is ill-formed. But *is he – she walking downtown?* is a very common repair type, and it corresponds to a well-formed coordination: *is he or she walking downtown?* (Levelt, 1989).

SPEECH UNDERSTANDING

The canonical objective in speech understanding is to recognize the speaker's communicative intention. How does the listener induce that intention from the speaker's overt speech, a continuous flow of acoustic events?

Several component processes are involved here. First, there is the hearer's acoustic-phonetic analysis of the speech signal, that is, representing it as a *phonetic* not just an *acoustic* event. Second, there is phonological decoding, in particular finding the words that correspond to the phonetic events, and analysing the overall prosodic structure of the utterance. Third, there is grammatical decoding, parsing the utterance as a meaningful syntactic structure. Finally, there is discourse processing, interpreting the utterance in the context of the ongoing discourse, and in particular inferring the speaker's intentions. Let us review these processes in turn.

Acoustic-phonetic analysis

It is very hard, if not impossible, to listen to speech as if it were just a string of chirps, buzzes, hums, and claps. We just cannot help perceiving it as speech. In this so-called "speech mode" (Liberman & Mattingly, 1985) we interpret the acoustic event as resulting from a speaker's articulatory gestures as a phonetic event. There is no unanimity in the literature, though, about what kind of representation the listener derives. According to Liberman and Mattingly, the listener derives the speaker's intended articulatory gestures (even if they were sloppy). Others argue that listeners have special detectors for distinctive events in the speech signal, such as for onsets, for spectral peaks, for the frequencies and motions of formants. The detection of such acoustic events may suffice to derive the presence or absence of phonetic features, such as voicing, nasality, vowel height, stridency, and so on (Stevens & Blumstein, 1981).

Speech segments, clusters, and syllables have characteristic distributions of phonetic features. Hence, if such feature detectors are reliable, they may provide sufficient information for effective phonological decoding. Opinions differ, however, about their reliability. The speech signal is highly variable, dependent as it is on speech rate, sex of the speaker, sloppiness of speech delivery, reverberation or noise in the room, for example. Even if the listener can partial out such effects of the speech context, acoustic-phonetic analysis will often be indeterminate. Still, it may well be sufficient for the purpose. Not every word has to be recognized in order to derive the speaker's intentions. And where a really critical word is missed, the interlocutor will say "what?" or signal difficulty of understanding in other ways.

For an excellent review of acoustic-phonetic processing, see Pisoni and Luce (1987).

Phonological decoding

Whatever the precise character of the phonetic representations, they are the listener's access codes to the mental lexicon. How does a listener recognize words in connected speech? A major problem here is to *segment* the speech, to find out where words begin and end in the continuous flow of speech. There are, basically, two routes here.

The first one is the bottom-up approach, that is, to build on cues in the phonetic representation. Cutler (1990) has argued that English listeners will, by default, segment speech such that there are word boundaries right before stressed syllables. It is a statistical fact of English that 85 per cent of the meaningful words that one encounters while listening begin with a stressed syllable. The segmentation strategy will, therefore, be quite successful. Cutler's theory has meanwhile found substantial experimental support. Also, there are speech sounds that tend to occur at the ends of words, such as [-ng] and [-nd] for English. Speakers may use such phonotactic properties of their language to predict word boundaries.

The second route is top-down. We often recognize a word before it ends. But that means that we can predict the word's end, and hence the upcoming word boundary. That gives us a handle on where to start recognizing the subsequent word.

Given that we know a word's beginning, how do we recognize it? According to the *cohort theory* (Marslen-Wilson, 1989), a small word-initial feature pattern (corresponding to about two segments of the input word) activates all words in the mental lexicon that match it phonologically. Assume the input word is *trespass*, and the cluster [tr] has become available. This will activate all words beginning with [tr], such as *tremble*, *trespass*, *trestle*, *trombone*, etc. This is called the "word-initial cohort". As more phonetic information becomes available, the cohort is successively reduced. When the vowel [ε] is perceived, all items not sharing that vowel, such as *trombone*, are deactivated. This process continues until a single candidate remains. For *trespass* this happens when [p] is reached. The segment [p] is, therefore, called the *uniqueness point* of *trespass*. A word's uniqueness point depends on its word-initial lexical alternatives. For most words the uniqueness point precedes the word's end.

For an optimally efficient system, the word's uniqueness point would also be its *recognition point*. There is good experimental evidence in support of this hypothesis (e.g., Frauenfelder, Segui, & Dijkstra, 1990), though the recognition point may slightly anticipate the uniqueness point in case syntactic or semantic information disambiguates the item from its remaining alternatives (Zwitserlood, 1989). Hence, it will often be possible for a listener to anticipate the upcoming word boundary.

Phonological decoding serves not only the recognition of words, but also their groupings into prosodic constituents, such as phonological and intonational phrases. These constituents carry important information about the syntax of the utterance, and about the communicative intentions of the speaker (cf. Levelt, 1989).

Grammatical decoding

As words are successively recognized and prosodically grouped, the listener will as much as possible interpret these materials "on-line" (Marslen-Wilson & Tyler, 1980). Each recognized word makes available its syntactic and semantic properties. There is, then, concurrent syntactic parsing and semantic interpretation, each following its own principles, but interacting where necessary.

In this connection, one should distinguish between local and global syntactic parsing. Local parsing involves the creation of local phrase structure, combining words into noun phrases, verb phrases, etc. There is increasing evidence that local parsing can run on word category information alone (Frazier, 1989; Tyler & Warren, 1987). We have little trouble parsing "jabberwocky" or semantically anomalous prose such as *the beer slept the slow guitar*. Here we construct phrase structure exclusively by recognizing the words' syntactic categories (Art, Adj, N, V). However, successful local parsing is highly dependent on the intactness of phonological phrases, as Tyler and Warren (1987) could show. For instance, in the above anomalous prose, one should not create a prosodic break between *the* and *slow*, or between *slow* and *guitar*.

Global syntactic parsing, however, interacts with semantic interpretation. In global parsing, semantic roles are assigned to syntactic constituents, and this is to a large extent governed by the verb's argument/function mapping. When the meaning of words or phrases contradicts the semantic roles they should carry, global parsing is hampered (Tyler & Warren, 1987).

One important aspect of global parsing is the resolution of anaphora. In the sentence *the boxer told the skier that the doctor for the team would blame him for the recent injury*, the anaphor *him* can refer back to *the boxer* and to *the skier*, but global syntax prohibits its referring to *the doctor*. Indeed, experimental evidence shows reactivation of both *boxer* and *skier*, but not of *doctor* when the pronoun *him* is perceived. Such reactivation can also be measured for so-called null-anaphors as in *the policeman saw the boy that the crowd at the party accused t of the crime*. Here there is measurable reactivation of *boy* at position *t* (the syntactic "trace" of *the boy*; see Nicol & Swinney, 1989). But also in this respect global parsing is semantically facilitated, for instance if the anaphor's referent is a concrete noun (Cloitre & Bever, 1988).

Grammatical decoding doesn't remove all ambiguity (for instance, the pronoun *him* above is not fully resolved). Here, further discourse processing is needed.

Discourse processing

Partners in conversation construct mental models of the state of affairs they are talking about (Johnson-Laird, 1983; Seuren, 1985). Indefinite expressions (such as in *there is a dog in the room*) make them introduce a new entity (a dog) in the model. Definite expressions (such as *the room* in the same sentence) make them look up an already existing entity. The new information in the utterance is then attached to whichever entity it concerns.

Identifying referents is a major accomplishment of human language processing, still unmatched by any computer program. The problem is that referring expressions can be highly indirect. How can a waitress in a restaurant interpret the referent when her colleague says *the hamburger wants the bill*? Nunberg (1979) argued that there are "referring functions" that map a *demonstratum* (like the hamburger) on to the intended referent (the person who ordered it). But the range of possible referring functions is almost unlimited. Clark, Schreuder, and Buttrick (1983) and Morrow (1986) have argued (and experimentally shown) that such demonstratum-to-referent mapping depends on the mutual knowledge of the interlocutors and on the saliency of entities in their discourse models.

Indirectness is the hallmark of discourse interpretation. As mentioned above, what is said often relates quite indirectly to what the speaker intends to convey. It is not only politeness that governs such indirectness. All figures of speech, whether polite or not, require the listener to build a bridge from the literal to the intended. This holds equally for metaphor (Sperber & Wilson, 1986), irony (Clark & Gerrig, 1984), and hyperbole (Grice, 1975).

Finally, whereas acoustic-phonetic, phonological, and grammatical decoding are largely automatic processes, discourse processing requires the listener's full attention. In that respect, it is on a par with the speaker's conceptual preparation. As interlocutors we are concerned with content. The processing of form largely takes care of itself.

READING

The invention of writing systems, whether logographic, syllabic, or alphabetic, is probably the most revolutionary step in human cultural evolution. It added a powerful means of storing and transmitting information. With the invention of printing, it became a major mechanism for large-scale dissemination of knowledge in a culture.

But equally surprising as this ability to map spoken language on to a visual code is our capacity to efficiently process such a code. When skilled, we silently read five or six printed words per second; this is about twice the rate of conversational speech. This ability has not given us any selective advantage in biological evolution; the invention of writing systems is as recent as about 5,000 years ago. Rather, the ability to read must be due to a happy coincidence of other pre-existing faculties of mind.

One of these is, of course, language. As readers we largely use our parsing potential for spoken language. Visual word recognition feeds into the lemma level of Figure 1. As lemmas are successively activated by the printed words, further syntactic, semantic, and discourse processing operates roughly as for spoken language. There are, admittedly, differences too. There is, for instance, no prosody to help syntactic parsing; instead there is punctuation. Also, there is no external enforcement of rate as there is in speech perception.

Another pre-existing faculty on which reading is parasitic is our enormous ability to scan for small meaningful visual patterns. In a hunter's society these were probably animal silhouettes, footprints, and so on. Words (if not too long or too infrequent) are recognized as wholes; a skilled reader processes a word's letters in parallel. Much ink has been spilled on the question whether the letters individually or the word as a whole activate a phonological code in silent reading, that is, the word's lexeme (see Figure 1). Such phonological recoding indeed exists. But it is only for low-frequent words that this "phonological route" is of any help in lemma access (Jared & Seidenberg, 1991). However, this silent "internal speech" probably does play a role in further syntactic and semantic parsing; it is a way of buffering successive words for further processing.

The ability to scan is optimally used in reading. The basic cycle is this: the reader fixates a word for, on average, one-fifth of a second. The fixation is roughly between the beginning and the middle of the word. During this period lexical access is achieved. In addition, there is some perception of the next word in the periphery of vision. Sometimes this suffices to recognize that next word as well on the same fixation (but the fixation will then last somewhat longer). Usually, however, the information from the periphery of vision is used only to plan a saccadic eye movement (a jump of the eye) to that next word. The size of the saccade depends on the length of the next word; the average saccade is about eight characters in size. The new word is fixated, and the cycle starts all over again.

When a word is quite infrequent, or when the reader has trouble integrating it in the developing syntax or semantics, the fixation duration can be substantially longer. Also, the reader may backtrack and refixate an earlier word when there is serious trouble in comprehension.

For a major review of the reading process and its disorders, see Rayner and Pollatsek (1989).

SIGN LANGUAGE

Contrary to written language, the sign languages of deaf people are not parasitic on spoken language. They are autonomous languages in the visual mode. Their mere existence shows that our faculty of language is not crucially

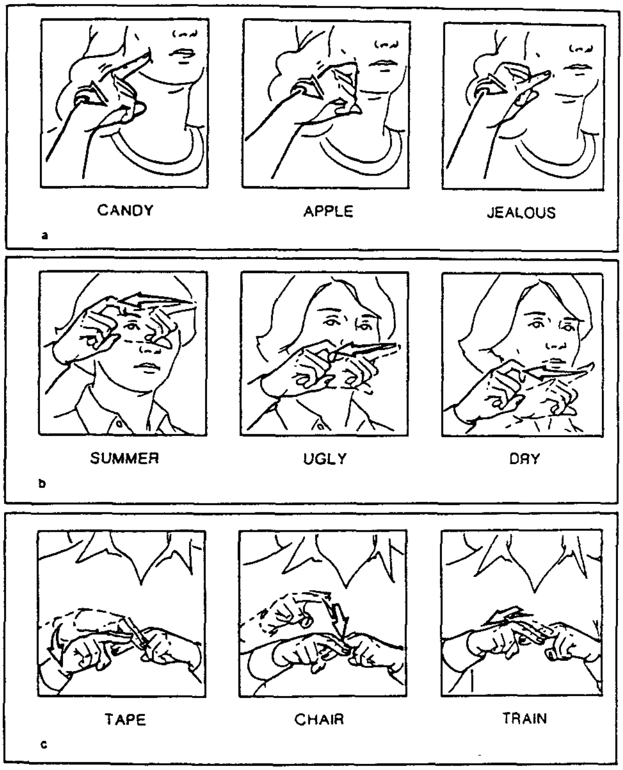

Figure 3 Minimal contrasts between signs in American Sign Language: (a) hand configuration, (b) place of articulation, (c) movement
Source: From Klima and Bellugi, 1979

dependent on our ability to speak. Deaf children who grow up in a signing deaf community acquire their language at the same age and in roughly the same stages as hearing children do.

Just as words, signs have form and meaning. The articulators of sign language are the hands, the face, and the body. Where words contrast phonemically (for instance in voicing: *bath* vs *path*), signs contrast in hand configuration, in place of articulation and in hand movement (see Figure 3). Also, facial features may distinguish between signs.

Although the first coining of a sign is often iconic, its meaning is eventually independent of its form, as it is for words in spoken languages. As a consequence, sign languages are mutually unintelligible, just as spoken languages are (contrary to what Wundt suggested in *Die Sprache* – see above).

Sign languages are rich in morphology (for inflection and for derivation of new signs) and have full-fledged recursive syntax. Many syntactic devices are spatial in character. Anaphora, that is, referring back to an earlier introduced entity, is done by pointing to the locus in the signing space (in front of the body) where the original referent was first "established". In American Sign Language the sign for transitive verbs either moves from subject to object locus, or from object to subject locus. Each verb has its own "mapping function" (like in spoken language, see above). For the structure and use of British Sign Language, see Kyle and Woll (1985).

There is increasing evidence that a sign language is subserved by the same areas of the brain that sustain spoken language. Poizner, Klima, and Bellugi, (1987) showed that damage to anterior areas of the left hemisphere in native signers resulted in a style of signing highly comparable to the agrammatism of so-called Broca's patients. Similarly, a form of fluent aphasia resulted when the damage was in a more posterior area of the left hemisphere, comparable to the fluent aphasia of so-called Wernicke's patients. Damage in the right hemisphere left the signing intact, but patients lost the ability to sign coherently *about* spatial relations, such as the layout of their apartment. Their spatial representations were damaged, but not their spatial language.

FURTHER READING

Garman, M. (1990). *Psycholinguistics*. Cambridge: Cambridge University Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: Massachusetts Institute of Technology Press.

Miller, G. A. (1991). *The science of words*. New York: Scientific American Library.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.

Slobin, D. I. (Ed.) (1985). *The crosslinguistic study of language acquisition* (2 vols). Hillsdale, NJ: Lawrence Erlbaum.

REFERENCES

Arbib, M. A. (1981). Perceptual structures and distributed motor control. In V. Brooks (Ed.) *Handbook of physiology: The nervous system. Motor control* (vol. 2, pp. 1449–1480). Bethesda, MD: American Physiological Society.

Austin, J. L. (1962). *How to do things with words.* Oxford: Clarendon.

Beattie, G. (1983). *Talk: An analysis of speech and non-verbal behaviour in conversation.* Milton Keynes: Open University Press.

Bock, J. K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.) *Handbook of psycholinguistics.* New York: Academic Press.

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition, 21*, 47–67.

Browman, C. P., & Goldstein, L. (1991). Representation and reality: Physical systems and phonological structure. *Haskins Laboratory Status Report on Speech Research*, SR-105/106, 83–92.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage.* Cambridge: Cambridge University Press.

Chomsky, N. (1968). *Language and mind.* New York: Harcourt Brace & World.

Clark, H. H. (1979). Responding to indirect speech acts. *Cognitive Psychology, 4*, 430–477.

Clark, H. H., & Gerrig, R. J. (1984). On the pretence theory of irony. *Journal of Experimental Psychology: General, 113*, 121–126.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1–39.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstratives. *Journal of Verbal Learning and Verbal Behavior, 22*, 245–258.

Cloitre, M., & Bever, T. G. (1988). Linguistic anaphors, levels of representation, and discourse. *Language and Cognitive Processes 3*, 293–322.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407–428.

Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In G. Altmann (Ed.) *Cognitive models of speech processing* (pp. 105–121). Cambridge, MA: Massachusetts Institute of Technology Press.

Dell, G. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93*, 283–321.

Dennett, D. C. (1991). *Consciousness explained.* Boston, MA: Little, Brown.

Frauenfelder, U., Segui, J., & Dijkstra, T. (1990). Lexical effects in phoneme processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance, 16*, 77–91.

Frazier, L. (1989). Against lexical generation. In W. D. Marslen-Wilson (Ed.) *Lexical representation and process* (pp. 505–528). Cambridge, MA: Massachusetts Institute of Technology Press.

Fry, D. (1977). *Homo loquens.* Cambridge: Cambridge University Press.

Galton, F. (1879). Psychometric experiments. *Brain, 2*, 149–162.

Garrett, M. F. (1975). An analysis of sentence production. In G. Bower (Ed.) *Psychology of learning and motivation* (vol. 9, pp. 133–177). New York: Academic Press.

Grice, H. P. (1968). Utterer's meaning, sentence meaning and word meaning. *Foundations of Language, 4*, 225–242.

Grice, H. P. (1975). Some further notes on logic and conversation. In P. Cole (Ed.) *Syntax and semantics: Pragmatics* (vol. 9, pp. 113–127). New York: Academic Press.

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General, 120*, 358–394.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.

Klima, E. S., & Bellugi, U. (1979). *The signs of language*. Cambridge, MA: Harvard University Press.

Kyle, J. G., & Woll, B. (1985). *Sign language*. Cambridge: Cambridge University Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: Massachusetts Institute of Technology Press.

Levelt, W. J. M. (Ed.) (1992a). Lexical access in speech production. Special issue of *Cognition*, *42*, 1–316.

Levelt, W. J. M. (1992b). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*, 1–22.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*, 122–142.

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

Linde, C., & Labov, W. (1975). Spatial networks as a site for the study of language and thought. *Language*, *51*, 924–939.

Marslen-Wilson, W. (1989). Access and integration: Projecting sound onto meaning. In W. D. Marslen-Wilson (Ed.) *Lexical representation and process* (pp. 3–24). Cambridge, MA: Massachusetts Institute of Technology Press.

Marslen-Wilson, W., & Tyler, L. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1–71.

Miller, G. A. (1991). *The science of words*. New York: Scientific American Library.

Morrow, D. G. (1986). Places as referents in discourse. *Journal of Memory and Language*, *25*, 676–690.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*, 5–19.

Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy*, *3*, 143–184.

Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, *25*, 21–52.

Poizner, H., Klima, E. S., & Bellugi, U. (1987). *What the hands reveal about the brain*. Cambridge, MA: Massachusetts Institute of Technology Press.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice-Hall.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, *50*, 696–735.

Saltzman, E., & Kelso, J. A. S. (1987). Skilled action: A task-dynamic approach. *Psychological Review*, *94*, 84–106.

Seuren, P. A. M. (1985). *Discourse semantics*. Oxford: Basil Blackwell.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds) *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale, NJ: Lawrence Erlbaum.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition.* Oxford: Basil Blackwell.

Sternberg, S., Wright, C. E., Knoll, R. L., & Monsell, S. (1980). Motor programs in rapid speech: Additional evidence. In R. A. Cole (Ed.) *Perception and production of fluent speech* (pp. 507–534). Hillsdale, NJ: Lawrence Erlbaum.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds) *Perspectives on the study of speech* (pp. 1–38). Hillsdale, NJ: Lawrence Erlbaum.

Tyler, L., & Warren, P. (1987). Local and global structure in spoken language comprehension. *Journal of Memory and Language, 26*, 638–657.

Wundt, W. (1900). *Die Sprache* (2 vols). Leipzig: Kröner.

Zwitserlood, P. (1989). The effects of sentential-semantic context in spoken-word processing. *Cognition, 32*, 25–64.