Connection Science
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713411269

Cryptotype, Overgeneralization and Competition: A Connectionist Model of the Learning of English Reversive Prefixes

Ping Li; Brian MacWhinney

Online publication date: 01 July 2010

To cite this Article Li, Ping and MacWhinney, Brian(1996) 'Cryptotype, Overgeneralization and Competition: A Connectionist Model of the Learning of English Reversive Prefixes', Connection Science, 8: 1, 3 — 30

To link to this Article: DOI: 10.1080/095400996116938

URL: http://dx.doi.org/10.1080/095400996116938

Connection Science, Vol. 8, No. 1, 1996, 3–30

Cryptotype, Overgeneralization and Competition: A Connectionist Model of the Learning of English Reversive Prefixes

PING LI & BRIAN MACWHINNEY

(Received for publication December 1992; revised paper accepted February 1996)

This study examined the role of covert semantic classes or 'cryptotypes' in determining children's overgeneralizations of reversive prefixes such as un- *in* *unsqueeze *or* *unpress. *A training corpus of 160 English verbs was presented incrementally to a backpropagation network. In three simulations, we showed that the network developed structured representations for the semantic cryptotype associated with the use of the reversive prefix* un-. *Overgeneralizations produced by the network, such as* *unbury *or* *unpress, *match up well with actual overgeneralizations observed in human children, showing that structured cryptotypic semantic representations underlie this overgeneralization behaviour. Simulation 2 points towards a role of lexical competition in morphological acquisition and overgeneralizations. Simulation 3 provides insight into the relationship between plasticity in network learning and the ability to recover from overgeneralizations. Together, these analyses paint a dynamic picture in which competing morphological devices work together to provide the best possible match to underlying covert semantic structures.*

keywords: Connectionist model, language acquisition, cryptotype.

1. Introduction

In one of the classic papers of early cognitive linguistics, Whorf (1956, p. 71) argued that links between language and culture are often most fully revealed in covert grammatical categories. Using the English reversive prefix *un-* as an illustration, Whorf called attention to the simple fact that English speakers can produce a wide range of verbs with this prefix, such as *uncoil*, *uncover*, *undress*, *unfasten*, *unfold*, *unlock*, *untie* or *untangle*. However, there are many other seemingly parallel forms that are not allowed, such as **unbury*, **unfill*, **ungrip*, **unhang*, **unpress*, **unspill*, **unsqueeze* or **untighten*.[1] Why are some of these derivations permitted and others not? Whorf believed that there was a category underlying all these formations that made its presence known only through the restrictions that it placed on the prefix *un-*. In contrast to 'overt' or 'phenotypic' grammatical

P. Li, Department of Psychology, University of Richmond, Richmond, VA 23173, USA. B. MacWhinney, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA. E-mail: ping@psy.cuhk.hk.

0954-0091/96/010003–28 $7.50

categories, such as the past tense (*-ed*) or the plural (*-s*), this category was not marked by a surface morpheme, but only by its effects on the licensing of possible combinations. Because this category functions only covertly, Whorf called it a 'cryptotype'.

Whorf further noted that, for this category: ''we have no single word in the language which can give us a proper clue to its meaning or into which we can compress this meaning; hence the meaning is subtle, intangible, as is typical of cryptotypic meanings''. None of the standard categories of Latin grammar can be used as a basis for a rule to tell us when we can use *un-* and when we cannot. The distinction is not grounded on some single feature such as 'transitivity', 'iterativity' or 'intentionality'. Instead, according to Whorf, the set of verbs that can be prefixed with *un-* seems to share a ''covering, enclosing, and surface-attaching meaning''. Only verbs that partake in this cryptotype are licensed to receive the *un-* prefix.

We believe that the study of small semantic fields, such as those underlying Whorfian cryptotypes, can have four important implications for connectionist models of language learning:

(1) *Understanding semantic structures.* The process of constructing working simulations for specific word derivations can force us to restate the notion of a cryptotype in detailed mechanistic terms. In these simulations, distributed representations play an important role in replacing the older analytic frameworks of categories and rules (Lakoff, 1987; MacWhinney, 1989).
(2) *Semantic grounding.* More generally, it is important that connectionist models of language learning be grounded on more complete and realistic semantic analyses (Cottrell & Plunkett, 1994). Analysis of small semantic fields is a good starting point for providing a detailed semantic grounding to neural network models.
(3) *Productivity and overgeneralization.* By attempting to model the empirical data on derivational overgeneralizations such as **unsqueeze* or **unspill*, we can deepen the link between models and complex developmental data (Bowerman, 1982; Clark *et al.*, 1995).
(4) *Competition.* By looking not only at a single prefix such as *un*, but also at competing prefixes such as *dis-* or *mis-*, we can obtain a better understanding of how cryptotypes work within the larger framework of language production.

Before presenting our simulations, we would like to consider each of these four issues in further detail. In each case, we are interested in ways in which this initial study of a single, limited semantic field can provide us with conceptual underpinnings for a more broadly based, semantically grounded, connectionist model of language processing and language acquisition.

1.1. *Understanding Semantic Structures through Connectionist Modelling*

Whorf's understanding of the cryptotype which licenses the reversive prefix *un-* is based on a ''covering, enclosing, and surface-attaching meaning''. Should this meaning be viewed as a single unit, as three separate meanings or as a cluster of related meanings? Do these notions of attachment and covering exhaust the subcomponents of the cryptotype, or are there additional underlying components? Subsequent analyses have suggested certain additional components not initially mentioned by Whorf. For example, Marchand (1969) and Clark *et al.* (1995) argue that all verbs that license *un-* involve a change of state. In addition, these verbs involve a transitive action that has a direct object. This transitive action typically reaches a terminal point in time, in which case it is called a 'telic' verb; alternatively, it reaches some end-state or result, in which case it is called an 'accomplishment' verb (Vendler, 1967). When the meaning of a verb does not involve a change of state or does not indicate telicity or accomplishment, then the verb cannot take *un*-. Therefore, verbs such as **unswim*, **unplay* and **unsnore* are ill-formed semantically, because the base forms involve continuous actions without a terminal point or end-state that could be reversed (Horn, 1988).

A connectionist implementation of the semantics of the reversive cryptotype provides us with a natural way of capturing these insights in a formal mechanism. In this implementation, there can be several 'mini-cryptotypes' which work together as interactive 'gangs' (McClelland & Rumelhart, 1981) to support the formation of the larger cryptotype. These mini-cryptotypes are not in competition; instead, they work in terms of summed activation to support the licensing of *un*- for a particular verb. For example, 'enclosing' verbs, such as *coil*, *curl*, *fold*, *reel*, *roll*, *screw*, *twist* and *wind*, all seem to share a meaning of 'circular movement'. Another mini-cryptotype includes verbs such as *bind*, *buckle*, *fasten*, *latch*, *leash*, *lock*, *strap*, *tie* and *zip*, which have a 'binding' or 'locking' meaning. A third mini-cryptotype includes 'covering' verbs such as *cover*, *dress*, *mask*, *pack*, *veil* and *wrap*. Finally, a fourth mini-cryptotype includes 'attaching' verbs, which usually involve hand movement, such as *clasp*, *fasten*, *hook*, *link*, *plug* and *tie*.

These mini-cryptotypes or mini-gangs can interact cooperatively, because they are closely related to one another. For example, the verb *screw* in *unscrew* may be viewed as having both a meaning of circular movement and a meaning of binding or locking, while the verb *zip* in *unzip* may be viewed as sharing both the 'binding/locking' meaning and the 'covering' meaning. Moreover, both *screw* and *zip* involve hand movements. In addition to such overlaps of semantic features, a verb may also have a feature in its inherent meaning at varying degrees of strength. For example, the verb *wrap* may be viewed as having the covering meaning. However, in some cases, the action of wrapping may also involve circular movements. These properties of feature overlap and degraded featural composition lend themselves naturally to the distributed representations used in neural networks. While it seems difficult for symbolic representations to come up with satisfactory accounts for cryptotypes, the distributed representations and the non-linearities of neural networks seem to be ideal for handling the elusiveness and gradience of these semantic structures.

1.2. Semantic Grounding

The second issue that we address in this work is the role of semantic grounding in neural network modelling. In the past decade, there has been heated debate on connectionist and symbolic models of learning in the acquisition of the English past tense (Hoeffner, 1992; MacWhinney & Leinbach, 1991; Pinker & Prince, 1988; Plunkett & Marchman, 1991, 1993; Rumelhart & McClelland, 1986). An important limitation of the connectionist simulations involved in this debate has been that the input to the network included only phonological information but no true semantic information. The use of only phonological information in these simulations was based largely on considerations of practicality and simplicity. However, this simplification is at odds with the basic emphasis that connectionist models place on non-modular cue interaction in word recognition and word production.

In an effort to broaden the connectionist approach to word formation, Cottrell and Plunkett (1991), Hoeffner (1992) and MacWhinney (1996a) have used randomly generated input that is supposed to represent semantic structure. Although this inclusion of a random, schematized semantic structure helps to bring us closer to a realistic simulation, it is clear that we would be on much more solid ground if we could code our input data in terms of descriptively meaningful semantic representations. Because of the primitive state of the art in computational lexicography, we cannot do this for large segments of vocabulary. However, if we work with small, well-defined areas of the lexicon, then it is possible to construct a reasonable and coherent semantic input set. The construction of such a realistic semantic input set is one of the principal goals of the current work.

1.3. Productivity and Overgeneralization

The third issue to which this work relates is the issue of productivity and overgeneralization in language development. If speakers simply learned verbs such as *untie* and *uncoil* by rote and showed no awareness of the productivity of the cryptotype, then we could easily dismiss Whorf's description as a figment of an overheated linguistic imagination. However, Whorf was careful to remind us that, despite the difficulty that linguists experience in characterizing this cryptotype, native speakers of Engish have an intuitive feel for which verbs can and cannot be prefixed with *un-*. He presented a thought experiment based on what is now a standard procedure in recent psycholinguistic investigations (see, for example, Gropen *et al.*, 1992; Pinker *et al.*, 1987). Whorf reasoned that, if we are told that *flimmick* means 'to tie a tin can to something', then we are willing to accept the sentence 'He unflimmicked the dog' as expressing the reversal of the 'flimmicking' action. However, if we are told that *flimmick* means 'to take apart', then we will not accept 'He unflimmicked the puzzle' as describing the act of putting a puzzle back together. Whorf took this as evidence in support of his claim that we all possess an intuitive grasp of the cryptotype that underlies morphological productivity for the reversive.

Children's overgeneralization errors also provide evidence of the reality of the cryptotype. Bowerman (1982, 1983, 1988) showed that the learning of *un-* goes through a four-staged developmental pattern. In the first stage, children treat *un-* and its base verb as an unanalyzed whole and produce *un-* verbs in appropriate contexts. This initial stage of rote control is analogous to the child saying *went* without realizing that it is the past-tense form of *go*. According to Clark *et al.* (1995), children talk about the reversal of actions long before they have acquired the productive use of *un-*. They rely on particles such as *off* and *back* or verbs such as *open* to express the notion of reversal. Although children understand the meaning of reversal at this early stage, this understanding does not automatically lead to productive uses of *un-*.

The second stage in the development of *un-* begins around age 3, with the first overgeneralizations in spontaneous speech. In an elicitation task, Clark *et al.* found that children's use of *un-* increased steadily with age from 3 to 5, with older children producing overgeneralizations such as **unbend*, **unbury*, **uncrush*, **ungrow*, **unstick* and **unsqueeze*. These elicited overgeneralizations match up well with spontaneous overgeneralizations such as **unarrange*, **unbreak*, **unblow*, **unbury*, **unget*, **unhang*, **unhate*, **unopen*, **unpress*, **unspill*, **unsqueeze* or **untake* (Bowerman, 1982; Clark *et al.*, 1995). At this stage, the overgeneralized *un-* verbs do not all respect Whorf's cryptotype.

In the third period of development, children begin to restrict overgeneralizations to forms that fit within the cryptotype, but whose adult forms do not exist, such as **unbury*, **unpress* and **unsqueeze*. During this same period, we also find certain 'overmarking' errors. For example, the child might say **unopen* and really only means to say *open*, or the child might say *unloosen* and really only means *loosen*. In such cases, the base forms *open* and *loosen* have a reversive meaning that triggers the attachment of the prefix, even when the action of the base meaning is not actually being reversed. Errors of this type include the forms **unopen*, **untangle*, **unplug* and **unloosen* which are attested in corpora in the Child Language Database Exchange System (CHILDES) (MacWhinney & Snow, 1985, 1990; MacWhinney, 1995), as well as in elicited errors reported by Clark *et al.* (1995). These overmarking errors are analogous to redundant past-tense marking in **camed* and redundant plural marking in **feets* (Brown, 1973). The CHILDES database includes other errors, most of which fit the cryptotype of *un-*, such as **unblow*, **unbuild*, **uncatch*, **uncuff*, **unhand*, **unlight*, **unpull*, **unstick* and **unzipper*, that can be found in the Brown, Clark, Gleason, Kuczaj, MacWhinney and Snow corpora. Appendix A lists examples of reversive errors in children's speech as reported by Bowerman (1982, 1988) and Clark *et al.* (1995), along with the glosses and the context in which they were produced.

In the fourth or final stage, children begin to display adult-like control of the reversives, and errors with these verbs decline.

1.4. *Competition*

The cooperative effects that support the operation of the *un-* cryptotype are matched by competitive effects from other negative prefixes. The commonly used negative prefixes in English include *de-*, *dis-*, *in-*, *mis-* and *un-*. Of these, the two reversive prefixes that are in closest competition are *un-* and *dis-*. These designate the reversal of the action specified by the base verb, as in *untie* and *disconnect*.

Unlike *un-*, the prefix *dis-* has received little discussion in the child language literature. However, *dis-* is equally interesting and important in our view. According to Horn (1988), *dis-* competed successfully against *un-* during the Middle English period to take overmarking of stative verbs such as *displease* or *distrust* (which had been *unplease* and *untrust*). As a result, the scope of *un-* was narrowed to only action verbs. As a consequence of historical change, verbs that take *un-* are typically Germanic in origin, whereas verbs that take *dis-* are typically Romanic in origin. In Modern English, *dis-* and *un-* still compete as alternative devices for marking reversal (Bauer, 1983; Marchand, 1969). This competition involves a close overlap in the basic function of the prefix and a semantic overlap between the cryptotypes involved in the verbal stems. For example, the base verbs in *disassemble*, *disconnect*, *disengage*, *disentangle*, *dismantle*, *dismount* and *disunite* all fit the cryptotypic meanings of binding, covering and attaching, which are also involved in the cryptotype for verbs that take *un-*. One result of this overlap is that many of the *dis-* verbs and *un-* verbs are synonymous: for example, *disconnect* vs *unlink*, *disentangle* vs *untangle*, *dismount* vs *unload*, *disengage* vs *uncouple* or *disjoin* vs *unyoke*. Another result of the close competition between *dis-* and *un-* is that some verbs allow both *dis-* and *un-*, but with different meanings: for example, *uncover* and *discover*. Finally, some *dis-* verbs have counterpart *un-* verbs in their past-participle forms, such as *disconnected* vs *unconnected*, *disconfirmed* vs *unconfirmed* and *disarmed* vs *unarmed.* Although the meanings of these pairs are not the same, they indicate the nature of competition between the two reversive prefixes.

Despite the fact that many verbs use the prefix *dis*-, it appears that *un*- is now far more productive for new verbs. Many of the uses of *dis*- in the lexicon are no longer available analytically. In some cases, there is no positive form of the *dis*- verb: for example, *discuss*, *dispel*, *disturb* and *distort*. In other cases, there is no apparent semantic relationship between the positive verb form and the negative *dis*- form: for example, *dismiss*, *dispose*, *dissolve* and *display*. In still other cases, the meaning of *dis*- is simply negation, and not reversal, as in *disagree*, *disapprove* and *disallow*. Furthermore, *dis*- is used for many abstract mental verbs that children are unlikely to use until they are much older. These facts suggest that the child may have to learn many of the *dis*- verbs by rote. The generalization of *dis*- to novel forms may, to some extent, be constrained by this kind of rote learning.

The limited productivity of *dis*- in the adult language does not necessarily suppress the child's early ability to generalize, especially if the child encounters both *un*- and *dis*- in the same kind of negative context. There are cases in Bowerman's (personal communication, 1992) data in which children replace *disenroll* with **unroll*, and *disarrange* with **unarrange*. In these cases, *un*- and *dis*- are clearly in competition. There is also evidence that children sometimes treat *dis*- as a separate form, even though it is not separable from the base form in the adult languages. For example, in Hall's data (CHILDES English database), the child says 'it's dising appear', showing that she treats *dis*- and *appear* as two distinct components. By examining the performance of *dis*- (together with *un*-) in our network, we may be better able to understand the processes involved in the acquisition of reversive prefixation. However, because there has as yet been no empirical report on the acquisition of *dis*-, our simulation results with this prefix are presented as generating hypotheses to be tested experimentally.

In summary, by examining these four issues in this study, we hope to provide insights into the representation of verb semantics and the learning of English reversive prefixes. To achieve this goal, we constructed an incremental backpropagation network to learn the reversive *un*- and *dis*- in three different sets of simulations.

2. Simulation 1

2.1. Method

2.1.1. Input corpus. In this study, the network was trained to map meanings of English verbs on to the different prefixation patterns. The input to the network was a corpus of verbs encoded as semantic feature vectors. A total of 160 verbs were selected from two sources: *Webster's New Collegiate Dictionary* and the corpus of Kucera and Francis (1967; henceforce K&F). Our data set consisted of 49 *un*- verbs, 19 *dis*- verbs and 92 'zero verbs', which take neither *un*- nor *dis*- (see Appendix B for a complete list of the verbs). Webster's contains other verbs prefixed with *un*- or *dis*-, but many of these—such as *unwish*, *unlive* and *disannul*—are unacceptable or unfamiliar to modern-day English speakers. The final selection of the 49 *un*- verbs and 19 *dis*- verbs was based on native-speakers' judgements of the acceptability of all the non-archaic *un*- and *dis*- verbs that appeared in Webster's. In the case of *un*-, 14 subjects were asked to rate how good each verb sounded to them on a scale of 1 to 7 (from 'completely weird' to 'perfectly natural') and only those with an average rating of 5.0 or above were selected. The 19 *dis-* verbs also excluded rarer, highly abstract items, such as *disentail*, *disfranchise* and *disinherit*; forms that are not clearly primarily verbal, such as *disadvantage*; and forms with no true base, such as *dispel*, *distort* or *disturb*. Finally, we randomly selected 92 zero verbs from Webster's that can be prefixed with neither *dis-* nor *un-*. Half of these were high-frequency verbs (above 100) in the K&F counts and half were low-frequency verbs (below 100). The relatively higher proportion of zero verbs as compared with *un-* and *dis-* verbs is intended to represent the distribution of these forms in the input to children.

Each of the 160 verbs was encoded as a set of semantic features with continuous values. Because there has been no systematic account of the *un-* and *dis-* verbs with respect to their semantic composition, the final selection of the 20 semantic features was based partly on the limited literature available on this topic (Whorf, 1956; Marchand, 1969; Levin, 1993), and partly on our own linguistic analysis (see Appendix C for a complete list of the features). These features include some general characteristics of actions (features 1–6; see Appendix C), relationships between entities (features 7–15) and joint properties of entities (features 16–20). They are designed to capture the semantic range of the verbs that can be prefixed with *un-* and *dis-*, as well as verbs that undergo no prefixation. These features, when combined in a distributed representation, provide a semantic basis for distinguishing verbs that can take the reversive prefixes from those verbs that cannot. Our feature coding focused on an attempt to capture basic linguistic and functional properties. It is possible that a more elaborate feature coding process could further emphasize the distinctions between individual verbs in a way that could facilitate aspects of learning and generalization (Plaut *et al.*, 1996; Plunkett & Marchman, 1991).

The assignment of particular features to particular verbs was based on empirical data. We presented 15 native speakers of English with the 160 verbs and the 20 semantic features, and asked them to evaluate each verb with respect to each feature, to determine whether or not the particular feature applies to that particular verb. Subjects rated the feature as being relevant to the verb if they thought that the feature was characteristic of, or typically involved in, the situation denoted by the verb. Therefore, for each subject, we had a feature-by-verb matrix of 0s and 1s (0 means that the feature is irrelevant to the verb, and 1 means that the feature is relevant). In this way, the averaged rating scores from the 15 subjects were the graded patterns used as input to the network. In this case, each verb was encoded as a vector of the 20 features, with values between 0 and 1. Although there were varying degrees of similarity among verbs represented in this way, the representation of a given verb was distinct from that of any other verb. In other words, no two verbs shared exactly the same values for all the 20 features. Some examples are given in Appendix C, along with the features.

To evaluate the validity of the results from our semantic judgement experiment, we conducted a hierarchical cluster analysis on the verbs encoded as feature vectors. The results show that synonymous words tended to group together as clusters, indicating that subjects were consistent in their coding of verbal features.

2.1.2. Network architecture. Our simulations used a standard three-layer backpropagation network (Rumelhart *et al.*, 1986). There were 20 input units encoding the verbal semantic features, six hidden units, and three output units

Table I. Vocabulary structure across prefixation patterns in Simulation 1

Total	*dis-*	*un-*	Zero
20	2	8	10
40	6	18	16
60	9	23	28
80	11	30	39
100	12	37	51
120	15	43	62
140	19	49	72

Note: The 20 zero verbs in the initial training are not included here.

representing *un-*, *dis-* or zero prefixation. All simulations reported in this study used this same basic architecture. The learning rate and momentum were held constant across all simulation trials. The simulations were conducted using the TLEARN program configured at the Center for Research in Language, University of California at San Diego.

2.1.3. Task and procedure. The task for the network was to learn to classify the verbs into three categories: those that can be prefixed with *un-*; those that can be prefixed with *dis-*; and those that cannot be prefixed with either.

We applied an incremental learning schedule to reflect more realistically the realities of lexical acquisition (Elman, 1993; MacWhinney & Leinbach, 1991; Plunkett & Marchman, 1993). Children typically build up their vocabulary in an incremental fashion, rather than learning all words at once. In the incremental schedule used here, lexical items entered the training corpus one by one, although with different rates at different learning stages. Before the network learned 60 verbs, the rate of vocabulary growth was one verb every five epochs of training. Between 60 and 100 verbs, the rate increased to one verb every three epochs of training. After 100 verbs, the rate was one verb every one epoch of training. This increasing rate was intended to reflect the accelerating function in children's vocabulary growth (Plunkett & Marchman, 1993). Table I presents the vocabulary increase process after the initial training in Simulation 1.

Prior to incremental training, the network was trained on 20 high-frequency zero verbs. This was carried out to reflect the fact that, before children learn negative prefixes, they have already learned some verbs that undergo no prefixation. After this initial training, the remaining verbs entered the training corpus one by one. The order in which they entered training was determined by a weighted random selection process. The weighting was based on the type frequency (*un-*, *dis-* and zero) and the token frequency of the verbs. The token frequencies of the zero verbs were rank ordered on a scale of 1 to 5, according to the K&F norms. A verb was assigned a rank of 1 if its K&F frequency count was 20 or below, and a rank of 5 if the count was above 500; ranks of 2–4 had K&F counts of 21–50, 51–200 and 201–500 respectively. The token frequencies of the *un-* and *dis-* verbs were randomly rank ordered, since most of these verbs were of low frequency in the K&F counts (simply because K&F counts were based on written data). When fed into the network for training, a verb was repeated a given number of times, according to its frequency rank as calculated above.

2.1.4. Data analysis. We analyzed the network's performance by examining the activation of the three output units (*un-*, *dis-* or zero). The root mean squared error (rms) score was used to determine the match between the network's output pattern and the predefined target pattern. If the rms fell below 0.25, then the output was deemed to be adequate as the correct target. This criterion is roughly equivalent to the activation of the target unit at or above 0.8, with the sum of the activations of the incorrect units not exceeding 0.2. All other patterns were considered to be incorrect.

The network's performance was also assessed at regular intervals with hierarchical clustering analysis (see Elman (1990) for application of this method in network analysis). This technique allows us to discover the relative semantic distances between different verbs that the network represents at the hidden-unit layer, across various time points during learning. Earlier, we used this method to study the semantic judgement results. In what follows, we will use this method to analyze the hidden-unit activation patterns, to determine if the network has developed meaningful and structured representations about the input–output relationships.

2.2. *Results and Discussion*

Figure 1 presents the percentages of the verbs that have been learned correctly after the initial training, across the vocabulary expansion process. The graph is broken down into three parts (Figures 1(a)–(c)) by the different prefixation types (i.e. *un-*, *dis-* and zero). One can immediately observe that the network quite ably learned both the *un-* and the zero verbs but failed to learn the *dis-* verbs. Although the performance with the *un-* and the zero verbs far exceeded that of the *dis-* verbs, the network still failed to learn some of the *un-* and zero verbs, even when all words entered training at the end (24% errors for *un-* and 26% for zero). This failure reflected the network's inability to recover from overgeneralization errors—a point to which we will return shortly.

The network acquired a distinct mapping for the *un-* verbs, by identifying covert semantic categories (i.e. cryptotypes) inherent in these verbs. The cluster analysis of the hidden units given in Figure 2 reveals that, by the 50-word level, the network had formed a structured internal representation. It should be noted that the capitalized marker after each verb in the figure is a mnemonic for the prefixation pattern of each verb; the actual input consists of the base verbs without any prefixes. In Figure 2, verbs that are closely related in meaning are grouped lower in the tree,[2] while clusters of verbs that are similar to other clusters are connected higher in the tree. In addition, we can observe two general clusters in this tree: one cluster for the *un-* verbs, and the other cluster for the zero verbs. If the network had not developed meaningful structures, then we would not expect to find meaningful clusters in the tree.

In Figure 2, most of the verbs in the *un-* cluster share the cryptotypic meaning of binding or locking: for example, *bind*, *chain*, *fasten*, *hitch*, *hook* and *latch*. The network's representation of this meaning was so strong that synonymous verbs in the other categories were also included in the *un-* cluster (such as *hold* and *mount*); hence, overgeneralizations of *un-* on these verbs. Clearly, these synonymous verbs were included because of their semantic similarity with the cryptotype. Because vocabulary growth was incremental, not all cryptotypic meanings were identified at the same time. Instead, mini-cryptotypes emerged at different times, depending on

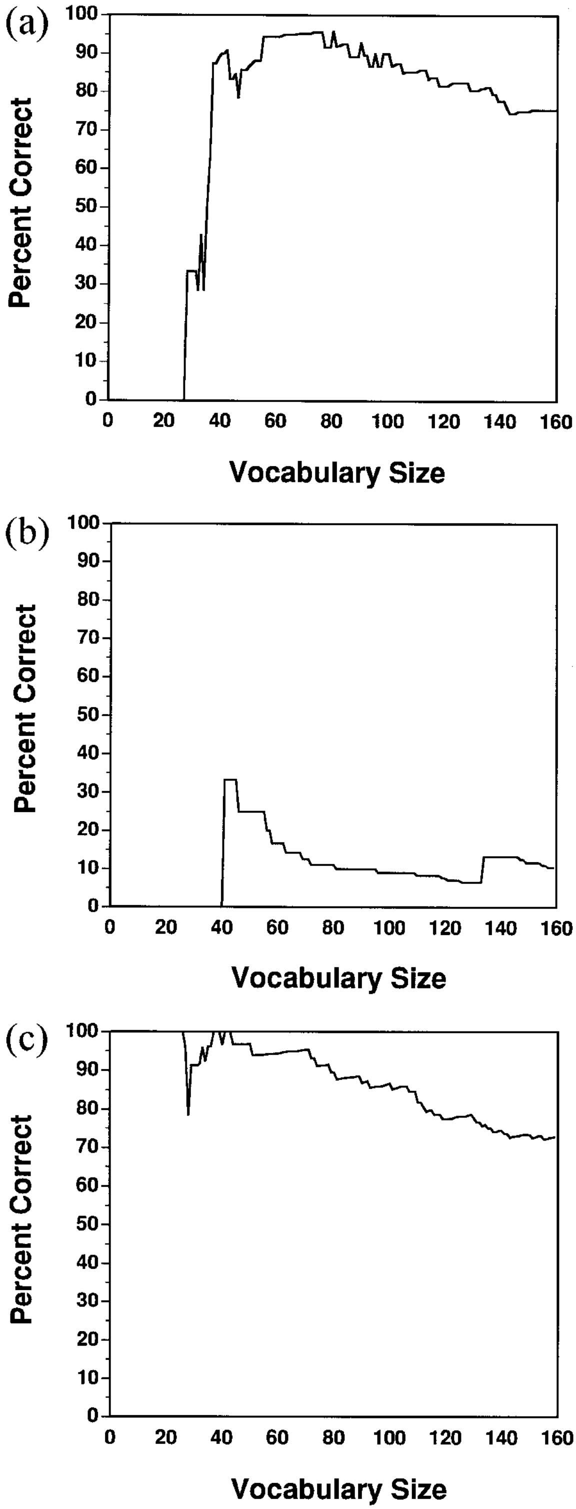

Figure 1. Per cent correct of prefixation as a function of vocabulary increase for (a) *un-* verbs, (b) *dis-* verbs and (c) zero verbs in Simulation 1.

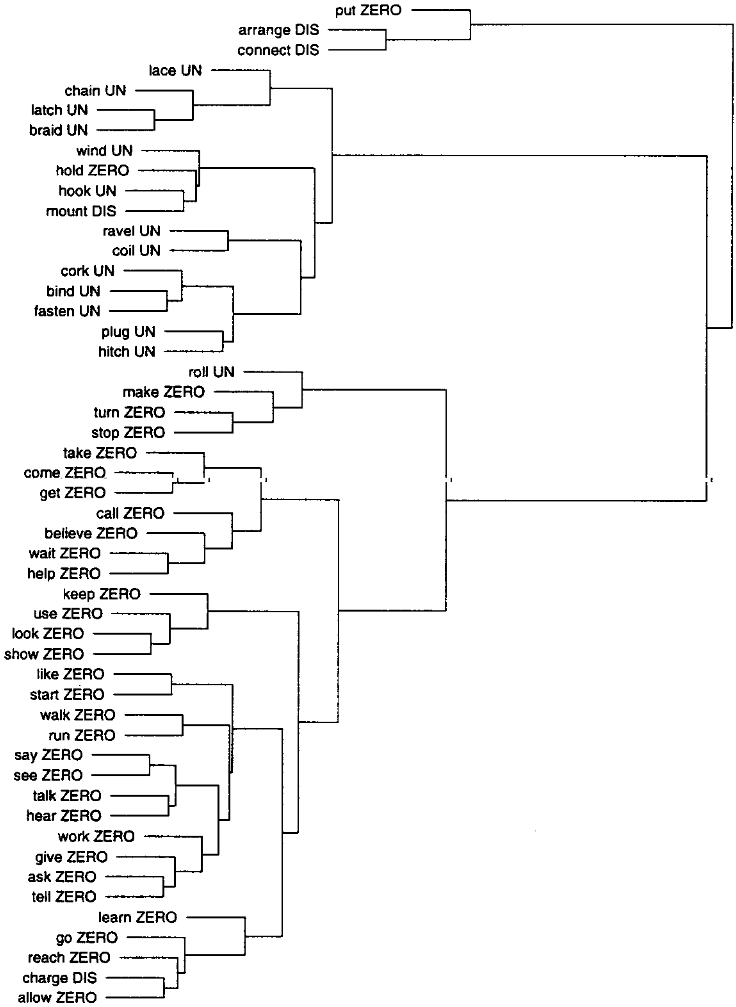

Figure 2. Hierarchical cluster tree of the network's representation of the semantic structure of verbs at the 50-word boundary in Simulation 1.

what words the network had learned at a given time. Figure 2 shows, for example, that the network had not yet developed a clear representation for the ''enclosing'' verbs that involve circular movements. The verbs *ravel* and *coil* were correctly categorized into the *un*- cluster, but the verb *roll* was incorrectly treated as a zero verb.

The network received no discrete label of the semantic category associated with *un*-, and there was no single categorical feature telling which verb should take which prefix. All the network received was semantic featural information distributed over different input patterns. Over time, however, the network was able to develop a structured representation for the mini-cryptotypes in the input–output mapping process. The structured representations in the network emerged as a function of its learning of the association between form and meaning, and not as a property that was given to the network by the modeller. An important implication of this result is that children, in learning to use the reversive prefix *un*-, also abstract the semantic regularities in the *un-* verbs through combinatory restrictions that the prefix places on these verbs (see Bowerman, 1982, 1983). The children are not learning a rule in this case, because the rule itself is not clear; in Whorf's words, the rule is 'subtle' and 'intangible'.

In contrast to the case with *un-*, the network developed no clear representation for the *dis-* verbs. Three of the four *dis-* verbs trained during the first 50 words were clustered with the *un-* verbs, and one with the zero verbs. The reasons for the network's inability to learn the *dis-* verbs are as follows:

(1) The network had seen only a few *dis-* verbs up to this point.
(2) The *dis-* verbs entered the learning process only sporadically (as a result of weighted random selection).
(3) The *dis-* verbs do not have a semantic structure as unified as that of the *un-* verbs.

How does a structured representation of the *un-* cryptotype influence the network's learning of reversive prefixes? Empirical research in child language indicates that there are two possible roles for the cryptotype. The first function of the cryptotype is to overcome overgeneralizations made at an earlier stage, if these overgeneralizations involve verbs that fall outside the cryptotype, such as **uncome*, **unhate* and **untake* (Bowerman, 1982). The second function of the cryptotype is to induce new errors. This occurs because, once children have identified the cryptotype, they will overgeneralize *un-* to all verbs that fit the cryptotype, irrespective of whether or not the adult form actually allows *un-* prefixation.

Our simulation results provide particular support for the second role of a cryptotype in leading to overgeneralizations. There were no simulated errors that constituted flagrant violations of the *un-* cryptotype, such as the forms **uncome* or **unhate* reported by Bowerman (1982). All overgeneralized verbs remained within the scope of the cryptotype. Overgeneralizations occurred after the network had developed some structured cryptotypic representation, including (in order of occurrence): **unhold*, **unpress*, **unfill*, **uncapture*, **unsqueeze*, **unfreeze*, **untighten*, **untack*, **unbury*, **unplant*, **unpeel* and **ungrip*. These results matched up very well with available empirical data. For example, errors such as **unbury*, **uncapture*, **unpeel*, **unpress*, **unsqueeze* and **untighten* all appeared in Bowerman's (1982) data. Other simulated errors, such as **unsplit*, **unmelt*, **unloosen* and **unstrip* reflect typical cases of children's overmarking of *un-* with verbs whose base form already indicates the reversal of the cryptotypic meaning, such as **unopen*, **unsplit* or **unapart* (see Bowerman, 1982; Clark *et al.*, 1995).

One of the two children discussed in Bowerman (1982) displayed the same patterns as those simulated in the network. The overgeneralizations that the child produced all fell into the cryptotype, and her acquisition of *un-* as a reversive prefix is closely associated with her discovery of the cryptotypic meanings of the *un-* verbs. In Clark *et al.*'s (1995) naturalistic data, the child's innovative uses of *un-* also respected the cryptotype from the beginning. Clark *et al.* noted that the child's use of *un-* matched the semantic characteristics of the cryptotype, even when the conventional meanings of the verb in the adult language did not: **unbuild* was used to describe the action of detaching lego-blocks; **undisappear* was used to describe the releasing of the child's thumbs from inside his fists. The child seemed to have recognized that *un-* marks the reversal of actions and that it can do so only for certain kinds of action (i.e. actions that fit the cryptotypic characteristics of binding, covering, enclosing and attaching).

Another child in Bowerman's (1982) data displayed a different pattern. She started to use *un*- productively before she recognized the cryptotype associated with the *un*- verbs. Only later on did she restrict the use of *un*- to verbs that fit the cryptotype, in which case the cryptotype helped her to recover from earlier errors. However, a detailed examination of the child's early errors revealed that she used *un*- in those cases to mean 'stop doing something', rather than the reversal of an action, such as in **uncome* and **unhate*. This 'stop doing *X*' meaning of *un*- could be a precursor of the reversive meaning of *un*-, and it is likely that the child came to recognize the reversive meaning of *un*- simultaneously as she recognized the cryptotypic meanings of the verbs. As Clark *et al.* (1995) pointed out, children can express the notion of reversal long before they have acquired the prefix *un*-, relying on negative particles (such as *off* and *back*) or general-purpose verbs (such as *open* or *undo*). Therefore, it is natural that, when learning to use *un*-, they pay attention to the cryptotypic constraints that *un*- places on the verb in terms of telicity, accomplishment and other features. This explanation is compatible with Clark's (1987) 'principle of contrast' or MacWhinney's (1989) 'principle of competition', which states that children tend to assign different functions to distinct forms.

The role of the cryptotype in inducing overgeneralizations can also be observed with the *dis*- verbs. Earlier, we observed that the network cannot learn the *dis*- verbs, as a result of the absence of a distinct cryptotype for these verbs, as well as the way in which they entered training. Interestingly, the network overgeneralized *un*- to a number of *dis*- verbs that shared the cryptotypic meaning of *un*-, producing errors such as **unassemble*, **unentangle*, **unmount* and **ununite*, all of which, in the adult language, should be prefixed with *dis*- instead of *un*-. These *dis*- verbs all entered the learning process when the network had already constructed a clear representation of the *un*- cryptotype. Although overgeneralizations on *dis*- verbs are rare in children's speech, the results show that, once the system starts to overgeneralize on the basis of the cryptotype, it does so to all the verbs that share the semantic characteristics of the cryptotype.

To summarize, this simulation has shown that the network exhibits learning patterns that closely resemble those of a human child. The model learns to extract the shared aspects of the semantic properties associated with *un*-; builds a structured representation of the semantic cryptotype; and uses this representation as a basis for productive and innovative use of the negative prefixes. Overgeneralization of *un*- is simply a result of such productive and innovative uses. The results also indicate that our network can use a distributed input to extract an internalized structured representation that expresses the 'subtle' or 'intangible' aspects of cryptotypes.

Simulation 1 suffered from two major mismatches to the empirical data and the network performance could not improve after even prolonged continuous training (an additional 500 epochs). First, the network was unable to learn the *dis*- verbs. However, children are able to learn these verbs. Second, the network was unable to recover from overgeneralizations that involved verbs that fall within the range of the semantic cryptotype (see Figure 1(c)). However, children are eventually able to restrict these overgeneralizations. In both cases, children can probably rely on verb-by-verb learning of the type discussed in MacWhinney (1996a,b) to rein in their overgeneralizations and to pick up individual verbs with *dis*-.

There are several possible reasons why our network did not show this type of learning ability. First, it is possible that the feature coding system that we have used in this particular simulation provides few resources for this type of verb-by-verb restriction. A richer coding may facilitate a greater separation between individual verbs. Second, it is possible that the absence of a phonological code made it difficult for the network to treat individual verbs differently. Third, it is possible that the difficulties that the network experienced in learning *dis-* verbs were due to the shape of the input training corpus. In the next simulation, we explore this final possibility. We discuss the first two possibilities later.

3. Simulation 2

To examine one possible cause of the difficulties that the network had in learning to use *dis-*, we modified the procedure by which input data entered the learning process. Simulation 2 was based on this new input procedure.

3.1. Method

The input data, the network architecture and the data analysis methods used in this simulation were identical to those in Simulation 1. The task and procedures were modified in the following ways.

First, we hypothesized that the intermittent nature of the training of *dis-* verbs in Simulation 1 (as a result of weighted random selection) made it impossible for the network to extract a semantic representation for this group of verbs. In this second simulation, we introduced the *dis-* verbs into training in a more focused manner, such that all *dis-* verbs were exposed to the network within the first 35% (i.e. 56 words) of the data set. This training schedule was intended to reflect the fact that children typically learn *dis-* and *un-* verbs separately (most *dis-* verbs actually occur later than *un-* verbs in children's speech, as in the CHILDES database, for example). The other types of verb entered training as in Simulation 1. Although this training schedule is less realistic than that used in Simulation 1, it allows us to assess the effects of focused training on the competition between the two ways of marking reversives.

Second, we wanted to see if the network could learn to use *un-* without immediately organizing its cryptotype. In the first simulation, the network identified the cryptotype for *un-* quite early and easily. We hypothesize that, if the network had more difficulty in forming cryptotypes, then perhaps it would initially overgeneralize *un-* outside the cryptotype, as was done by at least one of the children studied by Bowerman (1982). Identification of the *un-* cryptotype was made more difficult by introducing at the beginning a few *un-* verbs that do not clearly belong to the cryptotype (such as *undo*, *undelete* or *unscramble*); introducing at the beginning several *dis-* verbs that share the meaning of the cryptotype (such as *disassemble*, *disentangle* or *dismount*); and introducing all three types of verbs into training from the beginnning, without the initial training of zero verbs only. Table II presents the vocabulary increase process in Simulation 2.

3.2. Results and Discussion

Figure 3 presents the percentages of the verbs that have been learned correctly as a function of vocabulary increase, broken down by the different prefixation types (i.e. *un-*, *dis-* and zero). Three major results emerged from these data. First, unlike in Simulation 1, most of the *dis-* verbs (79%) were learned correctly by the 60-vocabulary-item mark, after which point learning tended to level off. Second, unlike in Simulation 1, a large number of zero verbs and *un-* verbs were mapped

Table II. Vocabulary structure across prefixation patterns in Simulation 2

Total	*dis-*	*un-*	Zero
20	6	8	6
40	13	14	13
60	19	21	20
80	19	32	29
100	19	42	39
120	19	49	52
140	19	49	72
160	19	49	92

incorrectly, even after all the words had entered training. Third, as in Simulation 1, the network could not recover from overgeneralizations of *un-* once it committed these errors.

Earlier, we identified several possible sources of the network's failure in learning the *dis-* verbs. These included the interspersed training procedure and the lack of semantic coherence of the *dis-* verbs. The results from this simulation showed that, if the *dis-* verbs entered training early on in a more focused manner, then the network was indeed able to learn these words.

In Simulation 1, the network overgeneralized *un-* to *dis-* verbs, based on their match to the cryptotype: for example, **unassemble*, **unentangle*, **unmount* and **ununite*. In Simulation 2, the network did not make such overgeneralizations, because these verbs had entered training at an early stage, before the system had developed a representation of the cryptotype for *un-*. To see how this works, let us consider that, in Simulation 1, *assemble* entered training as one of the last items in the vocabulary expansion process, when the network had already developed a firm representation of the cryptotype. In contrast, in Simulation 2, *assemble* was learned as a *dis-* verb shortly after it entered training as the third item, before any representation for the cryptotype could be formed; hence, no chance for overgeneralization of *un-*. The same account applies to verbs such as *entangle*, *mount* and *unite*. These results once again indicated the role of the cryptotype in the network's ability to overgeneralize.

A second major finding from Simulation 2 is that the revised training procedure led to problems in learning many zero and *un-* verbs. In Simulation 1, the network correctly learned 74% of the zero verbs and 76% of the *un-* verbs. In Simulation 2, it learned only 25% of the zero verbs and 51% of the *un-* verbs. Most of the errors involved mapping into incorrect categories; a smaller number involved low activations on all three output units.

Two major forces contributed to the network's poor performance on the zero verbs. First, as in Simulation 1, the majority of the errors still resulted from the network's overgeneralizations of *un-* verbs based on the cryptotype. Thus, simulated errors included **unbury*, **unclose*, **unfill*, **unfreeze*, **unhang*, **unhold*, **unloosen*, **unmake*, **unmelt*, **unopen*, **unpat*, **unpeel*, **unplant*, **unpress*, **unsqueeze*, **unstrip*, **untack* and **untighten*, which all fit the *un-* cryptotype, as in Simulation 1. Second, because there was no initial training in Simulation 2, the zero verbs entered training simultaneously with the *dis-* and the *un-* verbs, so were competing directly with them for distinct mapping from the outset. The focused training on the *dis-* verbs was actually disadvantageous to the zero verbs. As a

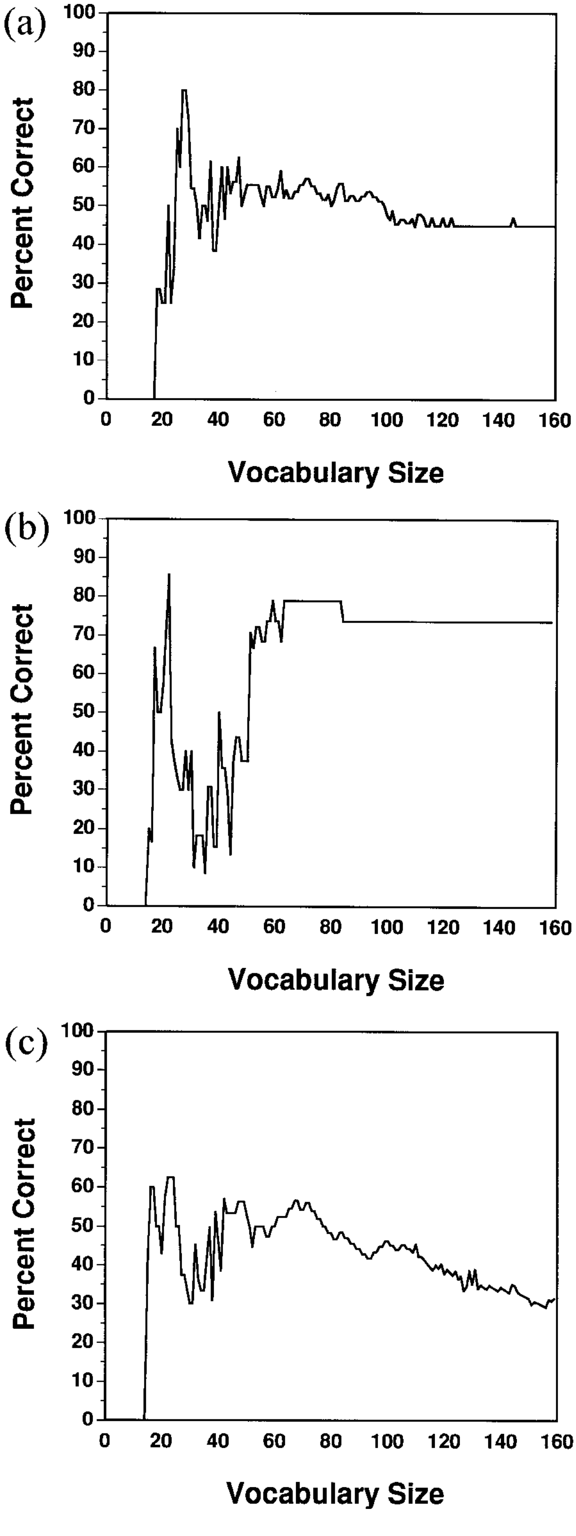

Figure 3. Per cent correct of prefixation as a function of vocabulary increase for (a) *un-* verbs, (b) *dis-* verbs and (c) zero verbs in Simulation 2.

result, 13% of the errors with the zero verbs were overgeneralizations of *dis-*, including the errors **dislearn*, **disinvite*, **display*, **distalk* and **diswrite*, all of which involve mental or cognitive activities. This pattern of overgeneralization indicates that the network tended to pick up a cryptotype for the *dis-* verbs that involved features for mental or cognitive activities. Such learning is a reasonable generalization from exposure to verbs such as *disaffiliate*, *disengage*, *disentangle*, *disintegrate*, *disprove*, *distrust* and *disunite*.

The change of the input also led to more overgeneralizations of *dis-* to *un-* verbs. Of the 49% errors with *un-*, 39% resulted from overgeneralizations of *dis-* or the inappropriate high activation of the *dis-* output unit. The network produced errors such as **disbraid*, **dischain*, **disclench*, **discoil*, **dislace*, **disleash*, **disscrew* and **dissnap*, all of which should be prefixed with *un-* in the adult language. These results suggest that the network first extracted the cryptotypic meanings of attaching, enclosing and binding from the *dis-* words (such as *disassemble*, *disconnect*, *disengage*, *disentangle*, *dismantle* and *disunite*), which it then used as a basis for overgeneralization to the *un-* verbs. This is the opposite of what we observed in Simulation 1, where the network overgeneralized *un-* to *dis-* verbs on the basis of semantic abstraction.

In this simulation, we modified the initial input sequences to make it more difficult for the network to discover the *un-* cryptotype. We hypothesized that, if the initial input is less favourable for the extraction of the *un-* cryptotype, then the network might overgeneralize *un-* before it could develop a committed semantic representation of the cryptotype. This did not happen, however. The initial input differences to the network did not affect its performance across Simulations 1 and 2. The results indicate that changes in the structure of the input delayed both the extraction of a cryptotype and overgeneralization, so providing support to the direct link between the growth of the strength of the prefix and the abstraction of its underlying conceptual structure, as found in Simulation 1.

Finally, as in Simulation 1, the network did not recover from overgeneralization errors with *un-*, and continued to produce high error rates for zero verbs, even at the end. Even after prolonged continuous training (an additional 500 epochs), the network did not converge on the correct mapping patterns. The network seemed to have discovered the relationship between the semantic properties of the verbs and their prefixation pattern (such as the cryptotype and *un-*), and settled on a firm mapping structure, by the time it started to overgeneralize.

To summarize, in Simulation 2, we have shown that the network can learn the *dis-* verbs, provided that these verbs are presented in sufficient quantity early in training. Early presentation of this less-productive prefix allows it to compete more effectively with the other reversive marking options. The cost of this early presentation is that the network overgeneralizes *dis-* to many zero and *un-* verbs. In agreement with Simulation 1, the results provide evidence for the role of the cryptotype in the acquisition of reversive prefixes. A comparison of Simulations 1 and 2 shows that overgeneralization depends on whether or not the network has developed a semantic representation of the cryptotype. For example, when the *dis-* verbs enter training after the development of a cryptotypic representation, the network overgeneralizes (for example, **unassemble* and **unmount*), as in Simulation 1; if they enter training before the cryptotype is recognized, then no overgeneralization occurs, as in Simulation 2. However, in Simulation 2, the network was still unable to recover from errors, once it overgeneralizes according to the cryptotype. Therefore, we deal further with the recovery problem in the next simulation.

Table III. Vocabulary structure across prefixation patterns in Simulation 3

Total	Zero	*un-*
20	11	9
40	21	19
60	32	28
80	42	38
100	49	51
120	49	71
141	49	92

4. Simulation 3

In the first two simulations, the network was unable to recover from overgeneralizations, despite repeated training. In this simulation, to study the effects of learning and recovery more directly, we removed the *dis-* verbs from training and exposed the network only to the *un-* and the zero verbs.

4.1. *Method*

The input data, the network architecture, the task and procedure, and the data analysis methods in Simulation 3 were the same as in Simulation 2. The only difference was that the *dis-* verbs were removed from training in this simulation. Table III presents the vocabulary increase process in Simulation 3.

4.2. *Results and Discussion*

Figures 4(a) and (b) present the percentages of the verbs that have been learned correctly for the *un-* and the zero verbs, respectively. Comparing these data with those from Simulation 2, we can see that the error rates for the zero verbs and for the *un-* verbs were significantly reduced. However, as in Simulations 1 and 2, the network continued to overgeneralize *un-* to zero verbs, based on the semantic cryptotype.

In this simulation, we found that the network could recover from overgeneralizations with a number of words, including **unbend*, **unbury*, **unhang* and **unsqueeze*, all of which appeared in Bowerman (1982). After the initial errors, these verbs began to resist *un-* prefixation as training continued. Incidentally, all the words that showed recovery were those that had entered the learning process at a very early stage. Apparently, the additional experience involved early on in coding the semantics of these verbs in distinction with other verbs gave them a greater distinctiveness in the network, which served to facilitate resistance to overgeneralization. A cluster analysis of the network's representation at the hidden layer revealed that, fairly early in training, the network had identified a structure of the *un-* cryptotype. Figure 5 presents the cluster tree of the network's representation at the 24-word boundary (*bury* entered learning as the 13th word, *bend* as the 15th word, *hang* as the 20th word and *squeeze* as the 24th word). As in Figure 2, the capitalized marker after each verb in Figure 5 is a mnemonic for the prefixation pattern of each verb; the actual input consists of the base verbs without any prefixes.

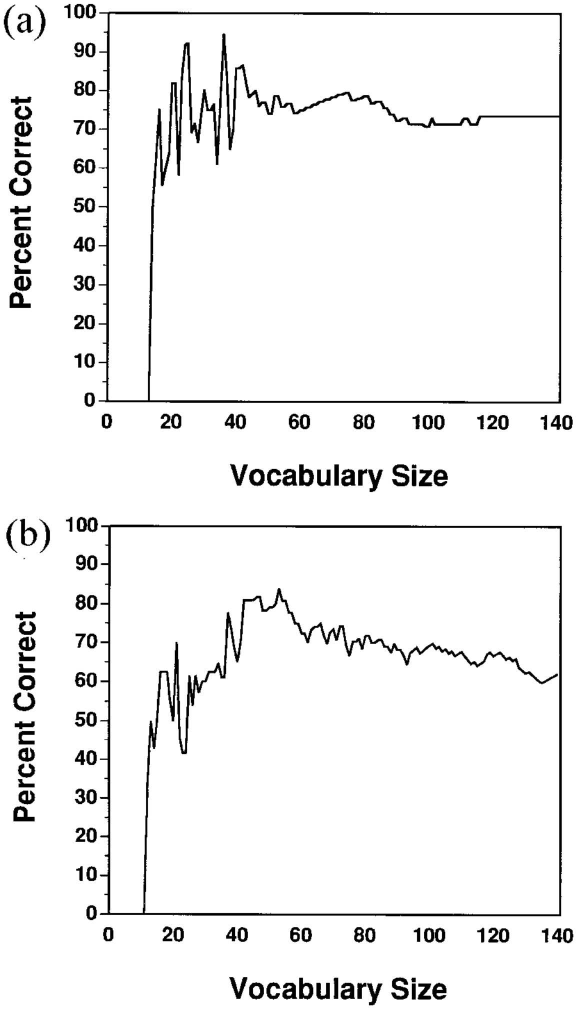

Figure 4. Per cent of prefixation as a function of vocabulary increase for (a) *un*-verbs and (b) zero verbs in Simulation 3.

As shown in Figure 5, the network developed separate representations for the *un*- and zero verbs at this point, although some zero verbs were miscategorized into the *un*- category (the overgeneralized forms). However, two factors make the representation incomplete and unstable. First, the category members in the zero clusters did not form clear semantic coherence; for example, *see* and *turn* were grouped together, but *come* and *go* were not. Second, although some verbs in the *un*- cluster were associated with the 'attaching' meaning (such as *link*, *tangle* and *plug*), others were not (such as *scramble*, *settle* and *do*). Compared with the representation in Figure 2, the network's representation of the *un*- cryptotype is only partial at this point. As a result, the overgeneralizations with zero verbs were based on the network's partial rather than stable and clear structure of the cryptotype (unlike the situation in Simulation 1, as revealed in Figure 2). Given such a structure, the network had a chance to recover from the overgeneralization errors. In contrast, in the previous simulations, the network did not show any signs

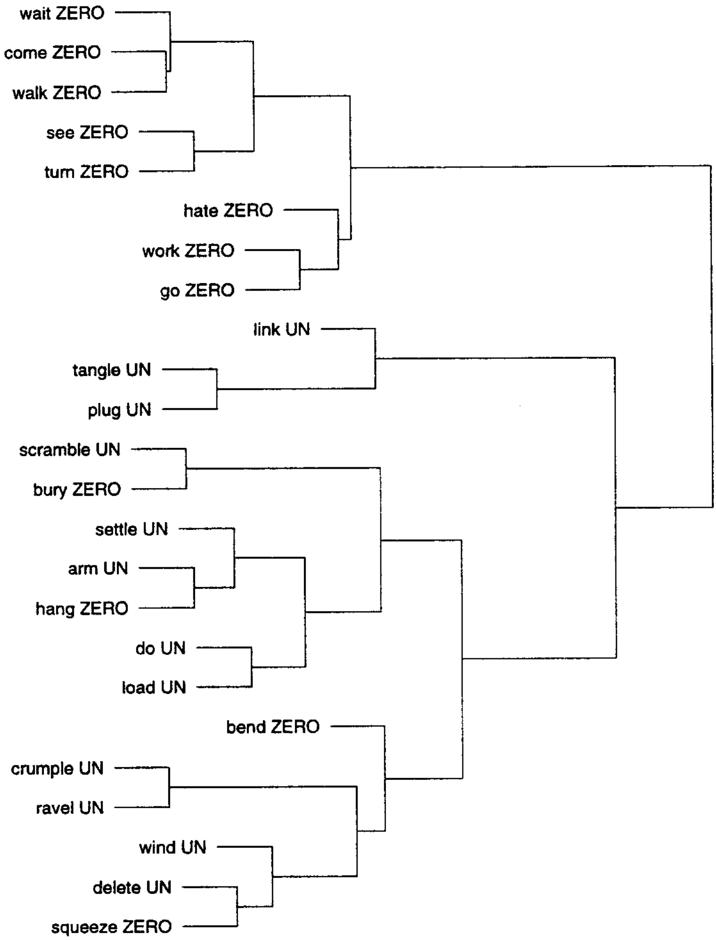

Figure 5. Hierarchical cluster tree of the network's representation of the semantic structure of verbs at the 24-word boundary in Simulation 3.

of recovery, probably because it had settled on a stable structure by the time it started to overgeneralize.

In this simulation, there were also many words that could not recover from overgeneralizations, including **unclose*, **undetach*, **unfill*, **unpress*, **unstrip* and **untighten*, all of which fit the cryptotype. Interestingly, these words were learned at a later stage, when the network already settled on a firm semantic structure for the cryptotype, as revealed by analyses of the network's hidden-unit activation patterns.

The discrepancy between errors that permit recovery and those that do not provides some evidence on the time course of the network's learning ability. Early in learning, the network has not built a complete or stable representation for the semantic cryptotype associated with *un-*. Overgeneralization errors at this stage have a chance to be corrected, because the network is still flexible or 'plastic' enough to adjust its error space in large sweeps. Later on, as the network learns more words, and as it settles on some stable semantic representations on which overgeneralizations are based, the network becomes increasingly inflexible and unable to make radical adjustments in the weight space. In other words, the network becomes entrenched in a state of weight configurations that makes further changes impossible (see Elman (1993, pp. 91–93) for a detailed discussion of how the learning algorithm determines weight adjustment over time). Thus, 'plasticity' is a property characteristic of early stages of learning, as in young children, while 'stability' is characteristic of later stages of learning.

To verify our analysis of the early versus late differences in the network's ability to recover, we retrained the same network with the following changes in training schedule. We exchanged the verbs that could recover from errors with the verbs that could not, in the order in which they entered the training process. Specifically, the verbs *bend*, *bury*, *hang* and *squeeze* that occurred early in the original training were moved to a later stage of learning in the new training, whereas verbs such as *close*, *detach*, *fill*, *press*, *strip* and *tighten* were now moved to the early stage. The result from this new round of training is informative. All the verbs that could not recover in the original training could now recover from the overgeneralizations after their initial errors. However, the verbs that recovered from overgeneralizations in the original training could no longer recover in this new training, because the overgeneralization errors were now based on a stable structure of the semantic cryptotype that the network built at the later stage of learning.

To summarize, the data from Simulation 3 provide converging evidence on the role of semantic cryptotypes in overgeneralization and recovery during the network's acquisition of the *un-* reversive prefix. The results are consistent with analyses of patterns of network learning in the domains of syntactic structures (Elman, 1993) and past-tense acquisition (Marchman, 1993). Those analyses, together with our results, paint a general picture of the early plasticity and late stability both in network's and in children's learning. Our results provide further evidence on the role of plasticity and stability in overgeneralization and recovery in language acquisition.

5. General Discussion

This study was designed to evaluate the detailed semantic basis for the overgeneralization and recovery processes in language acquisition. We wanted to relate these processes to semantic support within cryptotypes and competition between alternative devices. At the same time, we wanted to demonstrate ways in which connectionist models can benefit from a more complete semantic grounding. To achieve these goals, we built a network to map semantic input features to three prefixation patterns in an incremental learning schedule. We conducted three sets of simulations to examine the role of the semantic cryptotype associated with the use of reversive prefixes.

In Simulation 1, the network constructs a representation of a cryptotype for verbs that take *un-*, and it then uses this representation as a basis for productive and innovative use of the reversive prefixes. In Simulation 2, the network develops structured representations of the semantic cryptotype that underlies both *un-* and *dis-*. Results from these two simulations suggest that the role of the cryptotype is to induce overgeneralizations to verbs that fall within the realm of the cryptotype, rather than being to reduce overgeneralizations outside the cryptotype. Simulation 2 also highlights the role of lexical competition in morphological acquisition. It shows that direct competition between similar inputs (such as verbs that share the semantic cryptotype) leads to learning difficulties for dissimilar output (i.e. *dis-* versus *un-*). In Simulation 3, the network exhibits a higher level of early plasticity that promotes recovery from overgeneralization errors. Together, these results give us a picture of the learning dynamics of a network system that extracts semantic information, develops cryptotypic semantic representations, overgeneralizes competing morphological devices and eventually recovers from errors (see also Li, 1993). These results help us to understand how and why learners overgeneralize and recover from overgeneralizations.

Our analyses also provide a more precise account of how Whorf's cryptotype emerges from distributed representations, and how it affects overgeneralization and recovery. Whorf described the meaning of the cryptotype as 'subtle' and 'intangible', but he also noted that native speakers control it intuitively. Whorf did not explain how this could be true. In our view, the reason for the intangibility of the cryptotype is that the semantic features that unite different members of a cryptotype are represented in a complex distributed fashion (such as features overlap across categories), such that they are not easily subject to explicit symbolic analysis, but are accessible to native intuition. Therefore, the meaning of the cryptotypic members itself is not intangible, but the semantic relationships between the members are intangible, because they are inaccessible to symbolic analysis. Gender and classifier systems are further illustrations of the intangibility of such distributed patterns (Lakoff, 1987; MacWhinney, 1989; Köpcke & Zubin, 1984).

The meaning of a cryptotype constitutes a complex semantic network in which verbs differ from one another with respect to the following: how many features each verb contains; how strongly each feature is represented in the verb; and how strongly features within verbs are related to one another (all true with the input to our network). It is the relationships between the features that give rise to cryptotypes. For the child, learning reversive prefixes is not the learning of a symbolic rule for the use of a prefix with a class of verbs, but the learning of the connection strengths that hold between a particular prefix and a complex distributed set of semantic features across verbs. The learning system groups together those verbs that share the largest number of features and take the same prefixation patterns. Over time, the verbs gradually form clustered structures, with respect to both meanings and prefixation patterns.

Results from the present study also shed light on the role of the plasticity and stability of network learning in the network's ability to recover from overgeneralizations. If the network develops a firm structure of the cryptotype by the time it overgeneralizes, then it has little chance to recover from the errors. However, if the overgeneralizations are only based on partial or unstable structures that are present during a period when the weight space is not fully committed, then recovery is possible.

We should mention at least three factors that may have led to an underestimation of recovery capacities in our simulations.

(1) Our model has included only semantic representations of verbs. A model that also codes for phonological differences between verbs would probably aid the system's recovery from overgeneralizations (such as the use of phonological distinctive features in MacWhinney and Leinbach (1991)).
(2) It is possible that a richer, more distinctive semantic coding for our verbs would further aid in helping individual verbs resist overgeneralization and recover from overgeneralization errors.
(3) There is good reason to believe (MacWhinney, 1996b) that older language learners have access to a variety of secondary rehearsal and organizational but not our network systems that they can use to retune lower-level connections.

Each of these additional forces can be explored in further connectionist models and we have already begun to follow up each of these issues in work currently in progress (Li & MacWhinney, 1996). Eventually, consideration of these additional factors—and possibly others—will allow us to understand better the detailed mechanism of the processes of overgeneralization, recovery and competition across a wider variety of semantic fields.

Acknowledgements

This research was supported by a Direct Grant for Research from the Chinese University of Hong Kong. We are very grateful to Elizabeth Bates, Melissa Bowerman, Jeffrey Elman and members of the PDPNLP group at University of California, San Diego, for their insightful discussions during various phases of this study. Special thanks go to Cathy Harris for her constructive discussions on the relevant semantic features selected for this study. We would also like to thank Hong Li for running and analyzing the simulations; and Edward Yang for running the experiments with human subjects.

Notes

1. The verbal prefix *un-* should be carefully distinguished from *un-* with adjectives such as *unsure*: the former indicates reversal of action, while the latter roughly means 'not' and can attach to almost all adjectives that denote quality or state.
2. There were a few exceptions to this description. For example, *wind* was grouped with *hold*, *hook* and *mount*, and *go* was grouped with *reach*, *charge* and *allow* (rather than *walk* and *run*). These exceptions could result from the incompleteness or instability of the semantic structure that the network develops at an early stage (see more discussion on the instability issue later).

References

Bauer, L. (1983) *English Word-formation.* Cambridge: Cambridge University Press.

Bowerman, M. (1982) Reorganizational processes in lexical and syntactic development. In E. Wanner & L. Gleitman (Eds), *Language Acquisition: The State of the Art.* Cambridge: Cambridge University Press.

Bowerman, M. (1983) Hidden meanings: the role of covert conceptual structures in children's development of language. In D. Rogers & J. Sloboda (Eds), *The Acquisition of Symbolic Skills.* New York: Plenum.

Bowerman, M. (1988) The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? In J. Hawkins (Ed.), *Explaining Language Universals.* New York: Basil Blackwell.

Brown, R. (1973) *A First Language.* Cambridge, MA: Harvard University Press.

Clark, E. (1987) The principle of contrast: a constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition.* Hillsdale, NJ: Lawrence Erlbaum.

Clark, E., Carpenter, K. & Deutsch, W. (1995) Reference states and reversals: undoing actions with verbs. *Journal of Child Language*, **22,** 633–662.

Cottrell, G. & Plunkett, K. (1994) Acquiring the mapping from meanings to sounds. *Connection Science*, **6,** 379–412.

Elman, J. (1990) Finding structure in time. *Cognitive Science*, **14,** 179–211.

Elman, J. (1993) Learning and development in neural networks: The importance of starting small. *Cognition*, **48,** 71–99.

Gropen, J., Pinker, S., Hollander, M. & Goldberg, R. (1992) Affectedness and direct object: the role of lexical semantics in the acquisition of verb argument structure. In B. Levin & S. Pinker (Eds), *Lexical and Conceptual Semantics.* Cambridge, MA: Blackwell.

Hoeffner, J. (1992) Are rules a thing of the past? The acquisition of verbal morphology by an attractor network. *Proceedings of the 14th Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum.

Horn, L.R. (1988) Morphology, pragmatics, and the *un-* verb. In J. Powers & K. de Jong (Eds), *Proceedings of the 5th Eastern States Conference on Linguistics.* Ohio State University.

Köpcke, K.-M. & Zubin, D. (1984) Sechs Prinzipien für die Genuszuweisung im Deutschen: ein Beitrag zur natürlichen Klassifikation. *Linguistische Berichte*, **93,** 26–50.

Kucera, H. & Francis, W. (1967) *Computational Analysis of Present-day American English.* Providence, RI: Brown University Press.

Lakoff, G. (1987) *Women, Fire, and Dangerous Things.* Chicago, IL: Chicago University Press.

Levin, B. (1993) *English Verb Classes and Alternations: A Preliminary Investigation.* Chicago, IL: Chicago University Press.

Li, P. (1993) Cryptotypes, form-meaning mappings, and overgeneralizations. In E.V. Clark (Ed.), *Proceedings of the 24th Child Language Research Forum.* Center for the Study of Language and Information, Stanford University, CA, 162–178.

Li, P. & MacWhinney, B. (1996) In search of the hidden meaning: role of semantic and phonological representations in connectionist morphological learning (in press).

MacWhinney, B. (1989) Competition and lexical categorization. In R. Corrigan, F. Eckman & M. Noonan (Eds), *Linguistic Categorization.* New York: Benjamin Cummings.

MacWhinney, B. (1995) *The CHILDES Project: Tools for Analyzing Talk*, 2nd Edn. Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (1996a) Lexical connectionism. In P. Broeder (Ed.), *Cognitive Approaches to Language Learning.* Cambridge, MA: MIT Press.

MacWhinney, B. (1996b) Language specific prediction in foreign language acquisition. *Language Testing*, **12,** 292–320.

MacWhinney, B. & Leinbach, J. (1991) Implementations are not conceptualizations: Revising the verb learning model. *Cognition*, **40,** 121–157.

MacWhinney, B. & Snow, C. (1985) The Child Language Data Exchange System. *Journal of Child Language*, **12,** 271–296.

MacWhinney, B. & Snow, C. (1990) The Child Language Data Exchange System: an update. *Journal of Child Language*, **17,** 457–472.

Marchman, V. (1993) Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience*, **5,** 215–234.

Marchand, H. (1969) *The Categories and Types of Present-day English Word-formation: A Synchronic-diachronic Approach.* Munich: C.H. Beck'sche.

McClelland, J. & Rumelhart, D. (1981) An interactive activation model of context effects in letter perception: Part 1. An account of the basic findings. *Psychological Review*, **88,** 375–402.

Pinker, S. & Prince, A. (1988) On language and connectionism. In S. Pinker & J. Mehler (Eds), *Connections and Symbols.* Cambridge, MA: MIT Press.

Pinker, S., Lebeaux, D.S. & Frost, L.A. (1987) Productivity and constraints in the acquisition of the passive. *Cognition*, **26,** 195–267.

Plaut, D.C., McClelland, J.L., Seidenberg, M.S. & Patterson, K. (1996) Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, **103,** 1–60.

Plunkett, K. & Marchman, V. (1991) U-shaped learning and frequency effects in a multi-layered perceptron: implications for child language acquisition. *Cognition*, **38,** 43–102.

Plunkett, K. & Marchman, V. (1993) From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition*, **48,** 21–69.

Rumelhart, D. & McClelland, J. (1986) On learning the past tenses of English verbs. In J. McClelland, D. Rumelhart and the PDP research group (Eds), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. II. Cambridge, MA: MIT Press.

Rumelhart, D., Hinton, G. & Williams, R. (1986) Learning internal representations by error propagation. In J. McClelland, D. Rumelhart and the PDP research group (Eds), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. I. Cambridge, MA: MIT Press.

Vendler, Z. (1967) *Linguistics in Philosophy.* Ithaca, NY: Cornell University Press.

Webster's New Collegiate Dictionary, 9th edn. (1990) Springfield, MA: Merriam-Webster.

Whorf, B. (1956) Thinking in primitive communities. In J.B. Carroll (Ed.), *Language, Thought, and Reality.* Cambridge, MA: MIT Press. (Original manuscript from 1936.)

Appendix A: Overgeneralization Errors of *un-* Produced by Children

(1)	D, 2;8	M:	Did someone undo my belt? [after D pulled her belt undone]
		D:	No, no, I *unpulled* it because it wasn't tied yet.
(2)	D, 2;9	D:	It's *unflowing*. [= emptying; opening plug in bidet and letting water out]
(3)	D, 2;9	D:	No, no I was tighting my badge. I tighted my badge, and you should *untight* it. [= loosen; D wanted to take a PanAm badge off his shirt]
(4)	D, 2;10	D:	They've disappeared. [having hidden his thumbs by closing his fists on them]
		M:	Can you make them appear again?
		D:	I can't make it *undisappear*. [= reappear]
(5)	D, 3;0	D:	Oh your hair is *unpinning*. [as a hairpin fell out of M's hair]
(6)	D, 3;1	D:	First I *unbuild* it, okay? [needing to put his blocks in a bag to take upstairs]
(7)	D, 3;4	D:	I don't know what's in my stocking. I'll have to *unhang* it. [= take down; feeling his Christmas stocking]
(8)	D, 3;5	D:	Show me how you *uncatch* your necklace. [= undo the catch]
(9)	D, 3;10	D:	I'm *unpyjama-ing*. [= taking off; having put pyjamas over his clothes, and now proceeding to take them off]
(10)	D, 4;3	D:	Here's Duncan's airplane. D'you want me to *unblow* it? [= deflate]
(11)	D, 4;3	D:	. . . he *unstrings* the worms every day and throws them on the fire. [= taking off the string; telling a story]
(12)	D, 4;5	D:	Maybe it's for *unlighting* the flame . . . a faster way. [= extinguishing; speculating about a small knob on the stove]
(13)	D, 4;6	D:	. . . but the two big kids didn't know that the little one was *unknitting* the wool. [= undoing the knitting]
(14)	C, 3;9	C:	This is pooey that's coming out of here. [in tub, showing cup with water spouting out of the holes] And that's how to make it *uncome*. [blocking holes with hand]
(15)	C, 4;5	M:	I've been using them for straightening the wire. [C has asked M why pliers are on the table]
		C:	And *unstraighting* it? [= unbending]
(16)	C, 4;7	C:	I hate you! And I'll never *unhate* you or nothing!
		M:	You'll never unhate me?
		C:	I'll never like you.
(17)	C, 5;1	M:	Seems like one of these has been shortened, somehow. [M working on strap of C's backpack]
		C:	Then *unshorten* it. [= lengthen]
(18)	C, 5;1	C:	He tippitoed to the graveyard and *unburied* her. [telling ghost story]
(19)	C, 5;1	C:	I *unbended* this with stepping on it. [= straightened; after stepping on a tiny plastic three-dimensional triangular roadsign, squashing the angles out of it]
(20)	C, 6;0	C:	Wait until that *unfuzzes*. [watching freshly poured, foamy coca cola]

(21)	C, 6;11	C:	How do you make it sprinkle? [C trying to figure out how kitchen faucet works] [after getting it to sprinkle] How do you make it *unsprinkle*?
(22)	C, 7;11	C:	I'm gonna *unhang* it. [taking stocking down from fireplace]
(23)	C, 4;9	C:	You can take it *unapart* and put it back together. [C manipulating a take-apart toy. Here un- has migrated to the wrong part of speech]
(24)	C, 4;11	C:	Will you *unopen* this? [wants F to take lid off styrofoam cooler]
(25)	C, 5;6	C:	. . . *unpatting* it down. [as C pats ball of ground meat into hamburger patty]
(26)	E, 3;2	E:	Why did you *unclothes* her? [M has taken C's clothes off]
		M:	Why did I what?
		E:	Why did you *unclothes* her?
		M:	Why did I what?
		E:	Um . . . why did you take her clothes off?
(27)	E, 3;2	E:	I can't *untight*. [= loosen; E struggling with tight overall strap]
(28)	E, 3;10	M:	I have to capture you. [grabbing E in game]
		E:	*Uncapture* me! [trying to pull loose]
(29)	E, 3;11	E:	How do you *unsqueeze* it? [C coming to M with clip earring hanging from ear]
		M:	What?
		E:	How do you *unget* it . . . undone?
(30)	E, 4;7	E:	I know how you take these apart. *Unsplit* them and put 'em on. [holding up chain of glued paper strips]
		M:	How do you unsplit them?
		E:	Like this [pulling a link apart]
(31)	E, 4;7	E:	Will you *unpeel* the banana? [giving banana to M]
(32)	E, 4;11	E:	. . . and then *unpress* it out. [showing M how to get playdough out of a mould]
		M:	How do you unpress it out?
		E:	You just take it out.
(33)	E, 4;7	E:	You slip it across . . . and you *unslip* it like this [showing M how to work clasp on coin purse; as E says 'slip', she moves the two metal parts past each other so that purse closes; as she says 'unslip', she opens it]

Note: D, C and E are the initials of three children's names, and M denotes mother. Ages are given in 'years; months'. The results are taken from Bowerman (1982, 1988) and Clark *et al.* (1995) (reproduced with permission of the authors).

Appendix B: Verbs Used in the Input Corpus

un- *Verbs*

arm	cork	lace	screw
bandage	cover	latch	settle
bind	crumple	leash	sheathe
bolt	curl	link	snap

braid	delete	load	strap
buckle	do	lock	tangle
button	dress	mask	tie
chain	fasten	pack	twist
clasp	fold	plug	veil
clench	hinge	ravel	wind
clog	hitch	reel	wrap
coil	hook	roll	zip
		scramble	

dis- *Verbs*

affiliate	connect	infect	place
appear	continue	integrate	prove
arrange	embark	locate	trust
assemble	engage	mantle	unite
charge	entangle	mount	

Zero Verbs

affect	find	move	slip
agree	free	obey	solve
allow	freeze	open	speak
approve	get	pat	spill
ask	give	pay	split
become	go	peel	sprinkle
begin	grip	plant	squeeze
believe	grow	play	stand
bend	hang	pose	start
break	hate	possess	stop
bring	hear	press	straighten
bury	help	pull	strip
call	hold	put	tack
capture	invite	reach	take
clear	keep	release	talk
close	learn	remove	tell
come	lift	reverse	tighten
confirm	like	run	turn
deprive	live	say	use
detach	look	see	wait
embarrass	loosen	separate	walk
expel	make	show	work
fill	melt	sit	write

Appendix C: Semantic Features and Feature Vectors as Representations of Verbs

Semantic features	*connect* (*dis*-)	*link* (*un*-)	*turn* (zero)
(1) Mental activity	0.1	0.1	0.0
(2) Manipulative action	0.7	0.9	0.6
(3) Circular movement	0.0	0.0	0.6
(4) Change of location	0.1	0.1	0.5
(5) Change of state	0.3	0.3	0.1
(6) Resultative	0.3	0.4	0.3
(7) A affects B	0.2	0.3	0.5
(8) A touches B	0.9	0.9	0.1
(9) A distorts B	0.1	0.0	0.3
(10) A contains B	0.0	0.1	0.1
(11) A hinders B	0.1	0.2	0.1
(12) A obscures B	0.0	0.0	0.0
(13) A surrounds B	0.2	0.3	0.1
(14) A tightly fits into B	0.6	0.7	0.0
(15) A is a salient part of B	0.5	0.7	0.1
(16) A and B are separable	0.6	0.4	0.1
(17) A and B are connectable	0.7	0.8	0.0
(18) A and B are interrelated	0.5	0.6	0.0
(19) A and B are in orderly structure	0.3	0.5	0.0
(20) A and B form a collection	0.6	0.5	0.0

Note: In the semantic judgement experiment, subjects were given more detailed descriptions of these features in sentence format.