

A neural correlate of syntactic encoding during speech production

Peter Indefrey^{*†}, Colin M. Brown^{*}, Frauke Hellwig^{*}, Katrin Amunts[‡], Hans Herzog[‡], Rüdiger J. Seitz[§], and Peter Hagoort[¶]

^{*}Max Planck Institute for Psycholinguistics, NL-6500 AH Nijmegen, The Netherlands; [†]Institute of Medicine, Research Center, Jülich D-52428, Germany; [‡]Department of Neurology, University of Düsseldorf, Düsseldorf D-40225, Germany; and [§]F. C. Donders Centre for Cognitive Neuroimaging, NL-6525 EK Nijmegen, The Netherlands

Communicated by Willem J. M. Levelt, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, March 9, 2001 (received for review December 20, 2000)

Spoken language is one of the most compact and structured ways to convey information. The linguistic ability to structure individual words into larger sentence units permits speakers to express a nearly unlimited range of meanings. This ability is rooted in speakers' knowledge of syntax and in the corresponding process of syntactic encoding. Syntactic encoding is highly automatized, operates largely outside of conscious awareness, and overlaps closely in time with several other processes of language production. With the use of positron emission tomography we investigated the cortical activations during spoken language production that are related to the syntactic encoding process. In the paradigm of restrictive scene description, utterances varying in complexity of syntactic encoding were elicited. Results provided evidence that the left Rolandic operculum, caudally adjacent to Broca's area, is involved in both sentence-level and local (phrase-level) syntactic encoding during speaking.

An average speaker, when asked to describe, for example, a scene of a little girl drawing a round geometrical shape, is able to start an utterance like "a child is drawing a circle" after about 1 s, completing it within another 2–3 s. According to current models of language production, in this brief time the speaker has passed through a number of processing stages (1–6) (see Fig. 1). "Conceptual preparation" involves, among many others, the decision to make a statement about the child and not the circle (as in the passive sentence "a circle is drawn") and to discard the child's gender as irrelevant in the present discourse. Also at this stage the speaker maps the visual concept of a round shape to the lexical concept "circle." Then, in the "syntactic encoding" stage, the preverbal message is linguistically encoded by retrieving the corresponding words ("lemmas") from the mental lexicon and arranging them in a grammatical order. This process uses the stored syntactic information of words, such as word class and grammatical gender, to compute a syntactic structure that specifies the relations of words in a sentence and determines their order and inflectional markings. This computation is done in a highly automatic and efficient manner. Speakers never produce utterances like "drawing a circle child a are." During later processing stages the stored information on the sounds of words is retrieved. These "phonological codes" undergo further transformations that finally produce a code that can be executed by the articulatory system.

Because of the difficulty of controlling the conceptual processing of longer utterances, previous neuroimaging work on language production has concentrated on single words (7). Consequently, almost nothing is known about the cerebral substrates of sentence-level production processes. In this positron emission tomography (PET) study, we investigated the cortical activation induced by syntactic encoding during speaking. In the paradigm of restrictive scene description we elicited naturally produced responses with different degrees of syntactic encoding but constant and limited conceptual processing demands.

Materials and Methods

Tasks. Restrictive scene description involved asking subjects to view animated scenes and describe them in three different prespecified ways: (i) in a full sentence, (ii) with a sequence of noun phrases that had local syntactic structure but no sentence-level syntactic structure, or (iii) with a sequence of single words having no syntactic relationship. The noun phrase condition was included to assess whether any cerebral activations observed in the full-sentence condition should be attributed to sentence-level syntactic processing only or also to local syntactic processing of the noun phrase. Fig. 2 shows examples of three frames of one animated scene and of the different descriptions in German that were required of the subjects in different blocks. To minimize conceptual and naming ambiguities, the animated scenes did not involve people performing actions, but a fixed set of colored two-dimensional geometric objects. These objects could perform two specific actions upon one another. These were to "go next to" another object, or "launch" another object; "launch" meant to set another object in motion by impact. The objects were a circle (der Kreis—masculine gender), an ellipse (die Ellipse—feminine gender), and a square (das Viereck—neuter gender). The colors were red, blue, and green. Color assignment to objects varied randomly. There were always two objects that could only be distinguished by their color, to make naming of color plus shape the most natural description. The actions were performed by one or two of the objects. Subjects were instructed to name all participants in an action, their respective colors, and the action itself. In all response conditions, the order in which the objects were to be named depended on their role in the action (i.e., whether they themselves acted or were acted upon). This rule ensured equal conceptual processing of the scenes across conditions.

Participants were trained in the task 1 week before PET measurement. Training began by introducing the objects and the actions. After being instructed on how to describe the scenes in the different response conditions, subjects practiced each response condition in two blocks of 24 scenes, using the same stimuli as during PET measurement, but in a different order.

Experimental Procedures. For each trial, an animated scene was presented for 1,660 ms in the center of a Digital VT340 monitor screen, subtending a visual angle of 8° both vertically and horizontally. The resulting configuration of the geometrical objects remained on the screen during the response utterance.

Stimulus presentation began ≈60 s before PET scanning and

Abbreviations: PET, positron emission tomography; rCBF, regional cerebral blood flow; BA, Brodmann area.

[†]To whom reprint requests should be addressed at: Max Planck Institute for Psycholinguistics, Postbus 310, NL-6500 AH Nijmegen, The Netherlands. E-mail: indefrey@mpi.nl.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

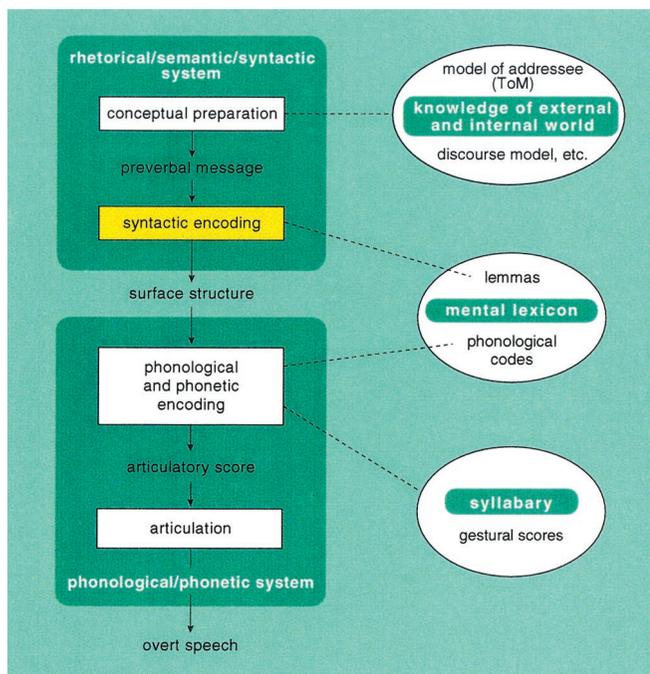


Fig. 1. Processing stages in speech production (adapted from ref. 6).

lasted for 3 min. During this time on average 21 scenes were presented. We applied two different presentation rates (eight scenes per min and six scenes per min) to control for the nonsyntactic (lexical, phonological, phonetic, and articulatory) processing load of the additional grammatical markers that subjects were required to produce in the sentence and noun phrase conditions. The increase in the overall language production rate (number of syllables per scanning period) that was induced by the fast presentation rate compared with the slow presentation rate was the same as that in the sentence condition compared with the single word condition. Therefore, possible hemodynamic effects due to differences in the overall language production rate between the sentence condition and the single word condition could be assessed by comparing the faster presentation rate with the slower presentation rate.

Twelve PET scans per subject were performed. The response conditions were in the order ABCCBAABCCBA; the assignment of sentence, noun phrase, and single word conditions to the positions A, B, and C was balanced across subjects. The visual stimuli were presented in a fixed order that was reversed for half of the subjects. The presentation rate changed every three scans; half of the subjects started with the slower rate, the other half with the faster rate.

Subjects. All participants (six females and six males) were consistent right-handers according to their scores on two handedness tests (8, 9). They were in the age range of 23 to 38 years, with a mean age of 26.8 years. All were native speakers of German, were in good health, and gave written informed consent in accordance with the Declaration of Helsinki. The study was approved by the Ethics Committee of the Heinrich Heine University Düsseldorf.

PET Data Acquisition and Analysis. PET data were recorded with an EXACT HR+ PET camera (Computer Technologies, Knoxville, TN). Scanning started at the time of i.v. injection of the tracer into the right brachial vein. Reconstructed activity images were created for a period of 40 s, starting with tracer arrival in the brain. For each scan, ≈ 550 MBq [^{15}O]butanol was injected as a

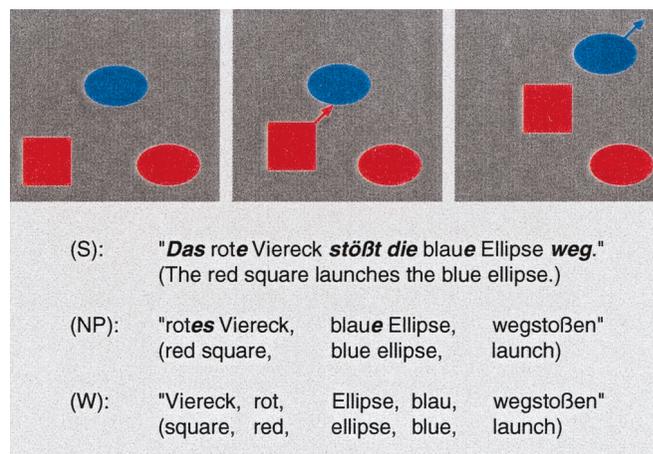


Fig. 2. Example of an animated stimulus scene. In this scene the red square launches the blue ellipse. Arrows are added to indicate the movement direction of the objects on the computer screen. Stimuli of the same kind were used in all three conditions. Examples of the three response types are given below (S, sentence condition; NP, noun phrase condition; W, single word condition). The response types differed in the degree of syntactic encoding and the corresponding application of grammatical markers (printed in bold) in German. Local gender agreement marking on the adjective was required in noun phrase and sentence responses but not in the single-word responses. Only in the sentence condition did syntactic relations across several words have to be expressed by means of word order and inflection of the main verb.

bolus. A combined dynamic-autoradiographic approach delivered image volumes of quantitative regional cerebral blood flow (rCBF) (10).

For data analysis we used the statistical parametric mapping (SPM96) software provided by the Wellcome Department of Cognitive Neurology, London (11). The image volumes were realigned, normalized into standard stereotactic space (using the template of the Montreal Neurological Institute provided by SPM96), smoothed with a 10-mm (full width at half maximum) Gaussian filter, and corrected for residual within- and between-subject global cerebral blood flow variation by analysis of covariance. For statistical comparisons of activation-control contrasts we chose a strict Bonferroni-corrected threshold of $P < 0.05$, which is considered necessary when there is no *a priori* hypothesis for possible activation locations. Hypotheses about rCBF differences within a region of interest were tested by ANOVA and post hoc *t* tests for paired samples at a threshold of $P < 0.05$, corrected for number of comparisons.

Anatomical Localization Procedure. In the context of this paper Brodmann area (BA) 44 is defined as the overlap of BA 44 of at least five of 10 postmortem brains as described by Amunts *et al.* (12). In brief, this procedure involved the cytoarchitectonic mapping of BA 44 in 10 individual brains by means of an observer-independent technique. Three-dimensional reconstructions based on high-resolution MR scans of the 10 brains were anatomically standardized to the reference brain of the European Computerized Human Brain Database (13) by means of linear and nonlinear transformations (14). Finally, the number of brains that agreed in the assignment of BA 44 was determined for every voxel. For the projection of functional data onto the same reference brain, we mapped the brain template of the Montreal Neurological Institute to the template of the European Computerized Human Brain Database by means of the anatomical standardization procedure of SPM96 and applied the resulting transformation parameters to the statistical parametric maps. Visual inspection of the resulting positions of the functional

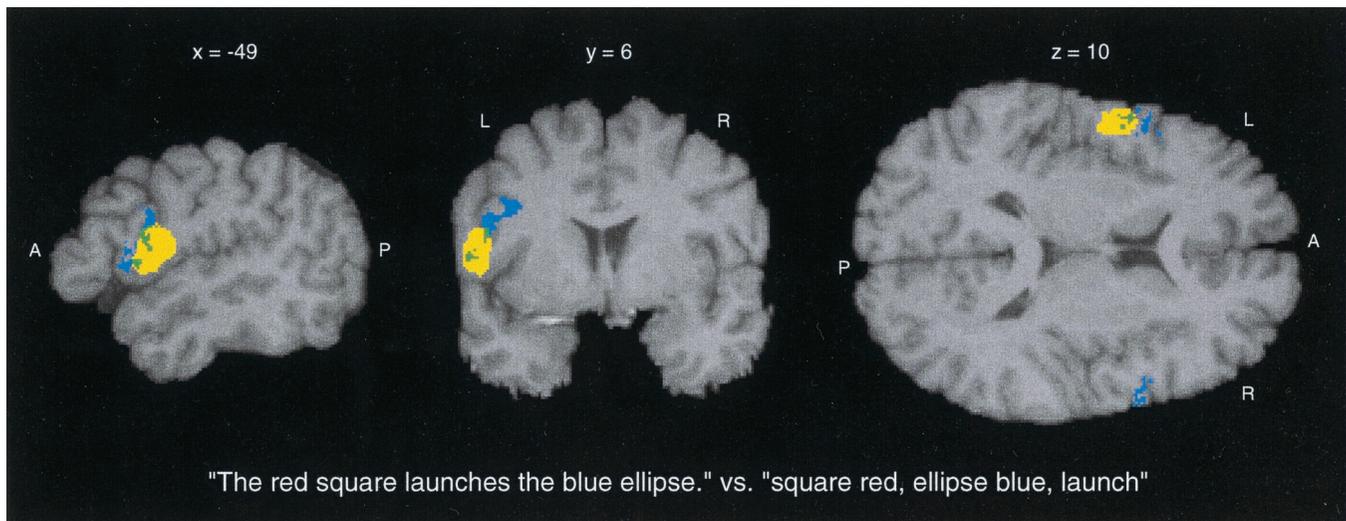


Fig. 3. Cortical activation of sentence relative to single word utterances. Significantly activated voxels are projected in yellow onto anatomical MR sections of a reference brain. For anatomical comparison, voxels belonging to BA 44 are projected in blue on the same reference brain. A smaller anterior portion of the activated volume (16.3%, shown in green) overlaps BA 44, and the larger part of the activation (83.7%) lies caudally adjacent to BA 44, most probably corresponding to the Rolandic part of BA 6. The maximally activated voxel was located at $x = -54$, $y = 6$, $z = 10$ (coordinates as given by SPM96). (Note that the depicted sagittal section is taken more medially to improve the visibility of the anatomical configurations of the posterior inferior frontal gyrus).

activations relative to anatomical landmarks found them unchanged.

Results

Behavioral Data. Response utterances were recorded on digital audiotape and analyzed for voice onset time (measured from the time at which the final configuration of geometrical shapes was reached) and response duration with the XWAVES speech-processing package. Voice onset times were 1,218 ms (SD 306 ms) for sentences, 1,286 ms (SD 307 ms) for noun phrases, and 1,227 ms (SD 299 ms) for single words. Response durations were 3,073 ms (SD 859 ms) for sentences, 3,073 ms (SD 855 ms) for noun phrases, and 3,074 ms (SD 860 ms) for single words. A 3×2 ANOVA with the within-subject factors condition and response variable (with the levels voice onset time and response duration) showed no significant main effect of condition ($P > 0.1$) and no significant interaction of condition by response variable ($P > 0.1$).

rCBF Data. Comparing the rCBF data of the two conditions that differed maximally in terms of syntactic encoding—i.e., sentences and isolated words—we found a single highly significant ($Z = 4.79$, $P = 0.019$ corrected) activation focus in the left anterior Rolandic operculum (15), caudally adjacent to BA 44 or Broca's area (see Fig. 3). There was no significant activation in the reverse comparison.

To determine whether the activated area was sensitive to syntactic processing only or also responded to changes in the overall language production rate, we calculated a 3×2 ANOVA with the within-subject factors response condition and presentation rate on the mean rCBF data in this region of interest (222 voxels). There was a significant main effect ($F = 16.739$, $df = 2$, $P = 0.002$) of the factor condition. There was no significant main effect of the factor presentation rate ($P > 0.1$) and no significant interaction of the two factors ($P > 0.1$). A post hoc t test comparing the mean rCBF in the single word condition at faster versus slower stimulus presentation rates showed no significant difference ($t = 1.328$, $df = 11$, $P > 0.1$, one-tailed).

To determine whether the activated area was sensitive to sentence-level syntactic processing only or also responded to local syntactic processing on the noun phrase level, we finally

compared the mean rCBF in this region of interest across all three conditions (see Fig. 4). There was a graded response, with sentences activating this region more strongly than noun phrases ($t = 3.148$, $df = 11$, $P = 0.005$, one-tailed) and noun phrases activating this region more strongly than isolated words ($t = 4.195$, $df = 11$, $P = 0.000$, one-tailed).

Discussion

Our results demonstrate a neural correlate of syntactic encoding during speaking. The overt production of sentences compared

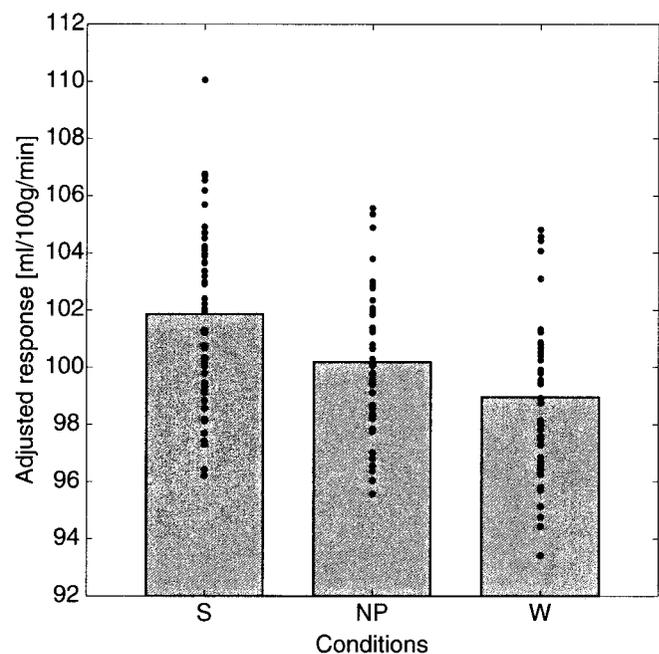


Fig. 4. Mean rCBF in the activated volume across conditions. Means are calculated for a region of interest comprising all 222 voxels (1,776 mm³) that were significantly activated for sentence relative to single word utterances at a single voxel threshold of $P < 0.001$ (the global cerebral blood flow was adjusted to 50 ml/100 g/min; S, sentence condition; NP, noun phrase condition; W, single word condition).

with the overt production of sequences of words that had no syntactic relationship induced a significant increase in the cerebral blood flow in the left anterior Rolandic operculum. More importantly, this syntactic activation focus showed a graded response depending on the complexity of syntactic processing. The syntactic encoding of noun phrases alone activated the same location, but to a weaker extent, as the syntactic encoding of full sentences.

The paradigm we used isolated syntactic encoding from other processing components. It involved the visual presentation of identical animated scenes in all response conditions. Correct responses required the same conceptual processing of the animated scenes in all conditions. This was confirmed by the reaction time data, showing no significant differences in the time subjects needed to initiate their spoken responses, indicating approximately equal cognitive demands of the three task variations. The identical durations of the response utterances in all three tasks, furthermore, suggest that the different intonation contours required by the three response types did not alter the amount of prosodic planning due to overall utterance length.

What might be considered as a potentially confounding factor was a slight increase in the number of syllables and words to be uttered in the syntactic conditions. This increase was due to the grammatical markers and function words that were necessary to express the syntactic relations of words within sentences and noun phrases and therefore could not be avoided. However, we have shown that a comparable increase in the number of produced syllables and words that was experimentally induced by a slightly faster stimulus presentation rate did not have any significant hemodynamic effect in the observed syntactic activation area. It is therefore highly improbable that the nonsyntactic processing load of the grammatical markers and function words contributed to the observed rCBF increases in the sentence and noun phrase conditions.

In how far can our findings be generalized to naturally produced spoken language? Although the sentences elicited in the experiment were more constrained than sentences speakers typically produce in naturally occurring contexts, these constraints were mainly operative at the conceptual level, reducing

the number of possible concepts. The paradigm thus preserved the essential properties of normal, conceptually driven language production. From one trial to the next, subjects accurately described scenes varying in the number of actors and the transitivity of the verb. This variation required subjects to use several different syntactic structures according to the demands of the scene. Because the scenes were quasirandomly ordered, it is unlikely that subjects could have used nonlinguistic response strategies in syntactic encoding.

Given our results, one should expect lesions of the left opercular cortex outside Broca's area to cause disturbances of the syntax of speech, a symptom called agrammatism. There are indeed clinical data suggesting that long-lasting impairment of syntactic encoding does not occur with isolated lesions of Broca's area but requires an involvement of adjacent opercular areas (16). On the other hand, agrammatism has been reported for lesions of a much more widespread set of areas in the left perisylvian cortex (17–19). The reason for this discrepancy between clinical and neuroimaging data, which has also been observed for syntactic comprehension, is still unclear. Possible reasons include a heterogeneity of processing impairments covered by the term “agrammatism” and individual variations in the location of neural tissue supporting specific processing components.

Activations of the opercular part of the left inferior frontal gyrus, next to the Rolandic operculum, have been found in hemodynamic studies of syntactic comprehension (20–24). Although this observation might be taken as support for the notion of a common syntactic processor, other locations reported for syntactic comprehension are several centimeters away (25–27). Investigation of encoding and parsing of identical syntactic structures in the same subjects is needed to come to firm conclusions on this question.

We have demonstrated a cortical region supporting one specific component, syntactic encoding, in the fast cascade of processes resulting in overt speech. This cortical region, which was identified as the anterior Rolandic operculum, is used to structure individual words into phrases and sentences expressing complex thoughts.

- Levelt, W. J. M. (1989) *Speaking: From Intention to Articulation* (MIT Press, Cambridge, MA).
- Garrett, M. (1975) in *Psychology of Learning and Motivation*, ed. Bower, G. (Academic, New York), pp. 133–177.
- Stemberger, J. P. (1985) in *Progress in the Psychology of Language*, ed. Ellis, A. W. (Erlbaum, Hillsdale, NJ), Vol. 1, pp. 143–186.
- Dell, G. S. (1986) *Psychol. Rev.* **93**, 283–321.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M. & Gagnon, D. A. (1997) *Psychol. Rev.* **104**, 801–837.
- Levelt, W. J. M. (1999) in *Neurocognition of Language*, eds. Brown, C. & Hagoort, P. (Oxford Univ. Press, New York), pp. 83–122.
- Indefrey, P. & Levelt, W. J. M. (2000) in *The New Cognitive Neurosciences*, ed. Gazzaniga, M. (MIT Press, Cambridge, MA), 2nd ed., pp. 845–865.
- Oldfield, R. C. (1971) *Neuropsychologia* **9**, 97–113.
- Steingrüber, H. J. (1971) *Z. Exp. Angew. Psychol.* **18**, 337–357.
- Herzog, H., Seitz, R. J., Tellmann, L., Schlaug, G. & Müller-Gärtner, H.-W. (1996) *J. Cereb. Blood Flow Metab.* **16**, 645–649.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D. & Frackowiak, R. S. J. (1995) *Hum. Brain Mapp.* **2**, 189–210.
- Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. & Zilles, K. (1999) *J. Comp. Neurol.* **412**, 319–341.
- Roland, P. E. & Zilles, K. (1996) *NeuroImage* **4**, S39–S47.
- Schormann, T. & Zilles, K. (1998) *Hum. Brain Mapp.* **6**, 339–347.
- von Economo, C. (1930) *Z. Ges. Neur. Psychiatr.* **130**, 774–781.
- Mohr, J., Pessin, M., Finkelstein, S., Funkenstein, H., Duncan, G. & Davis K. (1978) *Neurology* **28**, 311–324.
- Basso, A., Lecours, A. R., Moraschini, S. & Vanier, M. (1985) *Brain Lang.* **26**, 201–229.
- Vanier, M. & Caplan, D. (1990) in *A Cross-Language Narrative Source Book*, eds. Menn, L. & Obler, L. K. (Benjamins, Amsterdam), pp. 37–114.
- Willmes, K. & Poeck, K. (1993) *Brain* **116**, 1527–1540.
- Stromswold, K., Caplan, D., Alpert, N. & Rauch, S. (1996) *Brain Lang.* **52**, 452–473.
- Dapretto, M. & Bookheimer, S. Y. (1999) *Neuron* **24**, 427–432.
- Kang, A. M., Constable, R. T., Gore, J. C. & Avrutin S. (1999) *NeuroImage* **10**, 555–561.
- Embick, D., Marantz, A., Miyashita, Y., O'Neil, W. & Sakai, K. L. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 6150–6154. (First published May 16, 2000; 10.1073/100098897)
- Friederici, A. D., Meyer, M. & von Cramon, D. Y. (2000) *Brain Lang.* **74**, 289–300.
- Caplan, D., Alpert, N. & Waters, G. (1998) *J. Cognit. Neurosci.* **10**, 541–552.
- Caplan, D., Alpert, N. & Waters, G. (1999) *NeuroImage* **9**, 343–354.
- Caplan, D., Alpert, N., Waters, G. & Olivieri, A. (2000) *Hum. Brain Mapp.* **9**, 65–71.