Language

Willem J.M. Levelt

Language is the species-specific communication system of *homo sapiens*. Any normal person acquires at least one natural language, whose basic structure is fully developed at the age of six or seven. The innate ability to converse has provided us with the capacity to share moods, emotions and information of almost any kind, to plan joint action, to instruct, to educate our offspring and generally to transmit culture. The basic property of language that allows for this sheer infinite use is its *generativity*. Although the basic ingredients of a language are finite in number, their combinatorial possibilities are without limit. Each language has a small and fixed set of distinctive speech sounds, called *phonemes* (in particular consonants and vowels), for known languages ranging from as few as 11 to 141; not a single one of these is shared by all languages. A child acquires its local repertoire essentially during the first year of life. Also, each language has a limited set of meaning-bearing elements or *morphemes*, the basic ingredients of words. (The English word *follow* consists of one morpheme, *follow-ed* consists of two morphemes, where *follow* is a word itself, but *-ed* is only a bound morpheme). A normally educated English-speaking adult may easily know some 20,000 morphemes; over 95% of them are words themselves.

A language's generativity is 2-fold, lexical and syntactic. We can combine morphemes to create new words (*vaccins→vaccinatepre→vaccinate→prevaccination→anti-prevaccination* or *five-million-three-hundred-fifty-two-thousand-six-hundred-seventy-nine*); this is lexical generativity. Many of these complex words are also stored in memory; the just-mentioned adult may know some 40–80,000 words. If that adult is 20 years old, he or she will, on average, have acquired some 6 to 12 words *per day* since birth (Miller, 1991).

We can combine words to create new phrases and sentences (*it is raining→my sister thinks it is raining→Peter believes that my sister thinks it is raining→if Peter believes that my sister thinks it is raining, he is mistaken→...*). This is syntactic generativity. Many thousands of these phrases and sentences are idiomatic and stored in memory (*good morning, kick the bucket, the fat is in the fire*, etc.).

Both lexical and syntactic generativity are strictly rule-governed within a language. These rules display a fairly limited number of patterns over the languages that have been analyzed (which is a minority of the 5–10,000 existing languages). This small set of basic patterns is called *universal grammar*; it is a challenging conjecture that universal grammar is somehow genetically preprogrammed in the human brain (Chomsky, 1980).

1. Language in discourse

In discourse two or more interlocutors interact by means of language. This joint activity is usually intentional, a means to accomplish something, such as to establish a joint belief, to plan some action, to invite sympathy, or whatever. This typically involves taking turns, which makes it possible for participants to limit attention to one contribution at a time and to receive attention when contributing themselves. The study of discourse and conversational analysis are active fields of inquiry (Clark, 1996). Turn taking implicates that interlocutors contribute by acts of both speaking and listening. In the following we will consider these processes in turn. Figure 1 . presents the basic architecture of spoken language use, i.e., speaking and listening.

Basic architecture of speaking (message generation, formulating, articulating) and speech comprehension (acoustic-phonetic processing, parsing and discourse processing).

2. Speaking

The generation of fluent speech involves the cooperation of various processing components, each operating in relative autonomy (Levelt, 1989). *Conceptual preparation.* In order to achieve his communicative goals, the speaker should express information that will affect the interlocutor in the intended manner. A major aspect of conceptual preparation consists of selecting and organizing such effective information for expression. In order to achieve this, speakers can resort to a rich repertory of rhetorical devices, such as asserting, commanding, requesting, promising, threatening, apologizing, etc. A major variable here is the directness by which the intention gets expressed. A direct command (such as *lend me ten dollars*) may be far less effective than an indirect request (such as *could you lend me ten dollars?*) where the speaker only checks the interlocutor's *ability* to perform the desired action. Such dances of politeness are very similar across cultures. The speaker's ultimate message to be expressed must consist of *lexical concepts*, such as the notions "lend" or "dollar"; they are concepts for which there are words in the language.

Grammatical encoding. A first step in formulating the message is to select the appropriate words from the mental lexicon and to arrange them morpho-syntactically. In normal fluent speech, a speaker selects some two or three words per second from a mental lexicon that contains tens of thousands lexical items. At this speed of processing, the error rate is still negligible, somewhere around 1‰. Lexical selection is modelled as a process of activation spreading from concept nodes to word (or *lemma*) nodes in a *lexical network*. Since activation can spread among related concepts, the rare selectional error tends to be semantic in nature (like *penny* for *dollar*). However, lexical selection may be in jeopardy in aphasia. Each selected lemma has a syntactic specification. It is a noun (count, mass, proper name, etc.), verb (transitive, intransitive, etc.), adjective or other syntactic category. These specifications trigger the syntactic procedures that generate the appropriate syntactic pattern. These procedures loose their automaticity in Broca patients.

Phonological encoding. Upon selection of a word, its phonological make-up (its *lexeme*) is activated in the mental lexicon. The word's segments (consonants, vowels) and its metrics (such as its accent pattern) are independently retrieved. Word finding problems often occur at this level. In the so-called *tip-of-the-tongue* state, we may become aware of the word's accent pattern or its initial segment. Most anomia patients are specifically handicapped in accessing word forms (with rather intact lemma access). Normal speakers are slower in retrieving low-frequent word forms than high-frequent word forms.

The metrical frames of adjacent words are often merged. In *police demand it*, for instance, the weak-strong frame of *demand* merges with the weak frame of *it*, to create a three-syllable weak-strong-weak frame. The retrieved segments are attached to this larger frame in temporal succession "from left to right", on the fly producing the appropriate syllabification: *de-man-dit*. Syllables often straddle lexical boundaries (as is the case for *-dit*). Occasionally, segments end up in the wrong frame or position, creating speech errors such as *heft lemisphere*.

Upon formation of a phonological syllable (such as *de*, or *man* or *dit*), its articulatory specification or *gestural score* is retrieved from memory (from a *syllabary*) or independently composed. Accessing syllabic articulatory patterns can be particularly disordered in jargon aphasias.

Phonological encoding also involves the generation of larger metrical groupings and of intonation contours. These are expressive of syntactic structure, attentional focus and mood. The boundary tone of an intonational contour (i.e., the pitch movement of the final syllable) is highly expressive of the speaker's intention to continue or to invite a turn switch.

Articulation. Gestural scores are executed by the articulatory system. It consists of the respiratory system, which provides the source of acoustic energy, the laryngeal system (with the vocal folds as its central parts), which controls voicing and loudness, and the supralaryngeal system or vocal tract, whose resonance chambers (pharyngeal, oral, nasal) control the timbre of vowels and consonants and whose articulators (tongue, velum, lips) control the proper articulation of speech segments. In fluent speech we produce 10–15 segments per second. This hair-raising speed proceeds from the simultaneous, overlapping gesturing of the different articulators. Their coordination is controlled by a wide range of cerebral structures. This coordination is the subject of theories of speech motor control.

3. Understanding speech

The mechanisms involved in speech understanding are depicted in the right half of Figure 1. The listener's input is the acoustic pattern produced by the speaker's articulators. In the ideal case, they will reveal the speaker's intention. This involves various steps of analysis on the part of the listener.

Acoustic-phonetic processing. We need no speech recognition to tell a speech sound from a non-speech sound. The categorizing of an acoustic signal as "speechy" is an immediate sensation, which signals the existence of a specialized acoustic-phonetic processor. It interprets a sound pattern as a phonetic event, consisting of temporally distributed articulatory features, such as voicing, nasality, stridency, sonority and place of articulation. The resulting phonetic representation forms the input to further parsing of the speech signal.

Phonological decoding. A first major problem in speech decoding is the segmentation of connected speech into words. Different from written language, there are no obvious gaps between spoken words. This is a major stumbling block for artificial speech recognition systems, but not for human language users. In order to achieve segmentation, listeners use the metrics of their language. English listeners, for instance, use as a heuristic that strong syllables (usually accented and containing full vowels) are highly likely to be the initial syllables of words. The strategies are different for Japanese or French listeners (Cutler and Butterfield, 1992).

Given a word-initial stretch, word recognition can be initiated. According to *cohort theory* all words beginning with that stretch are activated in the mental lexicon. When you hear the initial stretch of *slender, sle-*, the word *slender* gets activated, but also *sled, sledge, slept.* However, as soon as the next segment (*n*) appears, this active "cohort" can be reduced to just one; *slender* is the only word in the lexicon that corresponds to *slen-*. The segment *n* is the "uniqueness point" of *slender*. There is substantial experimental evidence that, in clear speech, words are recognized around their uniqueness point (Marslen-Wilson, 1989). The word (or lexeme) is then selected from the lexicon and its syntactic (lemma) and semantic (concept) properties become available for further parsing. Although lexical selection is the main target of phonological decoding, the listener will also use the prosody of the utterance to group the words in smaller or larger phrases, to focus on accented words, etc.

Grammatical decoding. Parsing the utterance proceeds incrementally, "on the fly", as successive words are recognized. Though syntactic and semantic analysis proceed hand in hand, each follows its own principles and is, presumably, subserved by dedicated neural substrates. This relative autonomy of syntactic and semantic parsing not only appears from reaction time measurements, but also from scalp-recorded event-related brain potentials (ERPs) measured during sentence understanding. When you listen to a sentence that contains a semantic incongruity, such as *The girl put the candy in her needle*, this evokes a component with a negative polarity, peaking at about 400 milliseconds after presentation of the incongruent word (*needle* in the example). This so-called N400 component has a reduced amplitude when the critical word is semantically congruent, such as *mouth* instead of *needle* in the above example. The N400 is a *normal* response, elicited by any meaningful word. Its morphology is modulated as a function of the amount of effort a listener or reader exerts in integrating a word in the prevalent semantic context. Its amplitude is, for instance, greater when the word *pocket* is presented instead of *mouth* in the above sentence.

A quite different electrical response is triggered by a syntactic anomaly. It is a positive polarity component, peaking around 500–600 msec after the critical word in the sentence. For instance, when you listen to the sentence *The child throw the toys on the floor*, the word *throw* evokes this positive polarity response. It is, however, absent when the syntactically correct form *throws* is used. This so-called syntactic positive shift (SPS) is evoked by quite diverse syntactic violations, but not by semantic trouble. SPS and N400 can be independently evoked, further testifying to the relative autonomy of semantic and syntactic processing in the brain.

The on-line nature of sentence parsing is especially apparent from studies in ambiguity and anaphora resolution. When attending to the sentence *I pulled the fish up to the bank* both meanings of *bank* become temporarily activated, but the semantic context suppresses the inappropriate meaning within 100–200 ms. Similarly, syntactic context can immediately restrict the interpretation of a pronoun, as in the case of listening to a sentence like *The boxer told the skier that the doctor of the team would blame himself for the recent injury*, where the word *himself* reactivates the meaning of *doctor*, but not of either *boxer* or *skier*.

Discourse processing. Although grammatical decoding derives a meaningful message from the input, further processing is often needed to derive from this message the speaker's communicative intention. Usually, the interpretation of an utterance is context dependent. Which skier or boxer is referred to in the above utterance? The listener can only know by relating the utterance to what the speaker has said before or to a shared perceptual situation. As a listener you build up a so-called *discourse model* of the state of affairs discussed; each new utterance is then interpreted in terms of the current state of this model. Discourse integration is often distorted in autism, schizophrenia or Alzheimer disease.

Self-monitoring. Finally, we are always listening to our own speech, comparing what we are producing to what we intended to produce. This continuous self-monitoring is not essential for the production of speech; however, when we detect a serious error of delivery, we can interrupt ourselves and make a self-repair (Levelt, 1989).

4. See also

Speech development
Psycholinguistics
Language, neurology of
Language development
Language evolution
Language, nonhuman

Author's website:
http://www.mpi.nl

5. Further reading

Caplan D (1992): Language: Structure, Processing and Disorders. publisher: MIT Press Cambridge, MA

Garnsey SM *Language and Cognitive Processes, 8, issue 4.* Special issue on event-related brain potentials in the study of language

Gernsbacher MA.(1994): Handbook of Psycholinguistics publisher: Academic Press San Diego, CA

Levelt WJM.(1993) Lexical Access in Speech Production. publisher: Blackwell Cambridge, MA

Pinker S (1994): The Language Instinct. publisher: The Penguin Press London

6. References

Chomsky N (1980): Rules and representations. Behav Brain Sci volume: 3 pp 1-61

Clark H (1996): Language Use publisher: Cambridge University Press Cambridge

Cutler A, Butterfield S (1992): Rhythmic cues to speech segmentation: Evidence from junction misperception. J Mem Lang volume: 31 pp 218-236

Levelt WJM (1989): Speaking: From Intention to Articulation.publisher: MIT Press Cambridge, MA

Marslen-Wilson W. (1989): Lexical Representation and Processes. publisher: MIT Press Cambridge, MA

Miller GA (1991): The Science of Words. publisher: Scientific American Library New York

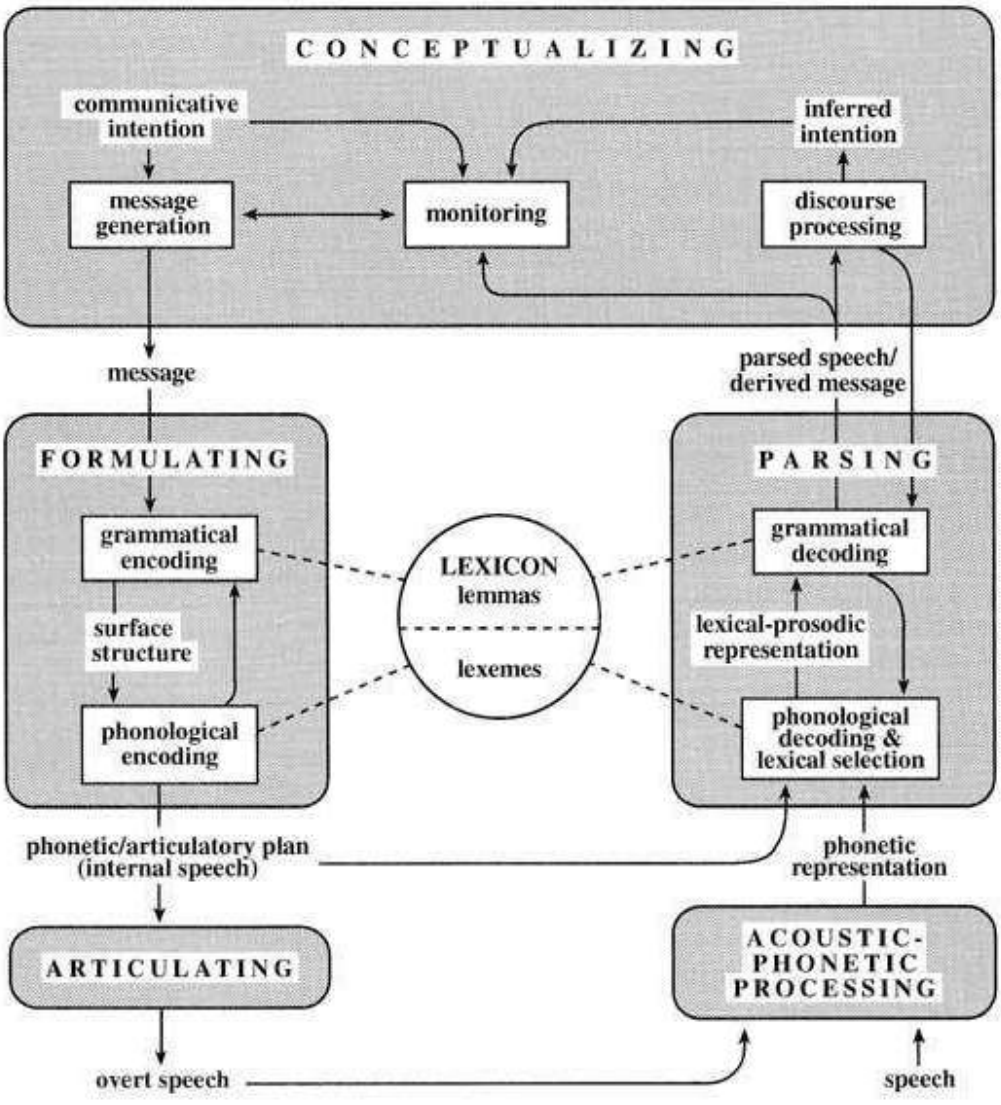

Figure 1. Basic architecture of speaking (message generation, formulating, articulating) and speech comprehension (acoustic-phonetic processing, parsing and discourse processing).