

Distributional Information in Language Comprehension, Production, and Acquisition: Three Puzzles and a Moral

Maryellen C. MacDonald
University of Southern California

One of the unwritten rules of psycholinguistics is that acquisition, comprehension, and production research each keeps to itself—the questions addressed in these three fields, and the researchers who ask them, overlap in only the most general ways. In acquisition, for example, researchers of necessity use comprehension and production measures in assessing children's progress, but the primary goal of much acquisition research is to understand how the child comes to acquire the grammar, or knowledge of the language, not how the child develops comprehension or production abilities. Production and comprehension research are similarly isolated; neither one digs deeply into the question of how the nature of these adult systems is constrained by the acquisition process, or whether production and comprehension processes exert significant constraints on each other. Research in each field has made a great deal of progress using this isolationist strategy, and there are clearly unique questions in each field for which the neighboring fields offer little insight. There do appear to be some important domains, however, where it appears that the isolationist approach is a distinct limitation. This is the theme of this chapter, which reviews three interrelated findings in comprehension, production, and acquisition research. In each case, the puzzling results in one field appear to have solutions in another. The intricate relationships between these puzzles hold important implications for the nature of the human language faculties and for the isolationist research strategies that currently dominate psycholinguistic research.

PUZZLE #1: A CURIOUS FINDING IN SYNTACTIC AMBIGUITY RESOLUTION

Syntactic processing, or *parsing*, is the subfield of language comprehension research that investigates how comprehenders uncover the syntax and meaning of sentences from the linear string of spoken or written input. Much of this research has used syntactic ambiguities, strings that temporarily have more than one syntactic interpretation, to study the parsing process. A great deal of recent parsing work shows the importance of lexical information in the ambiguity resolution process. This interest in lexical information is shared to some degree by every major approach to sentence processing but is represented most clearly within what is called the *constraint-based* approach, in which comprehension is achieved through the parallel satisfaction of multiple probabilistic constraints, including constraints from lexical representations. For example, MacDonald, Pearlmutter, and Seidenberg (1994) applied this approach to three major types of syntactic ambiguities in English and suggested that the initial interpretation of each of them is strongly guided by distributional information in the linguistic input concerning the relative frequencies of alternative lexical interpretations. An example of one of these ambiguities, the Main Verb/Reduced Relative ambiguity, is shown in (1):

1. a. *Temporary Main Verb/Reduced Relative Ambiguity*: The three men arrested . . .
- b. *Main Verb Interpretation*: The three men arrested the bombing suspects in a parking garage.
- c. *Reduced Relative Interpretation*: The three men arrested in the parking garage were wanted in connection with the bombing of the oil refinery.

In this case, interpretation of the ambiguity is constrained by the frequency with which the ambiguous verb (here, *arrested*) participates in transitive and passive structures, of which reduced relative clauses are a special type (MacDonald, 1994; MacDonald et al., 1994; Trueswell, 1996). Interpretation of this structure is also constrained by combinatorial lexical information, such as the plausibility of the initial noun phrase (NP) filling the agent or patient role of the verb (MacDonald, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Pearlmutter & MacDonald, 1992; Tabossi, Spivey-Knowlton, McRae, & Tanenhaus, 1994; Trueswell, Tanenhaus, & Garnsey, 1994). Similar simple and combinatorial lexical effects for this and other ambiguities have made a strong case for the importance of distributional information, particularly lexical information, in the earliest stages of syntactic processing, a key claim of constraint-based models of language comprehension.

One syntactic ambiguity stands out as a prominent exception to this pattern of lexical sensitivity. This ambiguity is a particular kind of modification ambiguity, in which a phrase can modify (or in syntactic structure terms, *attach to*) one of several elements earlier in the sentence. The particular modification ambiguity of importance here is one in which a prepositional or adverbial phrase can modify one of two verbs, as in (2). Example (2a) shows a fully ambiguous structure; (2b) shows an example in which verb tense disambiguates the sentence in favor of the *local modification* interpretation, in which the adverb *yesterday* modifies the nearby verb *left* rather than the more distant phrase *will say*; and (2c) shows a sentence with distant modification, in which *tomorrow* is not modifying the local verb, *left*, but is instead modifying the distant verb, *will say*.

2. a. *Verb Modification Ambiguity*: John said that Bill left yesterday.
- b. *Local Modification*: John will say that Bill left yesterday.
- c. *Distant Modification*: John will say that Bill left tomorrow.

English speakers have an extremely strong preference for local modification for this structure, as in (2b), and sentences like (2c) are typically perceived to be very difficult and awkward. Many researchers have assumed that this preference emerges from a general operating principle of the syntactic parsing mechanism to prefer the most local modification; such principles have been variously called Right Association (Kimball, 1973), Late Closure (Frazier, 1987), and Recency (Gibson, Pearlmutter, Consecogonzales, & Hickok, 1996). A striking feature of the verb modification ambiguity is that unlike so many other syntactic ambiguities in English, the choice of words in the sentence seems to have little effect on the local modification preference. Thus, we could change the verbs *say* and *left* in (2) to different verbs and change the modifying adverb to a different word or even a prepositional phrase like *in the morning*, and the local modification preference persists across all of these variations in the lexical content. This resistance to lexical influence is one of the primary reasons why these ambiguities have been thought to support a principle-based approach and to be the Achilles' heel of the constraint-based approach to syntactic processing, as lexical information is such a crucial kind of constraint within that framework.

On closer examination, the situation is both more complicated and more interesting than this short description would lead us to believe. If this ambiguity were truly immune to lexical effects, that fact would be a problem for every theory of sentence processing, as essentially every theory incorporates lexical information during at least some stage of the syntactic ambiguity resolution process. Thus, models in which lexical information does not affect initial interpretation but constrains later interpretation and

reanalysis, such as the Garden Path Model (Frazier, 1987) and Tuning (Mitchell, Cuetos, Corley, & Brysbaert, 1995), would be faced with the question of why the lexical information in (2) seems so ineffective in promoting the distant modification case, whereas lexical information in other ambiguities appears to have a strong effect in eventual interpretation. Fortunately for all theories, it turns out that there are real, although extremely subtle, effects of lexical information on the interpretation of these ambiguities. For example, Fodor and Inoue (1994) noted that when the modifying expression contains a negative polarity item, such as *any more*, which must modify a negated verb, this ambiguity is readily interpreted with distant modification, as in (3), where *any more* modifies the distant negated verb *doesn't tell* rather than the local verb *thinking*.

3. John doesn't tell us what he's thinking about any more.

The question still remains, however, why lexical information typically carries so little weight in the interpretation of this ambiguity. From the perspective of constraint-based theories, which stress the sensitivity to distributional information in the language, a likely answer is that there is something about the distributional information in the input that causes other constraints to be weighed much more heavily than lexical ones in interpreting sentences with this ambiguity. Of course framing the question in this way is not itself a solution, but it offers a particular direction for the research, namely identifying the nature of the relevant distributional information and exploring how this information constrains the ambiguity resolution process. This is the approach that I pursue here. The account is both distributional and rather Gricean in character, in the sense outlined by Frazier and Clifton (1996), in that I suggest that the sequence *verb . . . verb modifier* has come to be interpreted with the modifier modifying the local verb because this is essentially the only syntactic structure that can convey this meaning, whereas there exist better alternative syntactic structures to convey the distant modification interpretation. These alternative structures are "better" not from the point of view of comprehension, but from production. In other words, the solution to the puzzle of comprehending verb modification ambiguities is revealed in part through an understanding of how production processes work.

Phrase Length and Production Constraints

When there is an option for ordering phrases in alternative ways in English, choice of phrase order is strongly governed by the length of the phrases, such that the short phrase tends to be uttered first (Hawkins, 1994; Ross, 1967; Wasow, 1997). Some common examples include verb + particle con-

structions, as in (4), and *heavy-NP shift* constructions in (5). In both of these constructions, there is a general preference to place the direct object NP immediately after the verb, as in the (a) examples; an alternative ordering, as in the (b) examples, is quite awkward or ungrammatical, indicated by the asterisk. When the NP becomes particularly long, however, preferences to produce a short-before-long phrase order promotes an alternative phrase ordering in which the NP is not adjacent to the verb (V), as in the (c) examples. Most speakers of English rate the (c) versions to be at least as acceptable as the (d) versions, in which the V-NP adjacency is maintained despite the long NP.

4. a. Mary threw it out.
 b. *Mary threw out it.
 c. Mary threw out the old chicken salad that had sat in the refrigerator for four days.
 d. ?Mary threw the old chicken salad that had sat in the refrigerator for four days out.
5. a. Mary ate chicken for lunch.
 b. *Mary ate for lunch chicken.
 c. Mary ate for lunch the old chicken salad that had sat in the refrigerator for four days.
 d. ?Mary ate the old chicken salad that had sat in the refrigerator for four days for lunch.

Within most accounts of speech production, this preference for the short-long phrase ordering is assumed to stem from the incremental nature of speech production processes (Bock, 1987; Kempen & Hoenkamp, 1987). In this view, the ordering of words and phrases in production is constrained in part by these elements' accessibility, where more accessible words and phrases are ones that are higher in frequency, have more recently been primed, or have some other property that makes retrieval of word forms relatively easy (Bock, 1987). Several researchers have hypothesized that on average, short phrases should require less processing and be more accessible than long ones, so that on many occasions, a shorter phrase is ready to be articulated before a longer phrase (DeSmedt, 1994; Stallings, MacDonald, & O'Seaghdha, 1997).

Given these production constraints, speakers and writers will tend to produce utterances in which short phrases precede long phrases, creating particular distributional patterns in the language. Moreover, exceptions to this phrase ordering should tend to appear for a reason, for example, if the long-short order is necessary to convey a particular meaning; some of the modification ambiguities discussed earlier are exceptions of just this

sort. Consider the examples in (6), where the local site for modification, the second verb (*left*), is part of an embedded sentential complement of the distant verb (*said*). From the perspective of length and phrase ordering, such sentences are interesting in that a long (underlined) phrase, that John had left, precedes a short (italicized) phrase, *yesterday*, seeming to violate the length ordering constraints in English. This phrase order is necessary, however, to modify the second verb and convey the meaning that the leaving event, not the saying event, was yesterday. Alternative orders that follow the general short-long phrase ordering preference, as in (6b-c), do not convey the intended meaning; they instead unambiguously convey distant modification, such that the saying event was yesterday.

6. a. Bill said that John had left *yesterday*. (Long-short order)
- b. Bill said *yesterday* that John had left. (Short-long order)
- c. *Yesterday*, Bill said that John had left. (Short-long order)

This example shows that the technically ambiguous construction *verb . . . verb . . . modifier*, as in (6a), is actually the required ordering for expressing the meaning in which the modifier modifies the local site (the second verb, *left*). This construction can also be used to express the meaning in which the distant verb is modified by *yesterday*, but production constraints in English discourage this usage in favor of structures like (6b-c), in which the short phrase can precede the long one. Production constraints therefore create a distributional pattern in the language in which the long-short order is strongly associated with local modification rather than distant modification. Comprehenders have been shown to be exquisitely sensitive to distributional information of this sort, and they should therefore tend to interpret modification ambiguities in a way that is consistent with distributional information. If so, the local modification preference for structures such as (6a) emerges from comprehenders' sensitivity to distributional information, which in turn emerges from the incremental nature of speech production. There is therefore no need to postulate some local modification parsing principle such as Right Association (Kimball, 1973), Late Closure (Frazier, 1987), or Recency (Gibson et al., 1996) in order to account for interpretation preferences in this construction.

Not only does this account obviate the need for a parsing principle, it makes predictions for variations in the degree of local modification preference across individual sentences, whereas the parsing principles do not. The distributional information relevant to ambiguity resolution in this case specifically concerns those sentences in which a long phrase precedes a short phrase, because the distributional patterns emerge directly from length-sensitive production constraints. Pressure to utter the modifying adverbial expression before the embedded verb phrase (VP) should not

exist if the embedded phrase is not longer than the modifier, as in (7). Here, the modifying adverbial phrase (*very much* or *very slowly*) is longer than the embedded phrase (*swimming*), in contrast to (6). For sentences with short embedded phrases, therefore, comprehenders should not show a strong tendency to interpret a sentence-final ambiguous modifier in favor of local modification, because the length-based production constraint that promotes this interpretation in (6) is not present for sentences with short embedded phrases such as (7).

7. a. Short embedded phrase, Distant modification Mary likes swimming very much.
- b. Short embedded phrase, Local modification Mary likes swimming very slowly.

Thornton and MacDonald (1998) tested these length-based predictions in a self-paced reading experiment. They manipulated the length of the phrase containing the local verb in ambiguous verb modification constructions. In one condition, shown in (8), the embedded phrase was quite long and substantially longer than the modifying phrase at the end of the sentence. In another condition, shown in (7), the embedded phrase was only one or two words long and was as short or shorter than the modifying phrase. Thornton and MacDonald also manipulated the material in the modifying phrase so that it clearly modified either the distant verb (7a, 8a) or the local verb (7b, 8b), thereby providing a disambiguation to the temporary modification ambiguity.

8. a. Long embedded phrase, Distant modification Mary likes it when the dolphins at Sea World are swimming very much.
- b. Long embedded phrase, Local modification Mary likes it when the dolphins at Sea World are swimming very slowly.

Whereas a local attachment parsing strategy would predict shorter reading times for local modification than distant modification independent of the length of phrases in the sentence, the distributional hypothesis advanced here predicts an interaction, such that modification of the local verb is preferred over the distant only when the local verb is embedded in a long phrase. This long phrase creates a sentence with a long-short phrase order, which is the obligatory order for local modification but is strongly dispreferred for distant modification, owing to violations of the short-long constraint. When the local verb is embedded in a short phrase, however, this phrase and the ambiguous modifier are about the same length, so no ordering constraints are violated. In this case, either modification should be acceptable, and no differences in reading times are predicted.

These predictions were confirmed with a reliable Phrase Length \times Interpretation interaction in reading times at the disambiguation region. In the long condition, reading times at the disambiguation (*very much* vs. *very slowly*) were a reliable 20 msec per word longer for the distant modification than for the local modification. When the phrase containing the local verb site was short, however, reading times actually revealed a nonsignificant preference for distant modification. In other words, manipulations of phrase length dictated the presence or absence of a local modification preference in ambiguity resolution, supporting the importance of distributional information concerning phrase length in interpretation of this ambiguity.

Summary

By using an independently motivated account of production (Bock, 1987; Kempen & Hoenkamp, 1987) and an independent account of comprehenders' sensitivity to distributional regularities in the input (e.g., MacDonald et al., 1994; Tabossi et al., 1994), an approach to modification ambiguity emerges that does not require any local modification parsing principles in order to account for interpretation preferences. Moreover, this distributional account offers an explanation of why these preferences vary with the length of certain phrases in the sentence, in contrast to the parsing principles.

This account also sketches a direction for future research into the interpretation of other kinds of modification ambiguities. It offers an explanation for why there is a strong local modification preference for structures in which the local site is a verb phrase like (6) but much weaker interpretation preferences in two other constructions in which the local site is an NP, as in complex NP modification such as (9), in which both local and distant sites are NPs, and ambiguities in which a modifying prepositional phrase (PP) can attach to either a VP (the distant site) or an NP (the local site), as in (10). In each of these examples, the modifier and the head of the phrase being modified are underlined.

9. a. Distant modification: The cat on the rug with long whiskers
 b. Local modification: The cat on the rug with long tassels
10. a. Distant modification: Cynthia saw the woman from the balcony.
 b. Local modification: Cynthia saw the woman from Toledo.

The explanation for the stronger preferences in verb modification ambiguities than in the complex NP constructions (9) or VP-NP constructions (10) is that the verb modification ambiguity is created by embedding an entire sentence (containing the local verb site) within a VP (6). Embed-

ded sentences tend to be long, so that a large number of words typically intervenes between the distant site and the modifier. By contrast, in the complex NP construction (9) and the VP-NP construction (10), there is no embedded sentence, and the local site is an NP that tends not to be longer than the modifying phrase at the end of the sentence. Phrase length therefore exerts little effect on the ordering of phrases in the constructions exemplified in (9–10), so that distributional information concerning phrase length does not strongly promote local attachment in these cases.¹ The preferences that are observed for these constructions are largely due to other constraints, primarily lexical and discourse factors (Altmann & Steedman, 1988; Spivey-Knowlton & Sedivy, 1995; Thornton, MacDonald, & Gil, 1998).

The answer to the first puzzle, then, comes from an understanding of production processes and how these processes might provide distributional information relevant to language comprehension. The next puzzle arises from a deeper scrutiny of the very same production processes that lead to the solution of Puzzle #1. A closer look at length effects in production reveals that these processes may actually pose a problem for the incremental models of production that proved so useful for solving Puzzle #1. Again, framing the solution in terms of distributional information offers a perspective that is not a dominant theme in syntactic production research.

PUZZLE #2: EXCEPTIONS TO INCREMENTAL SPEECH PRODUCTION?

English is relatively inflexible in its ordering of phrases, but as we have seen, variations in the length of phrases promote some nonstandard (typically called *shifted*) phrase orders in production. One such structure, illustrated in (5c), has been termed *heavy-NP shift* because it appears when the direct object NP is very long or “heavy” (see Hawkins, 1994; Wasow, 1997, for review). More recent research has suggested that the relative length of the NP and the other constituent (a prepositional or adverbial phrase) is a better determinant of shifting than simply the length of the NP alone (Hawkins, 1994). Sensitivity to the relative length of the phrases is exactly what would be expected from an incremental production account, so that the choice of ordering of phrases in the VP appears to be constrained by

¹One prediction of this account is that manipulations of the length of the local NP site in (9–10) should affect attachment preferences in these constructions just as in the VP site attachment structure. However, it is very difficult to test this prediction, because manipulation of the length of the NP requires manipulating the amount of prenominal modification the NP receives, and this prenominal modification has been shown to affect attachment ambiguities in at least complex NP constructions (Thornton et al., 1998). Thus, any manipulation of length is confounded with another factor with demonstrated effects on ambiguity resolution.

how accessible each phrase is for the next phase of the production process (DeSmedt, 1994). Moreover, Firbas (1966) suggested that there is a discourse component to heavy-NP shift, in that previously described, or *given* information tends to precede new information in utterances. In this view, the length of the NP is not the basic motivation for shifting but rather a correlate of givenness; new information tends to be expressed using more words than previously given information. The incremental approach clearly has something to say here, in that the words and phrases conveying given information have typically been recently uttered in the discourse, and they will tend to be primed and thus more accessible than the words and phrases that convey the new information. The incremental approach is attractive in that it suggests that phrase length and givenness are not competing explanations for heavy-NP shifting, but rather that phrase accessibility, by virtue of length, givenness, or other factors, governs phrase ordering and thus heavy-NP shifting during production.

The puzzle concerning heavy-NP shifting and incremental production processes is that there seem to be some significant aspects of heavy-NP shifting that are not compatible with a strictly incremental approach to production processes. Stallings et al. (1998) investigated the role of verbs in shifting behavior in a series of experiments in which participants uttered sentences using a set of phrases that they saw on a computer screen. In critical trials, the phrases could be combined to make either a heavy-NP shifted sentence or a sentence with a "basic" V NP PP order in the VP. Stallings et al. measured the frequency with which speakers uttered shifted versus basic order sentences.

Stallings et al. (1998) manipulated the kind of verb that appeared in the sentence fragments on the screen. In one condition, the verb was a simple transitive verb that did not allow other argument structures (except possibly intransitive structures), such as *transferred* or *reviewed*. In a second condition, the verbs, in addition to allowing transitive structures, also permitted sentential complement constructions, in which one sentence, often introduced with *that*, is embedded in another sentence, as in *Mary revealed that the book was missing*. Such verbs are called NP/S verbs in the sentence comprehension literature, reflecting the fact that they can take both NP direct objects (the simple transitive structure) and sentential (S) complements. In all cases, the verbs appeared on screen with a subject, direct object, and PP (never an S-complement); thus, the verb type manipulation refers to the number and kinds of other structures afforded by verbs, and the critical stimulus items always formed simple transitive sentences.

Stallings et al. (1998) found that sentence fragments containing NP/S verbs were uttered in a shifted form about twice as often as those with the simple transitive verbs. This result, which was replicated across a series of experiments using three different production methods, indicates that lexi-

cal properties of the verb constrained the ordering of phrases later in the sentence. In other words, something about the distributional properties of verbs in transitive and S-complement structures affects whether transitive structures will be uttered in a shifted versus basic form.

Why would verb properties affect the ordering of postverbal phrases during sentence production? Stallings et al. (1998) hypothesized that each verb has a "shifting disposition" reflecting past experiences with various syntactic structures. A key feature of this argument is the observation that in heavy-NP shift, the verb and direct object NP are not adjacent to one another in the sentence. Stallings et al. suggested that the frequency with which a verb participates in various kinds of nonadjacent constructions affects a verb's disposition to allow shifting of the material in the VP. They identified two other nonadjacent constructions in addition to heavy-NP shift itself. One is the so-called verb-particle structures previously shown in (4), in which the verb and direct object are separated by a verb particle such as *up*, *out*, *on*, and so forth, as in *throw out the trash*, *clean up the room*, *reel in the fish*. The second kind of nonadjacent structure appears in S-complement constructions. When a verb that has a sentential complement is modified with a prepositional or adverbial phrase in English, the modification may appear between the verb and the complement, as in the examples in (11), where the underlined modifying phrases separate the verb from its S-complement.

11. a. The eccentric director reported in a loud voice that the cast party was canceled.
- b. The young woman discovered almost immediately that her new roommates were slobs.

Given that verb modification is a common occurrence in English, verbs that participate in S-complement constructions will often participate in the constructions shown in (11), which place the verb's complement nonadjacent to the verb. Stallings et al. (1998) hypothesized that a verb's participation in nonadjacent S-complement structures increases a verb's disposition to participate in other nonadjacent structures, including heavy-NP shift.² For example, the fact that the NP/S verbs *reported* and *discovered*

²Wasow (1997) suggested that the relevant generalization is not about nonadjacent structures but rather about the adjacency of verbs and prepositional phrases, that is, verbs that frequently occur adjacent to PPs are frequently found in shifted constructions. In support of this claim, he cited corpus data showing a large number of heavy-NP shifted sentences containing verbs that enter into verb-particle constructions. This account is simpler than the one proposed by Stallings et al. (1998), in that it refers to adjacency of two elements rather than the nonadjacency of two elements because of an intervening third element. The two accounts make different predictions concerning the rate of shifting in sentences with

participate in nonadjacent structures like those in (11) makes them more able to participate in the heavy-NP shift structures shown in (11), whereas simple transitive verbs such as *sang* and *uncovered*, which have not had the nonadjacent experience in S-complement structures like (12), are less compatible with shifting. This comparison between verb types is shown in (12-13); the NP/S verbs in (12) tend to be much better in shifted structures than the simple transitive verbs in (13) in both production measures and acceptability ratings (Stallings et al., 1998). Thus, the choice of syntactic structure in production, although clearly influenced by the accessibility of various planned constituents during the production process, is also strongly influenced by distributional information.

12. a. The eccentric director reported in a loud voice the songs that would be used during the opening act of the play.
- b. The young woman discovered almost immediately the richly embroidered sack of semi-precious stones.
13. a. The eccentric director sang in a loud voice the songs that would be used during the opening act of the play.
- b. The young woman uncovered almost immediately the richly embroidered sack of semi-precious stones.

Stallings et al. (1998) argued that shifting disposition was a property of the distributional information for each individual verb concerning its frequency of participation in nonadjacent structures. This sort of precise lexical knowledge, the frequency with which verbs participate in alternative syntactic structures, is exactly the sort of information that proponents of constraint-based accounts of language point to in accounting for a large number of syntactic processing phenomena (e.g., MacDonald, 1994; MacDonald et al., 1994; McRae et al., 1998; Pearlmutter & MacDonald, 1992; Tabossi et al., 1994; Trueswell et al., 1994). Thus, production, although it does appear to have an incremental component, also appears to have a

biased-intransitive verbs, which are verbs that can be used transitively but are more often used intransitively, such as *walk*, *move*, *work*, and so forth. These verbs do not enter into verb-particle constructions, but they often occur adjacent to a prepositional phrase (e.g., *walk in the park*, *move to the left*), and thus Wasow's (1997) formulation predicts that these verbs should frequently appear in heavy-NP shifted sentences. The Stallings et al. (1998) nonadjacency account makes the opposite prediction: These verbs do not typically occur with direct objects in nonadjacent position, by virtue of the fact that they rarely occur with any direct objects, and they are therefore poor candidates for shifted sentences. MacDonald, Stallings, and O'Seaghdha (1998) presented corpus data indicating that the biased-intransitive verbs are in fact quite rare in shifted sentences, supporting the nonadjacency characterization of verb shifting disposition.

sensitivity to distributional lexical information in a way that has not typically played a dominant role in production theories to date.

These results suggest important parallels between production and comprehension in the use of distributional information, and they raise questions about the extent to which the same distributional information is used by both comprehension and production systems. A number of different scenarios are possible. For example, the system might be entirely modular, such that a person's comprehension system uses distributional information accrued from the history of sentences comprehended but not from those uttered by the person, whereas the production system uses distributional information only from the person's prior utterances. A more likely scenario is that prior comprehension and production events constrain both comprehension and production. Certainly, prior productions should influence comprehension, because the utterances that a speaker produces are also perceived and comprehended by this speaker. From the point of view of production, it is clear that prior comprehension experiences do have some effects on syntactic production (Levelt, 1989; Potter & Lombardi, 1998), although it is possible to imagine scenarios in which distributional information from prior productions are weighed more heavily than information from prior comprehension events. These topics will be interesting ones for future research, but whatever their outcome, it is clear that they have introduced a new puzzle, namely, how and why the adult speaker and comprehender could come to possess all of this tremendously detailed lexical information. In other words, now that we have seen that comprehension and production research have a great deal to say to each other, it becomes clear that both of these fields must also pay attention to research in language acquisition.

PUZZLE #3: HOW IS DISTRIBUTIONAL INFORMATION ACQUIRED?

The sorts of distributional information that we have discussed in comprehension and production are really phenomenally detailed, and these examples represent only a small sample of the kinds of information that have been shown to be important in language processing. A brief survey of some current findings in Table 6.1 attests to both the varied nature of distributional information and the rapidity with which new kinds of distributional information are being uncovered. Such varied information would seem daunting to learn, and indeed some researchers have suggested that it is beyond the capability of humans to attend to, store, and use such information efficiently (Mitchell et al., 1995). Clearly, an account of production or comprehension that relies so heavily on distributional informa-

TABLE 6.1
A Sampling of Distributional Information Hypothesized to be Important in Comprehension and Production

<i>Type of Information</i>	<i>Source</i>
Frequency of past tense versus past participle uses of a verb	MacDonald, 1994; MacDonald, Pearlmuter & Seidenberg, 1994; Trueswell, 1966
Frequency of transitive versus intransitive uses of a verb	MacDonald, 1994; MacDonald et al., 1994; Trueswell, Tanenhaus, & Garnsey, 1994
Frequency of transitive versus sentential complement uses of a verb	Garnsey, Pearlmuter, Myers, & Lotocky, 1997; Trueswell, Tanenhaus, & Kello, 1993
Frequency with which a verb takes S-complements with and without overt <i>that</i>	Trueswell, Tanenhaus, & Kello, 1993
Frequency of long-short phrase order in verb modification ambiguities	Thornton & MacDonald, 1998
Frequency with which verb and complement are nonadjacent	Stallings, MacDonald, & O'Seaghda, 1998
Frequency with which a verb co-occurs with a PP conveying an Instrument role	Taraban & McClelland, 1988; Spivey-Knowlton & Sedivy, 1995

tion cannot afford to ignore the question of how such information is acquired.³

An important related question is why such information is acquired—why would a speaker of a language spend a lifetime encoding a finely detailed history of many co-occurrences between words and structures? This question may actually not be that difficult to answer, or at least to frame the shape of a response, as the basis for an answer has long existed in the acquisition literature: Distributional information appears to be crucial for acquiring a large amount of (at least) language-particular information, so that an acquisition system that did not pay attention to distributional information would not in fact successfully acquire language. This view can be seen most explicitly in Bates and MacWhinney's (1989) Competition Model, but the central role for distributional information is not limited to this perspective. For example, a number of researchers have suggested that information in the speech signal can be used by children to aid in a number of acquisition tasks, including finding word boundaries in the speech stream and identifying lexical categories and syntactic structure (Gleitman & Wanner, 1982; Morgan, 1986; Morgan & Demuth, 1996; Saf-
fran, Aslin, & Newport, 1996). These accounts are inherently distributional—the information in the speech signal is probabilistic, and the child must attend to the distributional properties of the input over some time

³This is not to say that one must have a full account of the origins of distributional information in order to pursue constraint-based accounts of processing. Indeed, MacDonald (1997) argued that the origin question and the question of how distributional information is used can be profitably investigated separately.

in order to obtain useful information. Similar claims can be made about semantic bootstrapping accounts (Pinker, 1984, 1987), in which the child combines innate syntax-semantics linking rules with information in the environment to learn the syntactic properties of words in the language. The mapping between the world and the linguistic context is of course complicated and probabilistic (Gleitman, 1990), so that a distributional analysis is an important component for the success of this sort of account. Finally, syntactic bootstrapping approaches, in which mappings between syntactic structure and the world aid the child in acquiring semantic information about words (e.g., Gleitman, 1990; Landau & Gleitman, 1985), are also reliant on distributional information, as the mappings between the events and the syntactic structure of utterances are variable. An important emphasis in this work is on the role of noun information in the acquisition of verbs, particularly the identification of noun arguments of a verb. This sensitivity to the relation between verbs and nouns should also underlie the acquisition of distributional information affecting what Stallings et al. (1998) called a verb's *shifting disposition*—information about the (non)adjacency of verbs and their complement thought to constrain heavy-NP shifting. In other words, the very same information that is crucial in the acquisition of verb semantics in childhood appears to constrain the production of heavy-NP shift in the adult state.

This sort of claim, that the same general kinds of distributional information that guide the acquisition process in childhood continue to have an important role in adult performance, emerges from an alternative account of the nature of the acquisition process. Posed in the traditional way, the process of language acquisition is acquiring a grammar and a lexicon, and in this view it is mysterious why a lifetime of distributional information must be recorded after the grammar and lexicon are in place. Another perspective (e.g., Seidenberg, 1997) suggests that the acquisition process is not so isolated from the immediate demands of comprehension and production, in that the goal of the child is not acquiring an adult grammar but rather understanding others and being understood as much as possible in each communicative event. In this view, there is not an "acquisition device" separate from comprehension and production processes.

An illustration of these ideas can be found in a connectionist model of verb acquisition developed by Allen (1998). This model acquired verb semantics from the pairing of child-directed speech taken from the CHILDES (Child Language Data Exchange System; MacWhinney, 1995) corpus, particularly verb argument structure information, and an interpretation—the set of events accompanying the speech. For example, for the transitive sentence *Peter broke the pencil*, the model received the argument structure information that there were two arguments, one before the verb and one after, and that the verb was *break*. This information was paired

with the interpretation that there were two participants, one the agent and one the patient, and that the event consisted of a breaking event. The goal of the modeling effort was to use knowledge acquired from exposure to these pairings to activate both appropriate argument role interpretations and verb semantics for each utterance, including constructions on which the model had not been trained, such as *the bottle broke*. Allen's model performed well, exhibiting both the ability to supply role interpretations for novel constructions and to activate appropriate verb semantics for novel verbs given information about both the argument structure and the semantics of the arguments in the utterance.

More important for our purposes, Allen (1998) showed that the model took advantage of a great deal of distributional information in the input to acquire its verb representations. This information included, in approximately descending order of importance, the frequency with which a verb was used, the set of constructions the verb appeared in, the frequency with which a verb was used in particular constructions, the semantic relation between a verb and other verbs used in similar constructions, the combined frequencies of related verbs, and the size of the set of semantically related verbs. These factors combined to form neighborhoods of verbs with semantically mediated privileges of co-occurrence. As is obvious from comparisons between this list and the adult comprehension and production studies shown in Table 6.1, Allen's model acquired distributional information that looks very similar to the constraints that are used in adult performance. This similarity is not accidental, because the task of the model was not acquisition per se but rather a primitive version of what human adult comprehenders do, namely, assign a representation to each input sentence. In the course of assigning these representations, Allen's model passes activation across various levels of representation, and each utterance affects the weights between connections in the network. In this system, encoding of distributional information does not stem from some specialized acquisition mechanism but is rather an inevitable consequence of this kind of processing architecture when applied to the task of comprehension or production.

THE PATH THROUGH THE PUZZLES, AND A MORAL

In this chapter, I argued that a sensitivity to distributional information is an important link between comprehension, production, and acquisition. In the case of comprehension, I suggested that a set of seemingly puzzling modification ambiguities reveal comprehenders' sensitivity to distributional patterns concerning the relative length of phrases in the input. The distributional patterns are thought to emerge from pressures on the production system to

produce phrases in a certain order. The production system itself was also shown to be extremely sensitive to distributional information, such that information about the typical location of verb complements, probably acquired from both comprehension and production, constrains the choice of syntactic structure during production. This wealth of distributional information in the adult state forces us to grapple with acquisition issues concerning how such information comes to be encoded. The answer suggested here was that in many constraint-satisfaction theories of acquisition, production, and comprehension, each comprehension or production event changes the nature of the linguistic representation. Such changes are substantial in early phases of learning and very minor in later phases, but they still serve to encode the distributional patterns of the language as a natural consequence of the comprehension and production processes.

This view suggests a moral: If this general account is on the right track, then the acquisition, comprehension, and production processes have links between them that cannot be safely ignored. First, work in comprehension is likely to have a lot to learn from constraints imposed on the production system, well beyond the ones discussed here. For example, consider the problem of how a comprehender determines whether a phrase such as *she* or *the cat* refers to some new entity in the discourse or whether it is coreferential with some earlier-mentioned entity (and if so, what entity). Speakers' choices for nominal expressions, for example, *she*, *the cat*, *Trinity*, and so forth, appear to be guided by a number of constraints such as whether the referred-to entity is the topic of the discourse, whether the entity has been previously referenced, the grammatical role of the entity in the sentence in its current or past mention, and many other factors (Ariel, 1990; Givón, 1976; Gordon, Grosz, & Gilliom, 1993). These constraints clearly create distributional patterns in the input that are likely to be extremely helpful to the comprehender, and they clearly are not the only examples of the intricate interplay between comprehension and production processes.

Similarly, the vast amount of distributional information available in the adult state has implications for acquisition research. For example, there is some debate in this literature whether one kind of bootstrapping mechanism—syntactic, semantic, or prosodic—is *the* mechanism underlying lexical acquisition. Many researchers adopt the position that multiple kinds of bootstrapping are likely to be at work (e.g., Jusczyk, 1997; Morgan, 1986; cf. Pinker, 1984). The variety of distributional information evidenced in the adult state points toward the multiple constraints approach. Of course, demonstrating sensitivity to distributional information in the adult state is not the same thing as showing that a child actually uses such information at a particular time to solve a particular problem, but the fact that such a huge variety of information is demonstrable in adults minimally requires that comprehenders have encoded the information, and it encourages us

to place our bets in favor of a system in which a variety of constraints simultaneously shape the comprehension and production processes not only in the adult state but also in the child. Examples such as these suggest that the traditional distinctions between acquisition, production, and comprehension are not actually realized in language users, and that psycholinguistic researchers should find it increasingly profitable to break the traditional rule about keeping these subfields separate.

ACKNOWLEDGMENTS

This research was supported by National Science Foundation Grant SBR-9511270. I thank Joe Allen, Robert Thornton, and Lynne Stallings for helpful comments on an earlier draft of this chapter. Correspondence may be directed to Maryellen MacDonald, Hedco Neuroscience Building, University of Southern California, Los Angeles, CA 90089-2520.

REFERENCES

- Allen, J. (1998). *Argument structures without lexical entries*. Unpublished doctoral dissertation, University of Southern California, Los Angeles.
- Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30, 191–238.
- Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.
- Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–76). New York: Cambridge University Press.
- Bock, J. K. (1987). An effect of the accessibility of word forms on sentence structure. *Journal of Memory and Language*, 26, 119–137.
- DeSmedt, K. J. M. J. (1994). Parallelism in incremental sentence generation. In G. Adriaens & U. Hahn (Eds.), *Parallelism in natural language processing*. Norwood, NJ: Ablex.
- Firbas, J. (1966). On defining the theme in functional sentence perspective. *Travaux Linguistiques de Prague*, 2.
- Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research*, 23, 407–434.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and Performance XII: The psychology of reading* (pp. 559–586). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.
- Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58–93.
- Gibson, E., Pearlmutter, N. J., Conesco-Gonzales, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition*, 59, 23–59.
- Givón, T. (1976). Topic, pronoun, and grammatical agreement. In C. N. Li (Ed.), *Subject and topic*. New York: Academic Press.

- Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition*, 1, 1-55.
- Gleitman, L. R., & Wanner, E. (1982). The state of the state of the art. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3-48). Cambridge, England: Cambridge University Press.
- Gordon, P. C., Grosz, B. J., & Gilliom, L. (1993). Pronouns, names and the centering of attention in discourse. *Cognitive Science*, 17, 311-348.
- Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge, England: Cambridge University Press.
- Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.
- Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11, 201-258.
- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, 2, 15-47.
- Landau, B., & Gleitman, L. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, 9, 157-201.
- MacDonald, M. C. (1997). Lexical representations and sentence processing: An introduction. *Language and Cognitive Processes*, 12, 121-136.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.
- MacDonald, M., Stallings, L., & O'Seaghdha, P. G. (1998). *Distributional information and heavy-NP shift*. Manuscript in preparation.
- MacWhinney, B. (1995). *The CHILDES project: Tools for analyzing talk* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- McRae, K., Spivey-Knowlton, M., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283-312.
- Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, 24, 469-488.
- Morgan, J. (1986). *From simple input to complex grammar*. Cambridge, MA: MIT Press.
- Morgan, J., & Demuth, K. (Eds.). (1996). *From signal to syntax*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Pearlmutter, N. J., & MacDonald, M. C. (1992). Plausibility and syntactic ambiguity resolution. In *Proceedings of the 14th Annual Conference of the Cognitive Society* (pp. 498-503). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 399-441). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, 38, 265-282.
- Ross, J. R. (1967). *Constraints on variables in syntax*. Unpublished doctoral dissertation, MIT, Cambridge, MA.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926-1928.
- Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, 275, 1599-1603.

- Spivey-Knowlton, M. J., & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, 55, 227-267.
- Stallings, L. M., MacDonald, M. C., & O'Seaghdha, P. G. (1998). Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy NP shift. *Journal of Memory and Language*, 39, 392-417.
- Tabossi, P., Spivey-Knowlton, M. J., McRae, K., & Tanenhaus, M. K. (1994). Semantic effects on syntactic ambiguity resolution: Evidence for a constraint-based resolution process. In C. Umiltà & M. Moscovitch (Eds.), *Attention and Performance XV* (pp. 589-616). Cambridge, MA: MIT Press.
- Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language*, 27, 597-632.
- Thornton, R., & MacDonald, M. C. (1997). *The role of phrase length in interpretation of modification ambiguities*. Manuscript in preparation.
- Thornton, R., MacDonald, M. C., & Gil, M. (1998). *Pragmatic constraints on modifier attachment to complex NPs in English and Spanish*. Manuscript submitted for publication.
- Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, 35, 566-585.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, 33, 285-318.
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing—separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 528-553.
- Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and Change*, 9, 81-105.