OSU Working Papers in Linguistics 45 (1-10)

Interaction with autonomy: Defining multiple output models in psycholinguistic theory*

Julie E. Boland and Anne Cutler
jboland@ling.ohio-state.edu
anne@mpi.nl

Abstract: There are currently a number of psycholinguistic models in which processing at a particular level of representation is characterized by the generation of multiple outputs, with resolution involving the use of information from higher levels of processing. Surprisingly, models with this architecture have been characterized as autonomous within the domain of word recognition and as interactive within the domain of sentence processing. We suggest that the apparent internal confusion is not, as might be assumed, due to fundamental differences between lexical and syntactic processing. Rather, we believe that the labels in each domain were chosen in order to obtain maximal contrast between a new model and the model or models that were currently dominating the field.

Models of psycholinguistic processing typically consist of a number of levels loosely corresponding to levels of linguistic analysis. Even where a model deals only with the operations of one level - e.g. word recognition or parsing - some assumptions about its relationship to the other levels will usually be spelled out. In part, this is because models virtually always take a stand on one side or the other of the Great Divide in psycholinguistic theorizing - interaction versus autonomy.

Consider models of syntactic processing. One of the defining issues is whether syntactic choices are made with the benefit of relevant semantic knowledge. For example, both sentence fragments in (1) are syntactically ambiguous between a main clause structure (... *the book.*) and a reduced relative structure (... *by the lawyer was informative.*). However, (1b) is not semantically ambiguous: it is much more plausible that the evidence is being examined than that the evidence is examining something. Thus the main clause structure ought to be blocked for (1b) if semantic information can be used to decide between syntactic alternatives. On the other hand, the main clause structure is simpler, so it might be preferred if only syntactic information could be considered. Widely cited work by

*This paper was written while J. E. Boland was a Visiting Fellow at the Max Planck Institute for Psycholinguistics. We thank Dennis Norris, Michael Tanenhaus, and Lee Osterhout for their insights and comments on an earlier draft of this manuscript. The ordering of the authors is strictly alphabetical. Julie Boland, 137 Townshend Hall, 1885 Neil Ave., Columbus, OH 43210. Anne Cutler, Max Planck Institute for Psycholinguistics, Post Box 310, NL-6500 AH Nijmegen.

Ferreira and Clifton (1986) suggested that syntactic decisions such as these were based solely upon structural simplicity, supporting Frazier's (1978) autonomous model of syntactic processing. According to Frazier's model, the parser always constructs the simplest structure allowed by the phrase structure rules of the grammar. This initial parse uses only the major syntactic category (noun, verb, etc.) of the input, and is later checked against detailed lexical and semantic information.

(1) a. The defendant examined ...

b. The evidence examined ...

However, recent work suggests that semantic influences can affect syntactic choices (e.g. Altmann, Garnham, & Dennis, 1992; Britt, 1994; Boland, Tanenhaus, Garnsey, & Carlson, 1994; Pearlmutter & MacDonald, 1992; Trueswell, Tanenhaus, & Garnsey, 1994). For example, Trueswell et al. found that although processing difficulty arose when sentences like (1a) were completed with a reduced relative structure, no such processing difficulty arose when sentences like (1b) were completed with a reduced relative structure. They argue that, contrary to Frazier's (1978, 1987) claims, detailed lexical information is used to constrain the syntactic alternatives, and semantic information is used to select among them. Similar arguments are put forth in Boland, Tanenhaus, Garnsey, and Carlson (1994), based on their work on *wh*-questions.

As might be expected, proponents of the lexicalist constraint-based approach have adopted a position on the question of interaction versus autonomy. They describe their approach as an interactive system, opting for an interactive architecture on the grounds that multiple constraints, some of them non-syntactic, govern the selection of the initial syntactic structure (e.g. MacDonald, Pearlmutter, & Seidenberg, in press; Tanenhaus & Trueswell, in press).

The lexicalist constraint-based view is in some ways similar to the incremental interactive theory first proposed in Crain and Steedman (1985) and further refined in Altmann and Steedman (1988). In this model, syntactic alternatives are constructed in parallel within the constraints of lexical specifications, and a single representation is selected by the semantic system, using principles of referential support, a priori plausibility, etc. Thus there is a bottom-up generation of alternatives, with selection of a single structure left for a later stage of processing.

As the label given to the model makes plain, Altmann and Steedman (1988) considered their model to be interactive, noting that their results "support the interactive hypothesis" (p. 192). However, they explicitly described it as only weakly interactive. "According to this [weak] version [of the interactive hypothesis], syntax autonomously proposes analyses, while semantics and context merely dispose among the alternatives offered." (p. 205) They contrasted their position with strongly interactive models, which generate only the most plausible

structure(s), and with Frazier's (1978) autonomous model, which generates only the simplest structure.

The incremental interactive model clearly separates generation processes from selection processes. This distinction is not unique to the parsing literature, however; it is also a feature of many word recognition models. Models of visual word recognition began to adopt this approach in the 1970s. Becker (1976) proposed a "verification model" in which a rough physical analysis of the input extracts sensory features and compiles a set of candidate words having those features, and therefore compatible at least in part with the incoming stimulus; these words then are ranked in order of frequency and compared one by one against a stored sensory representation of the input; this comparison process is termed verification because it actually consists of the generation of predictions from the lexical representations of the candidate words and the verification (or otherwise) of these predictions against the input store. The "checking model" put forward by Norris (1986) likewise generates an initial candidate set on the basis of partially analyzed perceptual information. The set is continually updated as the perceptual analysis is refined, but in the meantime the candidates in the set are also checked for compatibility with the sentential or other semantic context constructed so far in the recognition process. There is in this case no intrinsic ordering within the candidate set; word frequency, contextual compatibility and perceptual information all operate in the same way, to increment individual candidate words' weightings and thus eventually to determine which candidate word first reaches a specified recognition criterion.

Models of spoken word recognition, too, may split the recognition process into separate stages. Norris' (1994) SHORTLIST model, as its name suggests, is one such; in this model the initial stage again generates multiple candidates compatible with the input, while in the second stage a process of competition (involving, again, adjustment of weightings for each candidate word) determines which of the shortlisted candidates eventually wins through to recognition. A principled difference between the visual and spoken word recognition situations is that in the former case it is reasonable to consider a rough initial analysis of the entire word as input to the candidate set generation, so that such features as word length may play a role at this stage; in the latter case the temporal dimension within which the input arrives means that the initial candidate set will be primarily determined by the initial portions of the stimulus, and word length cannot be initially apparent.

The cohort model of Marslen-Wilson and Welsh (1978), especially in its revised form (Marslen-Wilson, 1987), also allows for an initial stage in which only the perceptual input determines a subset of lexical entries. Selection among this set of activated candidates is then carried out by a later stage, which in the earlier version of the model was sensitive both to further accumulation of perceptual evidence and to contextual (syntactic and semantic) information, but in the later version operates on perceptual information alone, in parallel with a contextual integration stage.

As Marslen-Wilson (1987) points out, the concept of multiple output distinguishes such models from, for example, direct access models such as the logogen model (Morton, 1970) in which only one lexical entry will surmount a recognition threshold and be effectively accessed. Norris (1986) argued that incorporating multiple output makes word recognition models in effect more parsimonious, since multiple access is in any case required to deal with the phenomenon of lexical ambiguity. It is generally agreed that at least under certain conditions presentation of an ambiguous word will lead to momentary availability of its multiple senses (Conrad, 1974; Swinney, 1979; Tanenhaus, Leiman & Seidenberg, 1979); that is, the perceptual input alone does not suffice to identify a unique word candidate, and final selection must be made by comparison against the context (which, since the evidence suggests momentary availability of multiple meanings, must occur at a post-access stage). If the mechanism for simultaneous access of multiple candidates must exist in any case, to explain selection where the input *cannot* unambiguously determine the output of the recognition process, then architectural economy is best served by exploiting precisely that mechanism in all recognition processes, for unambiguous as well as for ambiguous words.

There is widespread agreement in the word recognition literature of recent years that Multiple Output approaches fundamentally embody autonomy of the lexical access process. Becker's (1976) model, to be sure, included a separate process of generation of a semantically appropriate candidate set of words (to be compared via the same verification process with the stored input), and hence allowed for semantic context to drive a lexical access process. But since then none of the Multiple Output models have allowed higher-level (syntactic or semantic) processes such freedom. Instead, autonomy is deemed to be preserved in such models in that the actual process of contacting a lexical entry is responsive solely to bottom-up perceptual information, and is not affected in any way by higher-level processing. Thus the checking model has "a completely bottom-up flow of information" and the "stages are completely autonomous" (Norris, 1986: 131); the revised cohort model constitutes "a fully bottom-up model where context plays no role in ... access and selection" (Marslen-Wilson, 1987: 71) because "both access and certain aspects of selection are autonomous processes, in the sense that they are driven strictly from the bottom-up" (Marslen-Wilson, 1987: 98); Shortlist is "a bottom-up autonomous model" (Norris, 1994: 231) in which all "top-down feedback ... is redundant" (Norris, 1994: 191).

Thus the psycholinguistic literature may appear, to a newcomer, to be prey to internal confusion. Both for parsing and for word recognition, Multiple Output models have been proposed which have basically the same architecture; but the position they adopt on the interaction/autonomy issue is not as similar as their structural similarities would seem to demand. In fact models in the two domains take up fundamentally incompatible positions: one is called interactive, the other autonomous. Researchers in each domain clearly agree on the criteria by which these labels are applied, and thus, within each area, there is no confusion; but we

believe that the reasons for the asymmetry provide an interesting subject of scrutiny.

The description that best fits the Multiple Output architecture depends on what one considers to be the defining features of autonomy and interaction, and different definitions have been established in word recognition and parsing. In the parsing literature, use of higher-level information to resolve lower-level decisions constitutes interaction. Given this definition, Multiple Output models are clearly interactive because higher-level information is used in the selection process. On the other hand, one might not consider a process interactive unless higher-level information actually affects the way that alternatives are generated within the system, ruling out certain candidates *a priori*, irrespective of their compatibility with bottom-up information. Autonomy would imply that processing operations at a given level proceed in the same way irrespective of whatever counsel might be deducible from higher-level considerations. It is this type of autonomy that has characterized the debate within the domain of word recognition. This is also the definition that Fodor (1983) used in his argument for modularity in mental processing: "a system [is] autonomous by being encapsulated, by not having access to facts that other systems know about" (p. 73). In these terms, Multiple Output models are clearly autonomous.

Altmann and Steedman (1988) indeed pointed out that the architecture of their model "does not compromise the modularity hypothesis of Fodor (1983) in any way" (p. 192). In fact, in terms of this definition of autonomy, Multiple Output models of parsing are more autonomous than Frazier's (1978, 1987) model, which is autonomous only with regard to initial syntactic analysis. When the initial analysis is inconsistent with thematic information, syntactic reanalysis occurs within, or is guided by, a thematic processor (Rayner, Carlson, & Frazier, 1983; Ferreira & Henderson, 1991). Note that it is not enough for the thematic processor simply to send an error signal to restart the syntactic processor, because the syntactic processor would automatically construct the simplest structure once again. Models such as the incremental interactive model (Altmann & Steedman, 1988) or the concurrent model proposed by Boland (1993), which produce parallel outputs, do not have this limitation. If necessary, the syntactic processor would reproduce the parallel outputs exactly as it had the first time, and the external selection processes would make the correct selection, guided by the knowledge of the previous mistake. Thus the parallel parser generates structures completely autonomously during reanalysis as well as during initial analysis.

The fact that parsing models and word recognition models have maintained different definitions of autonomy provides only a superficial explanation for the inconsistent labeling of Multiple Output models. The question then becomes: why has the parsing literature used one definition and the word recognition literature another? One possible reason is that there exist fundamental differences between lexical and syntactic processing, which justify adopting different definitions of autonomy. Traditionally, word recognition has been viewed as a lookup process, i.e. the access of stored lexical representations. Parsing, on the other hand, has

been viewed as a construction process, whereby representations are computed rather than being chosen from a store. Correspondingly, outputs of lexical processing have been assumed to coincide with the completion of the processing stage (i.e. recognition of the word), but outputs of syntactic processing have been taken to correspond to many incremental stages in the construction of a complete syntactic structure.

However, we believe that current models of both parsing and word recognition make the maintenance of such rigid distinctions no longer tenable. For instance, it is clear that processes which essentially involve simple lookup can do much of the work in parsing traditionally believed to require construction processes. There is abundant evidence that syntactic decisions make use of detailed lexical information that is accessed as part of word recognition. This research has focused primarily on verb-based information, such as subcategorization frames (e.g. McElree, 1993; Osterhout, Holcomb, & Swinney, 1994), verb control information (Boland, Tanenhaus, & Garnsey, 1990), and thematic roles (e.g. Britt, 1994; Mauner, Tanenhaus, & Carlson, in press; Stowe, 1989; Taraban & McClelland, 1988). Use of stored lexical information means that syntactic processing is more dependent upon access processes and less dependent upon construction processes than has often been assumed. MacDonald, Pearlmutter, & Seidenberg (in press) have taken the lexicalist approach to sentence processing even further, suggesting that the lexical entries of nouns, verbs, and words of other categories contain X-bar structures. The only construction that takes places in their model is the connecting of one X-bar structure to another.

On the other hand, models of word recognition - and, in particular, Multiple Output models - do not necessarily consist solely of lookup procedures. For instance, Norris' "checking model" (1986) of visual word recognition contains much more of a continuous element, in that the initial stage is continually outputting updated analyses to the checking stage. Likewise, the SHORTLIST model of spoken word recognition (Norris, 1994) provides for a continuous input from the initial generation stage to the competition/selection stage. In fact this continuous updating feature turns out to be an essential feature of SHORTLIST. In order to account for empirical data indicating that human listeners employ prelexical segmentation routines *in conjunction with* competition processes (McQueen, Norris & Cutler, 1994), the SHORTLIST model has been modified to include a prelexical segmentation procedure mimicking Cutler and Norris' (1988) Metrical Segmentation Strategy (Norris, McQueen & Cutler, in press). To achieve this, it proved essential that the updated output of the initial stage continually *replace* the previous output. Only with this replacement mechanism did the model exactly simulate the human empirical data. The continuous output feature of such models renders the notion of a simple lookup procedure, with its completion amounting to completion of the lexical stage of processing, inaccurate as a description of the word recognition process.

We do not mean to imply that there is agreement that lexical and syntactic processes are fundamentally alike; these issues remain the subject of hot debate,

and the traditionally held differences may in fact have influenced the adoption of different definitions of autonomy in the two domains. But whatever the outcome of the debate, there is no longer any logical force behind the argument that lexical and syntactic processing are so different that an identical architecture motivates opposite theoretical descriptions in the two domains.

Instead, we believe that considerations outside the architecture of processing models have influenced how Multiple Output models have come to be labeled. Multiple Output models, both of word recognition and of parsing, were introduced after other models had already, in effect, defined the territory. In each case, the Multiple Output model posed a challenge to the existing model, and was correspondingly assigned an opposing label.

The dominant model in syntactic processing in the 1980s, when the syntactic models discussed above were first mooted, was undoubtedly Frazier's (1978) model. Moreover, Frazier's model was particularly known for its position in the dominant theoretical debate in psycholinguistics, in that it was declared to be strictly autonomous (although, as we pointed out above, it is at least reasonable to claim that her model is not in fact maximally autonomous). Because Frazier's model was labeled as autonomous, the opposing models - which were indeed very different in structure - came to be termed interactive. Quite reasonably, the proponents of Multiple Output models wished to promote their approach as a genuine theoretical alternative to the currently dominant approach. The most obvious way to do this was to adopt a contrasting position within the dominant theoretical debate.

In the word recognition literature, too, we believe that labels were influenced by considerations of contrast. Just as autonomy could be said to be making the running in syntactic modeling, and hence be the position with which contrast could most easily be drawn, so were there models in word recognition which were market leader in much the same way, and these models were interactive. In visual word recognition, the dominant model prior to the emergence of Multiple Output models was Morton's (1970) logogen model, in which higher-level information from the context contributed directly to the activation of lexical candidates just as bottom-up information from incoming input did. In spoken-word recognition, the logogen model was also a contender, but the first model specifically devoted to the auditory case, the cohort model of Marslen-Wilson and Welsh (1978) was likewise interactive in that syntactic and semantic context was deemed capable of controlling the availability of potential candidate words. And finally, TRACE (McClelland & Elman, 1986), the most influential model of spoken-word recognition since the mid-1980s, again embodies interactive use of higher-level information in the word recognition process. Thus adoption of an autonomous stance again allowed proponents of Multiple Output models to achieve maximal contrast with the currently dominant models.

The interaction/autonomy debate has functioned as an effective energizer for psycholinguistics in the last few decades; it may have stimulated more research

than any other single issue. Placing one's contribution within this paradigm is *de rigeur*, but, as we have argued, the placement may not always be rigorously determined by architectural issues alone. Contrast with theoretical alternatives - that is, in effect, sociopolitical considerations - may play as large a role.

References

Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language, 31,* 685-712.

Altmann, G. T. M. & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30,* 191-238.

Becker, C. A. (1976). Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 556-566.

Boland, J. E. (1993). The role of verb argument structure in sentence processing: Distinguishing between syntactic and semantic effects. *Journal of Psycholinguistic Research, 22,* 135-152.

Boland, J. E., Tanenhaus, M. K., and Garnsey, S. M. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language, 29,* 413-432.

Boland, J. E., Tanenhaus, M. K., Garnsey, S. M., & Carlson, G. N. (1994). *Verb argument structure in parsing and interpretation: Evidence from wh-questions.* Manuscript submitted for publication.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure: The parsing of prepositional phrases. *Journal of Memory and Language, 33,* 251-283.

Conrad, C. (1974). Context effects in sentence comprehension: A study of the subjective lexicon. *Memory & Cognition, 2,* 130-138.

Crain, S. & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D. Dowty, L. Kartunnen, & A. Zwicky (Eds.) *Natural Language Parsing.* Cambridge: Cambridge University Press.

Cutler, A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 113-121.

Ferreira, F. & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25,* 348-368.

Ferreira, F. & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 30,* 725-745.

Fodor, J. A. (1983). *The Modularity of Mind.* Cambridge, MA: MIT Press.

Frazier, L. (1978). *On Comprehending Sentences: Syntactic parsing strategies.* Ph.D. Thesis, University of Connecticut. Distributed by the Indiana University Linguistics Club.

Frazier, L. (1987). Sentence processing. A tutorial review. In M. Coltheart (Ed.), *Attention and Performance XII: The psychology of reading.* London: Erlbaum.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (in press). The lexical nature of syntactic ambiguity resolution. *Psychological Review.*

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition,* 25, 71-102

Marslen-Wilson, W. D. & Welsh, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology,* 10, 29-63.

Mauner, G., Tanenhaus, M. K., & Carlson, G. N. (in press). Implicit arguments in sentence processing. *Journal of Memory & Language.*

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

McElree, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory & Language, 32,* 536 - 571.

McQueen, J. M., Norris, D. G., and Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 621-638.

Morton, J. (1970). A functional model for memory. In D. A. Norman (Ed.), *Models for Human Memory.* New York: Academic Press.

Norris, D. G. (1986). Word recognition: Context effects without priming. *Cognition, 22,* 93-136.

Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52,* 189-234.

Norris, D. G., McQueen, J. M., and Cutler, A. (in press). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 20, 786-803.

Pearlmutter, N. & MacDonald, M. C. (1992). Plausibility effects in syntactic ambiguity resolution. *Proceedings of the 14th annual meeting of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning & Verbal Behavior, 22,* 358-374.

Stowe, L. (1989). Thematic structures in sentence comprehension. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing.* Dordrecht: Kluwer.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645-660.

Tanenhaus, M. K., Leiman, J., & Seidenberg, M. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning & Verbal Behavior, 18*, 427-441.

Tanenhaus, M. K. & Trueswell, J. (in press). Sentence comprehension. In J. Miller and P. Eimas (Eds.), *Handbook of Perception and Cognition: Vol. 11.* Academic Press.

Taraban, R. & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing. *Journal of Memory & Language, 27*, 597-632.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285-318.