Speech boundaries, syntax and the brain

Cyma Van Petten and Paul Bloom

Speech comprehension requires rapid decoding of grammatical relationships. Electrical scalp recordings show that the brain responds immediately to intonational cues signifying phrase boundaries. Thus, these cues may control initial decisions about sentence structure.

Language comprehension consists of the successful transformation of physical signals into abstract and meaningful information. From the same set of air-pressure waves, we might understand a new idea, determine the speaker's age, emotional state and educational level, and add a new word to our vocabulary. Understanding how the human brain accomplishes this feat requires a careful description of the stages of this transformation, beginning with the physical signal. In speech sounds, changes in frequency and amplitude occur on two time scales. Rapid changes occurring within 20–200 ms convey information about the individual phonemes that make up words, whereas slower prosodic modulations of frequency and amplitude typically extend over multiple words. Computer-synthesized speech demonstrates both the adequacy and the limitations of phonological information alone. Although intelligible, synthetic speech is difficult to understand because it lacks prosody or intonation cues.

Prosody has been granted a limited role in most theories of language comprehension. For instance, rising versus falling pitch toward the end of a sentence distinguishes questions from statements, and sarcastic intonation indicates that a speaker's intended meaning is the opposite of the literal one. In these traditional examples, prosody serves only to modify sentence meanings that have already been computed. In this issue of *Nature Neuroscience* (pages 191–196), Steinhauer and colleagues use event-related potential (ERP) recordings to provide persuasive evidence that prosody is far more central to language comprehension. Their data indicate that speakers use prosody to disambiguate potentially problematic sentences, and that listeners use this information almost immediately.

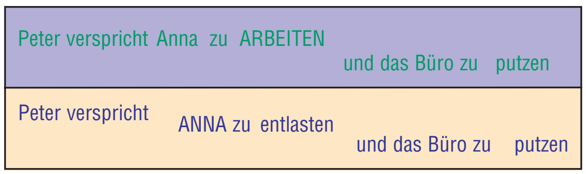

Fig. 1. An example of the German sentences used in the experiments of Steinhauer and colleagues, illustrating how the pauses between words differ in the two conditions.

To see how this works, consider the following poorly written sentence: "Since Jay always jogs five miles seems like a short distance to him." In the absence of a comma after "jogs", readers find this sentence difficult to interpret. They are led down the 'garden path' of assuming that a noun phrase ("five miles") following a verb ("jogs") is the object of that verb rather than the subject of a later verb ("seems"). Given this potential ambiguity, why don't readers simply postpone decisions about the interpretation of the noun phrase until they reach the end of the sentence? This strategy avoids the garden path, but at the cost of imposing a memory burden of holding the unassigned material in mind while continuing to read. Working memory capacity is severely limited[1], and it is needed for other aspects of sentence comprehension, such as linking pronouns to their antecedents[2]. Without contrary evidence (such as a comma), linking a noun with the earliest available verb is an efficient processing strategy. Moreover, the consequences of being led down the garden path by a less-than-thoughtful writer are fairly trivial—at worst, one might have to reread the sentence.

In contrast, listeners lack the luxury of instant replay and may be severely misled by an incorrect initial parse. So why aren't we chronically confused in our daily conversations? Steinhauer and colleagues began by documenting the auditory cues a speaker produces when saying a potentially ambiguous sentence—the verbal equivalent of a comma. They describe no less than four physical differences between German sentences analogous to the example above and those with the more typical verb-then-object structure (as in "Since Jay always jogs five miles this seems like a short distance to him"; **Fig. 1**). These differences in word duration, pause duration, pitch contour and loudness serve to group the words of each sentence into distinct intonational phrases. The critical question is whether this prosodic information can override the usual syntactic preference for assigning a noun to the nearest preceding verb.

Resolution of this question requires an immediate measure of comprehension. Querying listeners at the end of a sentence can reveal only their final analyses, which may not indicate whether they avoided the garden path, or have been there and back. The ERP recorded from the scalp provides a continuous measure of brain electrical activity with millisecond temporal resolution, and so it has been widely used to study language, for which timing is critical. Two components of the ERP elicited by words reflect comprehension difficulties. A negative voltage deflection peaking at about 400 ms after stimulus onset (N400) is small when a word is easily integrated with prior semantic context, but shows graded increases in amplitude as the prior context becomes less and less useful. A word like "sugar" will elicit no N400 at the end of a sentence like "I take my coffee with cream and...", a medium-sized N400 at the end of a sentence like "At the store, I bought a pound of..." and a large N400 in "I finally asked my boss for a ..."[3]. The N400 offers a rapidly changing index of the ease or difficulty of semantic processing, beginning well before a spoken word is even completely pronounced[4]. A second measure of comprehension difficulty is a positive voltage deflection following the N400, termed the P600; large P600s are

The authors are at the Department of Psychology, University of Arizona, Tucson, Arizona 85721, USA. email: bloom@u.arizona.edu and vanpettc@u.arizona.edu

triggered by words that violate rules of agreement, word order or other syntactic principles. The P600 may serve as a general index of surprise at encountering a grammatical problem, or the attempt to revise one's initial parse of a sentence[5].

Steinhauer and colleagues found that when sentences with typical and atypical syntactic structure were presented with normal intonation, there was no sign of an enlarged N400 or P600 during the atypical sentences. In other words, when subjects heard a sentence like "Since Jay always jogs five miles seems like a short distance to him", there was no sign of trouble at the word "seems". A critical third condition pitted the syntactic preference against the prosodic information. The authors spliced the first part of the sentence ("Since Jay always jogs five miles"), said with the intonation suggesting a 'verbal comma' between "jogs" and "five", to the second part of the more common structure ("this seems like a short distance to him"). This created a hybrid condition: initial prosodic cues suggested the atypical sentence structure, while the words specified the more typical verb-then-object construction. In this condition, a large N400 and P600 were elicited by the first word that indicated the sentence's true structure, although this structure would be entirely expected in written language. These results show that listeners take prosodic cues seriously and adjust their syntactic strategies accordingly.

Given the importance of intonational phrases in sentence processing, Steinhauer and colleagues went on to ask whether recognition of prosodic boundaries was reflected in scalp-recorded brain activity. The answer was yes: the authors report a distinct ERP component beginning shortly after the conclusion of each intonational phrase. This is an exciting finding because discovering boundaries is a fundamental process in multiple domains of perception and cognition[6]. The contours defining a visual object must be perceived before the object can be recognized, and neurons early in the visual cortical pathways respond to even occluded contours[7]. Early auditory processes partition the acoustic input into distinct streams of sound arising from different locations; otherwise we might amalgamate the sound of a dog barking next door with water running in the kitchen and a radio playing in the living room into one incomprehensible acoustic event[8]. The short latency of the 'closure positive shift' (CPS), as the authors call this new ERP component, suggests that detection of intonational boundaries is an equally primary process in speech perception. The specificity of the CPS to full intonational phrases remains to be discovered. Prosodic contours are also used to mark word boundaries within a phrase – the distinction between "green house" and "greenhouse". Segmentation of the speech stream into words is a basic problem that must be solved by infants[9,10], so developmental studies of the CPS may shed light on the perceptual strategies and brain processes underlying language acquisition.

The precursor to a complete understanding of human language is a definition of the right units of analysis. Linguists and psychologists have not always agreed on these fundamental issues: whether phonemic features, phonemes or syllables are the smallest speech units[11], or whether words, phrases or sentences are the primitive conceptual unit[12,13]. Physiological methods for monitoring brain activity bring a new and rich dimension of data to bear on these old questions. Other recent studies report brain potentials triggered at syntactic boundaries (clauses) in both spoken and written sentences[14,15]. The discovery of a brain event marking intonational units adds another tool that can be used to understand the complex interplay of phonological, prosodic, syntactic and semantic information in deriving meaning.

1. Miller, G. A. *Psychol. Rev.* **63**, 81–97 (1956).
2. Carpenter, P. A., Miyake, A. & Just, M. A. in *Handbook of Psycholinguistics* (ed. Gernsbacher, M. A.) 1075–1122 (Academic, New York, 1994).
3. Kutas, M. & Hillyard, S. A. *Nature* **307**, 161–163 (1984).
4. Van Petten, C., Coulson, S., Rubin, S., Plante, E. & Parks, M. *J. Exp. Psychol. Learn. Mem. Cogn.* (in press).
5. Coulson, S., King, J. W. & Kutas, M. *Lang. Cogn. Processes* **13**, 21–58 (1998).
6. Bloom, P. *Curr. Directions Psychol. Sci.* **5**, 90–94 (1996).
7. Von der Heydt, R., Peterhans, E. & Baumgartner, G. *Science* **224**, 1260–1262 (1984).
8. Bregman, A. S. *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA, 1990).
9. Jusczyk, P. W. *The Discovery of Spoken Language* (MIT Press, Cambridge, MA, 1997).
10. Bloom, P. in *Handbook of Psycholinguistics* (ed. Gernsbacher, M. A.) 741–780 (Academic, New York, 1994).
11. Remez, R. R. in *Handbook of Psycholinguistics* (ed. Gernsbacher, M. A.) 145–172 (Academic, New York, 1994).
12. Van Petten, C. *Psychophysiology* **32**, 511–525 (1995).
13. Bach, E. *Informal Lectures on Formal Semantics* (SUNY, New York, 1989).
14. Muller, H. M., King, J. W. & Kutas, M. *Cogn. Brain Res.* **5**, 193–203 (1997).
15. Kutas, M. *Psychophysiology* **34**, 383–398 (1997).

Coupling calcium to SNARE-mediated synaptic vesicle fusion

Sabine Hilfiker, Paul Greengard and George J. Augustine

Calcium triggers vesicle fusion, but the molecular mechanism is largely unknown. Now Ilardi and colleagues identify Snapin, a protein that may provide the missing link.

Neurotransmitter release involves a series of interactions between the membranes of synaptic vesicles and the presynaptic terminal, culminating in the calcium-dependent fusion of the two membranes. However, the molecular mechanism by which calcium triggers fusion remains largely unknown.

Sabine Hilfiker and Paul Greengard are at Rockefeller University, 1230 York Avenue, New York, New York 10021, USA. George Augustine is at the Department of Neurobiology, Duke Medical Center, Box 3209, Durham, North Carolina 27710, USA.
email: georgea@neuro.duke.edu

Among the many molecules implicated in neurotransmitter release, most attention has focused on two types of proteins, the SNAREs and synaptotagmin. In this issue of *Nature Neuroscience*, Ilardi and colleagues (pages 119–124) identify and characterize a novel presynaptic protein, called Snapin, which promotes the interaction of SNAREs and synaptotagmin and may be important for synaptic vesicle fusion.

Synaptotagmin, an integral calcium-binding protein of synaptic vesicle membranes, is thought to mediate the action of calcium in neurotransmitter release. Evidence for this notion includes the observa-