Shared lexical representations: evidence from first-language acquisition

Thomas Berg

To Bruce Derwing,
the lord of the linguistic-psychologists

There is a striking inconsistency in how similarity is treated in local connectionist models: phonological and semantic similarity are taken into account, whereas lexical similarity is plainly ignored. In this article, lexical similarity, in particular the ordering of segments within words, is shown to influence linguistic behaviour. A detailed analysis of one child's acquisition of German reveals that the correct pronunciation of /k/ in the same phonological environment (e.g. *Klammer* 'clothes peg' and *klein* 'little') is learned in different words at about the same time. This is interpreted to mean that the same horizontal link has been adjusted in the learning process. To account for identical links between segment tokens from different words, lexical items are claimed to have shared representations. More generally, memory representations are assumed to be superpositional in nature. The shared-representation hypothesis argues against local connectionism in favour of a model which endorses a one-to-many relationship between processing and linguistic units. In the final analysis, fully distributed representations may well be necessary to account for all of the data.*

1. Introduction

Local connectionist models are defined by a one-to-one relationship between linguistic units and processing units. Each bit of information is represented by one and only one mental node. The division of the processing system into various hierarchical levels, and the principle of establishing connections between them, allow for two different modes of representation. Information which is germane to a particular unit may be either *private* or *shared*. Let us take as an example the colour terms *blue* and *brown,* which begin with the same phoneme /b/. It is possible to regard the two /b/'s as private components of the two lexical items, treat them as unrelated, and accordingly reserve separate nodes for them. Alternatively, the two /b/'s may be viewed as identical and assigned to the same phoneme node. In so doing, the /b/ is made to serve both lexical items equally. The words *blue* and *brown* may thus be said to share (a part of) their phonological representation.

Rivista di Linguistica, 16.2 (2004), pp. 325-348 *(ricevuto nel dicembre 2004)*

Stemberger (1985) argued persuasively in favour of shared representations mainly because they create the competition that is characteristic of the human processing system. Hence, local connectionist frameworks typically take the form shown in Figure 1 (e.g. Dell & O' Seaghdha 1992). Semantic information is given in square brackets, lexical information in quotation marks, and phonological information in slashes.

Figure 1. A fragment of a connectionist network consisting of three levels.

The basic architectural principle underlying these models is that similarities among lexical items are expressed as identities at other levels. This principle is applied to the semantic and the phonological level alike. The shared representations of *brown* and *blue* manifest themselves in the identical [colour] node that they connect to at the semantic level as well as in the identical /b/ node that they connect to at the phonological level.

Curiously enough, local connectionist models display a certain inconsistency in adhering to the principle of shared representations. Whereas phonological and semantic information are both shared, lexical information is represented in private fashion: everything that is genuinely lexical is stored with each individual word. Since no other representational levels are assumed, there is no way in which lexical information can be extracted and, if identical across words, represented elsewhere. This representational format introduces a great deal of redundancy and minimizes the connectivity that is so clearly the centrepiece of connectionist models.

It is the purpose of this contribution to question the standard assumption of private representation at the lexical level. If my line of argument is accepted, it will lead to a fundamentally different view of the lexical level; if not, this article should be read as an invitation to take cognizance of, and explicitly defend, this inconsistency in local connectionist modelling. The next two sections detail the type of lexical information for which an adequate representation has to be found. Subsequently, predictions will be derived from the shared-representation and private-representation hypotheses and put to the test. Finally, the theoretical implications of the empirical findings will be considered.

2. Serial order as lexical information

The production of single words involves not only the selection of the phonemes that make up a word, but also a decision on their linear order. As linear-order information may be either lexical or phonological in nature, we have to consider both lexical and phonological ordering schemes. Let us imagine a language in which all words are monosyllabic and all syllables are of the CV type. Would it be necessary for linear-order information to be part of the lexical representation? McCarthy (1989) gave a negative answer to this question. Since the syllable structure stipulates that the consonant precedes the vowel and since there is only one consonant and one vowel, their order is exhaustively determined by phonological principles. The lexical representation may do without serial-order information and consist of an unordered set of two phonemes per word.

It is evident that this simple ordering scheme is incapable of dealing with more complex phonotactic structures. As soon as more than one consonant per syllable is allowed, an additional mechanism has to be invoked which specifies the order in which the two consonants appear relative to each other and to the vowel. The latter task may be accomplished by a CVC syllable template. For the former task, an entirely phonological sequencing mechanism is conceivable. For more than a century, linguists have entertained the idea of a phonological basis for segment order within the syllable (e.g. Sievers 1885/1901, Jespersen 1904). This phonological basis is expressed by the sonority sequencing generalization (e.g. Selkirk 1984, Clements 1990) whereby a specific order is imposed on any number of segments. Each segment (or segment class) is assigned a sonority value on the basis of its phonetic properties. The segments of a syllable are stipulated to form a sonority contour which rises steadily from the beginning to the middle of the syllable and falls steadily from the middle to the end. It is additionally assumed in recent theories that the rise is steeper than the fall. On this basis, the two segments /k/

and /l/ would be unambiguously assigned to the onset and the coda of a CVC syllable, respectively, rather than vice versa, because the lateral is more sonorous than the stop. In a CCV structure, the onset cluster would invariably be /kl/, never /lk/.

The sonority sequencing principle accounts for quite a few serial-ordering relations on a purely phonological basis (compare Anderson 1987). It is even conceivable that this principle is powerful enough to deal with ALL serial-ordering relations in a language whose phonotactic patterns are relatively restricted, and there would be no reason for discrediting such a hypothetical language as an impossible one.

On the other hand, the limits of this ordering scheme are all too obvious. It fails in multisyllabic words, and perhaps more significantly, it falls flat in monosyllabic words consisting of the same segments but requiring different orders (Lashley 1951). Take the example of *tip* and *pit*. No phonological principle can generate both orders. By implication, the serial-order information must be lexically represented. If the processing system must encode serial order lexically for some words, then it stands to reason that this must hold for all words, and not just those which happen to have neighbours with the same phonemes in a different order. Otherwise, it would have to be claimed that each word is compared in some mysterious way to all others and that the outcome determines whether linear order is lexically or phonologically coded. In view of the other limitations of the phonological ordering scheme (e.g. its restriction to monosyllabic words and to syllables with an adequate sonority profile, as in *ask* but not *axe*), this strategy cannot be considered a serious option. However, this is not to say that phonological information is entirely irrelevant in generating linear order. It might play a subsidiary role compared to lexical information.

It is worth reiterating that the structural patterns requiring lexical representation of serial-order information are not a necessary part of language. They could be dispensed with without reducing language to an impracticable tool for communication. The reason for lexical coding thus is not logical necessity but rather flexibility and efficiency. With lexical coding the system can generate *any* serial order irrespective of which phoneme orders have already been established in the word store.

In summary, both phonological and lexical ordering schemes have been argued to be viable in principle. Whereas the lexical scheme has virtually unlimited coding power, the phonological scheme can cope only with a restricted set of structural patterns. Natural languages show variability significantly beyond what is permitted by the phonological scheme. It seems fair to conclude that the appropriate representation of serial-order information is at the lexical level.

3. The coding of serial order

The claim that serial order is lexically coded hints at where the information is located, but it neither pins down the locus nor specifies how the information is coded.[1] Local connectionist models offer two distinct lexical loci for coding serial order: vertical links connecting the word and the segment levels, and horizontal links at the lexical level.[2] These options are represented in Figs. 2B and C. For the sake of contrast, phonological coding of serial order is shown in Fig. 2A. The sample word in all cases is *tip*. Phonological representations are phonemic rather than allophonic.

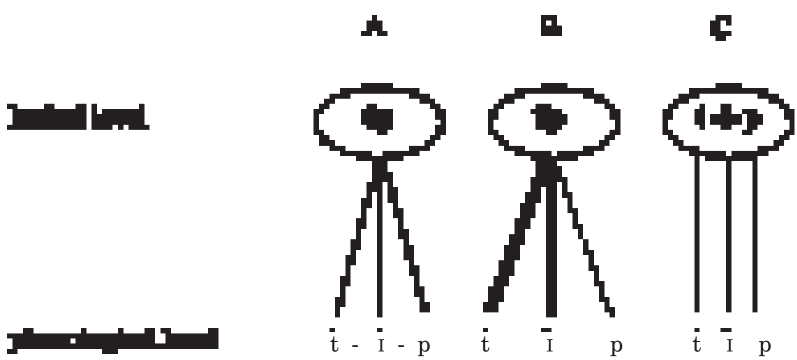

Figure 2. Three modes of coding the serial order of segments within words

Since Model 2A was rejected in the preceding section, it will not receive further attention. In Model 2B, the word node is a unitary concept linked up with three segment nodes. Serial order is coded by differential weighting of the links between the lexical and phonological levels (Houghton 1990). The earlier the occurrence of a segment in a word, the stronger the link. Thus, the link between "tip" and /t/ is stronger than that between "tip" and /ɪ/ which in turn is stronger than that between "tip" and /p/. As linkage strength is an index of the amount of activation that accumulates on a node, the alveolar stop amasses more activation than its contenders and is outputted in the first round. After its selection, it is clamped down by a self-inhibition mechanism and is no longer available. This allows the vowel to become the most highly activated node, to be selected and subsequently to undergo self-inhibition. The bilabial stop is produced in the final round.

Unlike Model 2B, Model 2C assumes a lexical node with a complex internal structure. In particular, it draws a distinction between the internal constituents of lexical items and elements at the segmental level. Since the former are used to access the latter, these constituents were termed segment programmes in Berg (1993). Another way of distinguishing between these units would be to regard the former as tokens and the latter as types. In any case, their name should not be taken to prejudice a phonological interpretation of the concept in question. What is claimed by this model is that the between-level links do not meet in one place at the word level, but make contact with different parts of the lexical item. To establish serial-order relationships, these parts are connected via horizontal links. In the sample word, the programme "t" is connected to "i" and "i" is connected to "p". No direct link is assumed between "t" and "p". Serial order is effected by spreading the most activation to the first phoneme; the next programme is also activated by way of the horizontal connection. After the first phoneme has undergone self-inhibition, the second programme ensures that the corresponding phoneme is the most strongly activated. At the same time, activation is passed to the third programme and so on.

There is an important parallel between Models 2B and C. The links responsible for serial order are assumed to be separate in both. The horizontal lines in Model 2C are as private as the vertical lines in Model 2B. This makes sense in the case of different segments and different orders. Take the words *pit* and *shed*. Since there is no segmental similarity between them, there is no basis for shared representations. But is the same decision appropriate for *pit* and *pin*? The horizontal model (2C) postulates separate segment programmes, and implicationally separate links, between the "p" and "i" in the two words even though the information is identical. By the same token, because the vertical model (2B) postulates distinct lexical nodes, the links connecting "pit" and /p/ or /ɪ/ must be distinct from the links connecting "pin" and /p/ or /ɪ/.

This mode of representation does not conform to the general philosophy of connectionism. An overarching principle guiding the construction of connectionist models is that identical information is coded by the same nodes. Thus, the identical beginnings of *pit* and *pull* lead to the creation of a segment node linking the two words. Similarly, no common representation beyond the segment nodes is assumed for *pit* and *pin* although the ordering relation between "p" and "i" is also one respect in which the two words are identical. Thus, identity is given an inconsistent treatment in local connectionist models. It is taken as essential in some areas but plainly ignored in others. *A priori*, it is not clear why serial order should require a different representation from other types of information. The following section will examine whether it is empirically justified to treat identity and non-identity of serial order indiscriminately in theory construction.

4. Child language as a testing ground

A hallmark of connectionist theory is that learning takes place in the links (McClelland, Rumelhart & Hinton 1986). By increasing linkage strength, more activation is relayed to the critical nodes and the likelihood of correct output is enhanced. Learning in this framework is a local process. Each link has to be modified individually. This principle serves as the starting point for an analysis of private versus shared representation. If two behaviours are acquired at the same time, they are more likely to be brought about by the same change in the network than by different changes. Of course, two distinct changes in the network are also possible, but they are made less probable by the requirement that they happen simultaneously. A simultaneous change in various places is also possible, but its likelihood is inversely proportional to the number of places at which a change has to be effected.

Applying this logic to the problem of serial order, the following predictions can be made. If the ordering relation between "p" and "i" is separately coded for the words "pin" and "pill", children should learn the transition from /p/ to /ɪ/ on a word-by-word basis. At a certain point, this transition would be mastered in one word but not in another. If, however, the same link is responsible for the ordering relation between "p" and "i" in all words beginning with /pɪ/, learning should be across the board, i.e. all relevant words should benefit from a single change in linkage strength and therefore evince an improvement towards the adult norm.

It should be stressed that these predictions are somewhat idealized. It is not the case that the 'quality' of a given output depends solely on the strength of one link. Nodes accumulate activation from many sources, and this introduces a good deal of variability. Even if serial-order information were shared among lexical items, it would be unrealistic to expect complete homogeneity in the data. However, this need not blur the empirical difference between the competing modes of representation. Whereas minor variation would be compatible with shared representations, major variation would not. Conversely, private representations allow us to expect major, not minor variation.

Testing these predictions requires detailed documentation of a completed sound change in child language. The best test would draw not only on information about all relevant words, but also on repeated occurrences of the same word within a minimum time interval. Of course, such an ideal is practically impossible to attain. Children, like adults, never produce all the items from their mental dictionary even within an extended period of time. Nor can they be made to do so under more or less ordinary speaking conditions. Repeating a list of words after the examiner is not the same as spontaneously producing an item with no recent acoustic image. Also, it is impractical to record all utterances over a prolonged time span; every corpus of spontaneous speech is of necessity fragmentary.

A suitable data collection procedure must strive for maximum sampling density and naturalness. The requisite depth in lieu of breadth can be attained only by single-case studies. In the present case, the naturalness criterion was met by selecting as the subject my daughter Melanie, with whom I interacted in ways that are typical of everyday family life. I refrained from eliciting responses of particular theoretical significance and contented myself with recording what Melanie uttered on her own initiative. As it is unthinkable to focus on all potential areas of change, the analysis was confined to one issue, the change from a voiceless alveolar to a voiceless velar stop in word -onset positions.

The following procedure was adopted. Each rendition of a word beginning with /k/ in adult language was noted on a daily basis. The child's rendition could be in conformity with the adult norm (i.e. a word beginning with [k]) or in conflict with it (i.e. a word beginning with any segment other than [k], most usually [t]). The two types of rendition (i.e. the correct and the incorrect one) were written down once per day and were termed *daytypes*. For example, six correct renditions of a given word would be given the same importance as one incorrect rendition on the same day. This method levels the frequency relations among the daytypes. However, since almost the entire period of acquisition is covered, an overall perspective may reflect at least part of the frequency information.

I observed my daughter over a period of almost twelve months (from age 3;4,9 to 4;3,20), spending an average of two to three hours a day with her. I either listened passively to what she said to other people, or to herself, or actively engaged in conversation with her; I immediately wrote down all utterances which satisfied the above criteria. After completing the data collection, all utterances (N = 11,224) were computer-coded on a number of potentially interesting criteria such as stress and context.

Melanie acquired Standard German in a monolingual context. She was a late talker, uttering her first adult-like words at 1;7, and she did not achieve full mastery of her phonological system until nearly five years of age. Her pronunciation was usually clear so that there was little difficulty in identifying the segments she used in her correct as well as incorrect renditions. In particular, this was true of the [k] - [t] contrast under investigation here: the alveolar stop, even as a substitute segment, sounded like a perfect alveolar stop to my native German ears.

The area of primary interest was the context in which the critical segment appeared. Since the velar or alveolar stop always occupied the word onset, only the subsequent context (either a vowel or a single consonant) was assumed to play a role.[3] Three-consonant clusters beginning with /k/ or /t/ do not exist in German. For example, (1-2) show adult-like renditions of the velar stop while (3-4) exemplify deviations from the adult model. The critical consonant is followed by a vowel in (1), the alveolar nasal in (2), the lateral in (3), and the rhotic in (4). Each utterance has a phonetic transcription of the relevant parts and an English gloss. The child's age is given in parentheses.

(1) und Papa muß essen kochen [kOx@ n] (4;2,13)
'and Daddy has to cook the meal'

(2) Du weißt nicht, was ein Knecht [knEçt] ist (4;2,15)
'You don't know what a servant is'

(3) Wir sind beide Tlowns [tlauns] (= Clowns [klauns]) (4;2,19)
'Both of us are clowns'

(4) Was heißt "Trankheit" [tra˜khaIt]? (= Krankheit [kra˜khaIt]) (4;2,23)
'What does (the word) "illness" mean?'

These four contexts exhaust the phonotactic possibilities of German, except for the cluster /kv/, which occurs in so few words in the data that a reliable analysis is precluded.

If the links responsible for serial order (i.e. the vertical connections in Fig. 2B or the horizontal connections in Fig. 2C) are particular to each word, we should expect no correlation between correctness of the velar stop production and subsequent context. Thus, even if two words begin with the same cluster, their onsets should be mastered to different degrees.

An examination of Melanie's utterances reveals that this is not the case. Subsequent context turns out to be the most reliable predictor of correct velar-stop production. When /k/ is followed by a vowel or alveolar nasal, it is articulated in an adult-like manner, but when it is followed by a lateral or rhotic, the child experiences great difficulty: [k] is usually replaced by [t] before the lateral and elided before the rhotic. Fig. 3 presents the rate of correct production as a function of subsequent context and time. The period under study was divided into weeks and the data from seven consecutive days were pooled. Correct (adult-like) productions are given as a percentage of total productions per time unit. Fig. 3 is complemented by Table 1 which presents the absolute numbers of correct and incorrect productions on which Figure 3 is based. Information on sample size may be useful to fully appreciate the role of context.

It is apparent from Fig. 3 and Table 1 that each context has its own pattern of accuracy. While the velar stop is mastered fairly quickly and reliably before a vowel or /n/, the /kl/ graph hovers

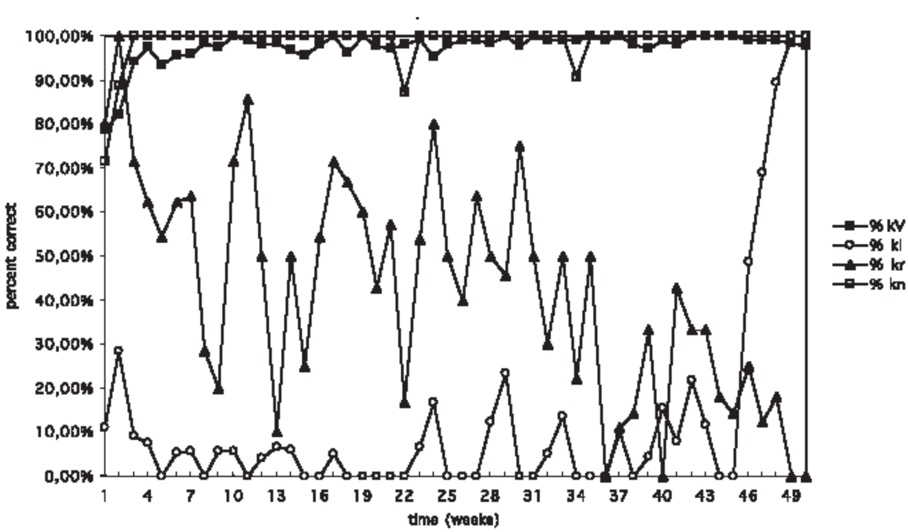

Figure 3. Percentage of correct /k/ productions as a function of subsequent context.

Table 1. Number of correct (c.) and incorrect (i.) productions of /k/ and /g/ as a function of subsequent context

Target	Context	_V c. i.	_l c. i.	_r c. i.	_n c. i.	_v c. i.
/k/		5140:162	118:815	186:256	319:4	7:9
		96.9%:3.1%	12.6%:87.4%	42.1%:57.9%	98.8%:1.2%	43.8%:56.2%
/g/		2850:306	55:476	133:393	0:1	–
		90.3%:9.7%	10.4%:89.6%	25.3%:74.7%	0%:100%	–

around a minimum level of accuracy[4] for most of the time and the /kr/ graph oscillates between 10% and 80% accuracy. The significant point is that the shape of the graphs is relatively constant over time. Either the accuracy level remains largely the same (as with /kV/, /kn/ and /kl/) or the oscillation persists (as with /kr/).

The results for /kV/ and /kn/ are the easiest to interpret: almost all words were pronounced correctly. In short, these words act homogeneously. Fig. 3 suggests, but does not prove, that the same is true of /kl-/ and /kr-/ words. Notice that although the child's lexicon expands over time, the /kr/ and /kl/ rates remain essentially unchanged, suggesting that production accuracy is independent of the size of the lexicon.

However, these two graphs do not rule out the possibility that the low accuracy for /kl-/ results from most words undergoing the /k/ → [t] substitution, while certain others are articulated flawlessly. There would then be two sets of /kl-/ words with quite disparate characteristics. The same might be true of /kr-/ words. The extreme vacillation between low and high production accuracy could reflect a preference for easy /kr-/ words on some occasions, and difficult ones on other occasions. It seems unlikely that Melanie chose her words on the basis of processing difficulty rather than semantics, but this hypothesis must be falsified if we are to reject the possibility that she had two classes of /kr-/ words.

Fig. 3 is also insufficient because weekly analysis is too coarse-grained. There are at least two possible scenarios not distinguished by Fig. 3. It may be that a given change affects one fraction of the relevant words on the first day and another on the seventh day, or the change may spread slowly from one word to another within one week. Although these two scenarios are quite different (i.e. abruptness in the first and gradualness in the second), they would produce the same weekly graph. What is needed therefore is a word-by-word analysis with maximally thin time slices. The thinnest time slice allowed by the data collection procedure is the day.

I have singled out /kl-/ words for closer scrutiny because the observational period spans the entire range from complete failure to perfect mastery. Table 2 presents the daytypes for each lexical item uttered at least once during a span of one month. To keep the size of the table within limits, all inflectional variants were subsumed under one head entry, e.g. the verbal infinitive and the uninflected form of the adjective. Pluses indicate an adult-like production, minuses a non-adult-like production. When both a plus and a minus occur in the same cell, the child gave a correct and an incorrect version of the same word on the same day. Since only one type was collected per day, it is not possible to provide the absolute number of productions for each cell. The period covered in Table 2 was selected by determining a reference point (see below) and including 15 days before and 15 days after that date.

The crucial observation to make about Table 2 is that the change from [t] to [k] affected several words at once. This change took place on one day when Melanie was 4;2,23. She correctly produced three words which she had rendered incorrectly during the two preceding weeks (*Kleid* 'dress', *klein* 'little' and *Klammer* 'clothes peg'). This moment is identified as 0 in Table 2. Two days later, two further items were uttered in their correct form, and another two words four days later.

In view of this global emergence of the velar stop, it may seem justified to advance the strong claim that *all* /kl-/ words underwent a change on the same day. While much can be said in favour of this conjecture, it is difficult to prove, because there are natural gaps in the data. As noted, no speaker utters all words every day. We cannot be sure how Melanie would have pronounced the other relevant items on that critical day. However, there is a significant developmental trend. Before that day, no word was pronounced correctly (but see below), and there are only misses in Table 2. After that day, correct pronunciations are clearly predominant, though not exclusively so. Three classes of words may be distinguished. Category A comprises items for which no minuses were recorded beyond the critical day (e.g. *Kleid* 'dress' and *Klötze* 'blocks'). The five words of Category B were rendered incorrectly only once (e.g. *Klo* 'toilet' and *klingeln* 'to ring'). Category C consists of the single word *klein* 'little' whose correct renditions are frequently accompanied by incorrect ones. The high number of /k/ → [t] substitutions beyond the critical day is almost certainly due to the high frequency of this word in Melanie's

Table 2. Production accuracy of /kl-/ words over a four-week period.

	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Kleid	-			-	-					-						+							+	+	+			+		+	+
klein	-	-	-	-	-	-	-	-	-	-	-	-	-	-		±	±	±	±	±	±	±	-	+	+	±	+		+	+	+
klettern					-	-										-						+							+	+	
klauen																-		+	+			+									
Klebe								-				-																+			
klemmen							-	-										+													
Klo											-			-						+	+	-					+				
Klappe		-	-					-																							
Clown						-						-													+						
klingeln						-		-														-									
kleckern													-		-			-													
Klöße															-	-															
Klammer	-											-				+															
klar									-																						
klipp																						+									+
Klipper																				±								+		+	
klapp																						-									
Klötze																								+							+
klingeln																												+			
Klasse																													+		

vocabulary. Because she is so used to rendering /klaɪn/ as [tlaɪn], she has great difficulty getting rid of the incorrect form.

The existence of variable pronunciations during the period of change in no way undermines the analysis offered here. It merely serves to show that the production problems cannot be solved instantaneously. The essential claim is that these problems are *attacked* for several, perhaps even for all, relevant words simultaneously. The word *simultaneously* should not be taken too literally, because the three words *klein, Kleid* and *Klammer* were not produced in adult-like manner at exactly the same time. It is thus theoretically possible to maintain that the correct pronunciation of these words is the result of three individual changes in the mental network. However, this conjecture dismisses the correct production of the three words within a couple of hours as purely coincidental. Given their consistently incorrect pronunciation before that point, it appears warranted to claim that the changes have a common cause.

It is now possible to provide a clear interpretation of the /kl/ and /kr/ graphs in Fig. 3. We can rule out the objection that the shape of the graphs is a compromise between the impact of two word sets, one causing pronunciation difficulty but not the other. As Table 2 demonstrates, there is a great deal of homogeneity in the data. All items that belong to the set of /kl-/ (or /kl-/) words display uniform behaviour. They change in a non-random fashion in that there is a moment before which all of them are incorrectly pronounced and beyond which most of them are correctly pronounced when they are attempted.

Despite certain inherent limitations of the data, there is good reason to conclude that the change of [t] to [k] operates on all /kl-/ words as a group. This uniformity can be attributed to the fact that all these items have the same consonantal onset. The constraints this finding imposes on the representational system will be explored in the next section.

5. Implications

We started from the assumption that linguistic change may be exploited as evidence of representational closeness. If a change affects one word at a time, lexical items may be argued to be representationally separate; if it affects several items simultaneously, representations are more likely to overlap. We may even put forward a stronger claim to the effect that representational overlap leaves pertinent words no choice but to change in unison. The empirical data speak in favour of the representational-overlap hypothesis. The pattern of change suggests close interrelationships among lexical items. This result is incompatible with the standard view in which lexical items are separately stored. What, then, does a more adequate representation look like? Since change was conceived of as taking place in the links, one possibility is to reserve identical links for different words. This can be achieved by placing identical parts of words 'on top of' one another, affording lexical representations a 'superpositional' structure. Figure 4 illustrates the representations of the sample words *clumsy* and *clever* as well as *ravage* and *savage.* In these example, identity of serial order is coded on horizontal links (see Fig. 2C). Phonetic symbols are used to represent segment programmes.

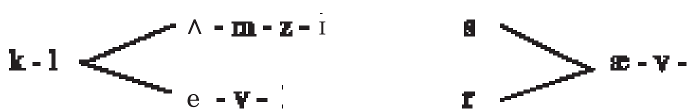

Figure 4. Shared representations of lexical items.

As shown in Fig. 4, each word is represented not by one node but by several, each coding some local serial-order relationship (e.g. "k-l"). Words with identical phoneme sequences share one or more nodes while words with non-identical sequences do not. The higher the number of identical sequences, the higher the number of shared nodes. In this conception, different words are not represented by different sets of nodes but by partially overlapping sets of nodes which are linked horizontally.

It should be emphasized that these representations mark a radical departure from local connectionist practice. The representations of words are no longer neatly separated; they overlap messily. Word nodes thereby lose their integrity, though not their individuality. Since no two words, homophones excluded, have exactly the same horizontal links, the system has no problem keeping them apart.

The representations in Fig. 4 deal in a straightforward manner with the empirical data. Because the same horizontal link underlies the onsets of the words *klein* and *Klammer,* strengthening this link cannot help but increase the availability of the velar stop. This is because the "k" receives more activation from the "l" after than before the change. More generally, performance on many different words can be improved by adjusting a single horizontal link.

Can the same effect be obtained with the help of vertical links as shown in Fig. 2B? The answer is clearly in the negative. In this model, the integrity of each lexical node is preserved. By implication, there cannot be a single link connecting two different word nodes with the segment level. That is, the vertical links must remain private by virtue of the separateness of the lexical nodes. It follows that this model is incapable of accounting for the homogeneity observed in Melanie's data (barring the improbable assumption that different links are changed in rapid succession). As a means of coding serial-order information, the idea of varying strength of vertical links thus appears less adequate than the notion of horizontal links.

Before the shared-representation hypothesis can be accepted, it is necessary to address a number of objections. In Fig. 4, the locus of serial-order information is the lexical level itself. However, it might be possible to locate serial-order information at a different level, as has been done for the coding of phonological and semantic information (see Fig. 1). Such a level would have to consist of nodes that represent all adjacent-phoneme pairs allowed by the language. For instance, there is a node for "k-l" and one for "k-r" but none for "k-b". A word like *stick* would have connections to three nodes at the serial-order level (i.e. "s-t", "t-i", "i-k"), as shown in Fig. 5.

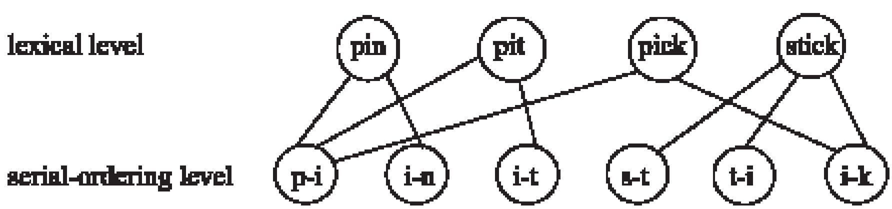

Figure 5. A fragment of a lexical network with a separate serial-ordering level.

In this model, Melanie would be presumed to have increased the efficiency of the "k-l" node during her learning process. Since this node subserves all words beginning with /kl/, the pronunciation would improve on all these words at the same time.

There is much to take issue with in this model. To begin with, it shifts the locus of learning from the links to the nodes. This is counter to the general philosophy of connectionism and would render this learning problem quite unlike all other learning problems, which have been identified in the links. Even more detrimental is an empirical argument: the creation of a new level introduces the possibility of between-level interactions and thus allows us to expect certain errors. For example, interaction between the word and the phoneme levels may lead to an exchange of phonemes within a word, as in the proper name *Kimura -> Mikura* (from Fromkin 1973). By analogy, if a serial-order level exists, we would expect switches of order nodes. Let us come back to the case of “stick”. If the first and the last order nodes swap places, we will end up with something like “iktist”. Such errors are completely unattested. Finally, the serial-order level is incapable of dealing with the repeated occurrence of a segment in the same word. In *assimilate* for example, it would not be clear after the generation of the “s-i” node whether to proceed with the “i-m” or the “i-l” node. As both alternatives would fit, the system would be stymied and efficient production would be hampered. For all these reasons, the idea of a distinct serial-order level should be rejected.

Melanie’s course of learning might be interpreted in other ways. One might create another level between the lexical and the segmental level, the consonant-cluster level, as illustrated in Fig. 6.

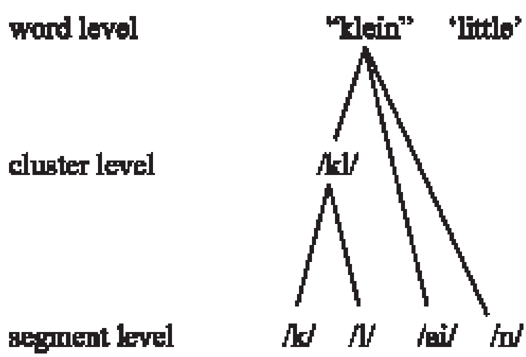

Figure 6. The enlargement of the processing network by a cluster level.

This model provides a novel perspective on Melanie’s difficulty with /kl-/ words. Her learning problem may be remedied by strengthening the connection between “klein” and /kl/ or that between /kl/ and /k/. The link between /kl/ and /l/ may be assumed to be fully developed as Melanie had no trouble articulating the lateral correctly (see (3)). The empirical data immediately exclude the link between “klein” and /kl/ as the source of the trouble. Since each word has a private connection to the cluster level, the repair has to proceed on a word-by-word basis and any improvement cannot occur across the board. As argued above, this is not how Melanie progressed. So we are left with the link between /kl/ and /k/. Indeed, if this very link was strengthened, all pertinent words would receive more activation and could be pronounced correctly. If this model were viable, it would obviate the need for shared representations at the lexical level.

However, a cluster level has no empirical support. One major justification for postulating distinct nodes is that they may be holistically involved in malfunctions. Whole-word slips are evidence of word nodes as much as segment slips are evidence of segment nodes. Therefore, a cluster level would entail the prediction that clusters customarily act as wholes in error processes. This is not, however, the case. While clusters may sometimes exhibit a monolithic behaviour, usually only one of their constituent parts is misordered. Berg (1994) reports that, averaged across the major cluster types of English, 73% of cluster errors were of this type. The two possibilities are illustrated below.

(5) clamage dame. for: damage claim. (from Fromkin 1973)

(6) craperies cleaned. for: draperies cleaned. (from Fromkin 1973)

As can be seen, the /kl/ cluster remains intact in (5) but splits up in (6). Since cases like (5) are the exception rather than the rule, it must be concluded that there is no distinct cluster level in the mental network.[5] If there was, we would expect clusters typically to behave like indivisible units, as we observe in the case of phonemes and morphemes. This argument is based on adult language but can be straightforwardly extended to child language. Generally speaking, children's networks are less structured than those of adults. Therefore, it is at least conceivable that the adult lexicon contains a cluster level while the child's lexicon does not, but it is highly unlikely that a cluster level exists in children's networks but not in those of adults.

Another strategy to salvage the private nature of lexical representations would be to locate the cause of Melanie's behaviour one level lower in the hierarchy, viz. in the links connecting the segment and the feature levels (not shown in Fig. 6). Suppose that the link between /l/ and the place-of-articulation feature [alveolar] is too strong, stronger in fact than the link between /k/ and [velar]. As /k/ and /l/ are simultaneously activated – they belong to the same planning unit – the alveolar-feature node will be more strongly activated than the velar-feature node at the moment when the production of the velar stop is imminent. As a consequence, the /k/ will surface as

[t]. This would be an instance of partial harmony, a phenomenon that is typical of child language and that figured quite prominently in Melanie's speech. Consider (7).

(7) [me:b@ l]. for: [ne:b@ l]
'fog' (from Berg 1992)

This example documents the replacement of /n/ by [m]. The intrusion of [m] can be accounted for by positing a hyperactive link between the bilabial stop and the feature [bilabial]. This hyperactivation in the place domain surpassed the activation of the intended feature [alveolar] and turned the /n/ into a bilabial nasal.

Let us suppose next that Melanie lessened the strength of the overloaded linkage between /l/ and [lateral] on her way to complete mastery. This would have a non-local effect at the lexical level: all relevant words would profit from it insofar as all /k/'s could be pronounced as [k]. Again, if this model is correct, there would be no need for shared representations.

In fact, this model also fails on empirical grounds. Excessive strength of the /l/ – [lateral] or of the /l/ – [alveolar] link predicts lateral or alveolar harmony in Melanie's output. For instance, *holen* [ho:l@ n] 'to fetch' should be rendered as [lo:l@ n]. No such forms occurred in Melanie's speech. The only form of consonant harmony that Melanie produced was bilabial. This harmony was not restricted to one bilabial consonant but held for all bilabial segments. If Melanie had an overly strong link between /l/ and [alveolar], we would likewise expect all links connecting alveolar segments with the feature level to transport an excessive amount of activation. Concretely, we would expect partial harmony to also arise within the onset cluster /kn/, which should turn up as [tn] because the alveolar feature of the nasal suppresses the velar feature of the stop. However, as depicted in Fig. 3, Melanie did not have the slightest difficulty with /kn/ clusters. We thus find little evidence to suggest that the cause of her development lies in the links between the two phonological processing levels.

The final alternative is that Melanie's output is under the control of phonotactic constraints. Her inability to pronounce /kl/ in adult-like fashion would be attributable to an output constraint on this cluster. By contrast, /kn/ is not subject to such a constraint and can be freely uttered. In this view, learning means overcoming phonotactic constraints (as espoused in the natural phonology approach to phonological acquisition, see Stampe 1969).

There are two objections to this view, one of a more theoretical and the other of a more empirical nature. First, the notion of phonotactic rules is quite alien to the spirit of connectionist models. These models do not provide for the explicit coding of rules but instead generate regularities through the simultaneous activity of a large number of nodes which bias the output towards the majority pattern. As Derwing & Skousen (1989) argued, this bias is conducive to analogical, not rule-based behaviour. Connectionist systems produce a probabilistic, not a deterministic output, as is characteristic of rules.

Secondly, by its very nature, a phonotactic constraint, especially if it has an articulatory basis, allows no exceptions. Melanie should never have produced a velar stop preceding /l/ before the age of 4;2,23. Again, the facts suggest otherwise. Although there was not a single correct /k/ before /l/ token during the two weeks preceding the critical day, Melanie occasionally got the velar stop right before the period covered by Table 2 (see Fig. 3). Her first correct production of /kl/ occurred in *Klöße* 'dumplings' when she was 3;4,10, that is about ten months before she eventually resolved this problem. These occasional adult-like renditions of /k/ before /l/ invite the conclusion that phonotactic constraints are not an adequate explanation of Melanie's behaviour. If they can be ignored on some occasions, one wonders why they are not ignored all the time, given that they keep the child from advancing towards the adult norm. It is worthwhile to add that occasional correct productions are exactly what one would expect from a connectionist perspective. If the usual level of activation of a given node is exceeded, as may happen on occasion for various reasons, this excessive activation may compensate for the weakness of the relevant link and lead to correct behaviour.

To sum up, none of the three alternatives stands up to serious test. They thus do not undermine the hypothesis that the best way to model Melanie's behaviour is to invoke shared representations for lexical nodes. The fact that Melanie's progress is not limited to individual words suggests that words are not isolated entries in the mental lexicon, but overlap to the extent that they code the same linear-order information. In particular, it is argued that words with identical sequences of two (or more) phonemes share the same lateral connections between the segment programmes. This step dissolves the integrity of lexical nodes.

6. Conclusion

In the foregoing, a case has been made for rejecting the one-to-one correspondence between the representational unit and the unit to be represented. One network node no longer codes one word but contributes to the coding of many words. This step likens the proposed model to distributed connectionism (e.g. Hinton, McClelland & Rumelhart 1986). However, a crucial difference remains. Whereas distributed representations assume a many-to-many relationship between linguistic and mental units, the model developed above assumes a many-to-one relationship. That is, one mental unit subserves many different linguistic units while the reverse assumption is not made. Since this latter assumption, viz. the coding of one linguistic unit by many mental units, is regarded as fundamental to distributed representations, the proposed model may be argued not to belong in this category. On the other hand, relinquishing the principle of one mental unit representing one linguistic unit clearly distances the model from local connectionism and makes it more congenial to distributed connectionism. The model advanced here is therefore best understood as a hybrid which combines aspects of both types of connectionism.

A major advantage of this hybrid model is that the representation of similarity is treated in a more consistent fashion. As mentioned, local connectionist models take account of similarity at the phonological and semantic levels but ignore similarity at the lexical level. This inconsistency vanishes in the present framework. With shared representations, all facets of similarity are taken into consideration and a great deal of redundancy in lexical representations is eliminated.

The remainder of this article will be devoted to exploring briefly whether it may ultimately be necessary to abandon the hybrid model in favour of fully distributed representations. To this end, we examine one aspect of Melanie's phonological development, thus far interpreted within a local connectionist framework, with a view to determining whether it can, or should, be reinterpreted within a fully distributed framework.

As noted above, Melanie's speech exhibited a harmony process whereby bilabial segments imposed their bilabiality on certain phonemes in the same planning unit (see (7)). The critical issue in the present context is that *all* bilabial segments behaved alike. Using a local connectionist model, Berg (1992) construed these harmony patterns as evidence for the claim that all links between the bilabial segments and the feature [bilabial] are overly strong. While this account is clearly compatible with the data, it suffers from one major weakness: one unitary phenomenon is assumed to be brought about by three distinct deficits. Bilabial harmony is put down to excessively strong links between /p/ and [bilabial], /b/ and [bilabial] as well as /m/ and [bilabial]. This explanation implies that it should be possible to find children who have only one or two of the three links with excessive strength. Significantly, to the best of my knowledge, children who exhibit a certain harmony type which is restricted to one segment but does not apply to other segments of the same phonological class, have not yet been reported on in the literature on language acquisition.

Theoretically speaking, two solutions suggest themselves. It may be contended that the underlying deficit is not in the links but in the nodes, particularly in the bilabial-feature node. This account would no doubt match the unitary nature of Melanie's harmony. However, as has been pointed out before, changes in the network are more appropriately modelled as taking place in the links rather than in the nodes (see Berg (1995) for empirical support for this claim). The alternative would be to code segments in distributed fashion whereby all bilabial segments would have shared representations. This would make the bilabial segments representationally more similar than they are in local models. A heightened representational similarity would let us expect a strong behavioural similarity, and this is precisely what Melanie's harmony patterns exemplify. Hence, it is not unreasonable to suggest that a fully distributed approach furnishes a more adequate account of the acquisition data than a local one.

Address of the Author:

Department of English Linguistics, University of Hamburg, Von-Melle-Park 6, 20146 Hamburg, Germany,
e-mail: thomas_berg@uni-hamburg.de

Footnotes

* Dedication and acknowledgements: I am pleased to dedicate this contribution to Bruce Derwing as the linguist-psychologist who has probably had a more profound influence on my thinking than anybody else in the field. When I was exposed to his work for the first time as a doctoral student, I thought to myself: 'Yes, this is it. This is the right way of looking at language.' I am honoured to publish this paper as part of a minifestschrift, although Bruce would clearly have deserved a real festschrift much earlier. – The comments of Pier Marco Bertinetto and Jussi Niemi on an earlier version are acknowledged with gratitude. Viola L'Hommedieu, Sandra Lund and Paulina Orzechowska have done a great job creating the artwork for this paper.

[1] And, of course, it remains silent on how this information is actualized in speaking (see Dell, Burger & Svec 1997).

[2] A third possibility is to shift the responsibility for serial order into the domain of structural nodes. These nodes have been variously called slots (Shattuck-Hufnagel 1979), wordshape header nodes (Dell 1988) or frames (Schade 1992). As a serial-ordering device, structural nodes suffer the limitations discussed in the preceding section. In addition, they require certain assumptions about the phonological status of segments (in particular, context-sensitive coding) which are not empirically sound. Therefore, this hypothesis will not be considered in the following.

[3] The prior context would imply an influence across the word boundary, which is inherently unlikely.

[4] This graph rectifies an embarrassing error that crept into Table 2 of Berg (1995). Fortunately, none of the conclusions reached in that paper are affected by this mistake.

[5] Cluster errors arise as a result of structural units which are gradually activated during the unfolding of suprasegmental representations in speaking (see Berg & Abd-El-Jawad 1996).

References

ANDERSON, John (1987), "The limits of linearity", in: ANDERSON & DURAND (1987:199-220).

ANDERSON, John & Jacques DURAND, eds. (1987), *Explorations in dependency phonology*, Dordrecht, Foris.

ARONOFF, Mark & Richard T. OEHRLE, eds. (1984), *Language sound structure*, Cambridge, MA, MIT Press.

BERG, Thomas (1992), "Phonological harmony as a processing problem", *Journal of Child Language* 19:225-257.

BERG, Thomas (1993), "The phoneme through a psycholinguist's looking-glass", *Theoretical Linguistics* 19:39-76.

BERG, Thomas (1994), "More on intersegmental cohesiveness", *Folia Linguistica* 28:257-278.

BERG, Thomas (1995), "Sound change in child language: A study of inter-word variation", *Language and Speech* 38:331-363.

BERG, Thomas & Hassan ABD-EL-JAWAD (1996), "The unfolding of suprasegmental representations: A cross-linguistic perspective", *Journal of Linguistics* 32:291-324.

CLEMENTS, George N. (1990), "The role of the sonority cycle in core syllabification", in: Kingston & Beckman (1990:283-333).

COOPER, William E. & Edward C.T. WALKER, eds. (1979), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*, Hillsdale, N.J., Lawrence Erlbaum.

DALE, R., C. Mellish & Michael ZOCK, eds. (1990), *Current research in natural language generation*, London, Academic Press.

DELL, Gary S. (1988), "The retrieval of phonological forms in production: tests of predictions from a connectionist model", *Journal of Memory and Language* 27:124-142.

DELL, Gary S., Lisa K. BURGER & William R. SVEC (1997), "Language production and serial order: A functional analysis and a model", *Psychological Review* 104:123-147.

DELL, Gary S. & Padraig G. O' SEAGHDHA (1992), "Stages of lexical access in language production", *Cognition* 42:287-314.

DERWING, Bruce L. & Royal SKOUSEN (1989), "Morphology in the mental lexicon: A new look at analogy", *Yearbook of Morphology* 2:55-71.

FROMKIN, Victoria A. (1973), "Appendix", in: Fromkin (1973:243-269).

FROMKIN, Victoria A., ed. (1973), *Speech errors as linguistic evidence*, The Hague, Mouton.

HINTON, Geoffrey E., James L. MCCLELLAND & David E. RUMELHART (1986), "Distributed representations", in: RUMELHART, MCCLELLAND & The PDP Research Group (1986:77-109).

HOUGHTON, George (1990), "The problem of serial order: A neural network model of sequence learning and recall", in: DALE, MELLISH & ZOCK (1990:287-319).

JEFFRESS, J. Lloyd, ed. (1951), *Cerebral mechanisms in behaviour*, New York, John Wiley.

JESPERSEN, Otto (1904), *Lehrbuch der Phonetik*, Leipzig, Teubner.

KINGSTON, John & Mary E. BECKMAN, eds. (1990), *Papers in laboratory phonology I*, Cambridge, Cambridge University Press.

LASHLEY, Karl S. (1951), "The problem of serial order in behavior", in: JEFFRESS (1951:112-136).

MCCARTHY, John J. (1989), "Linear order in phonological representation", *Linguistic Inquiry* 20:71-99.

MCCLELLAND, James L., David E. RUMELHART & Geoffrey E. HINTON (1986), "The appeal of parallel distributed processing", in: RUMELHART, MCCLELLAND & The PDP Research Group (1986:3-44).

RUMELHART, David E., James L. MCCLELLAND & The PDP Research Group, eds. (1986), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1*, Cambridge, MA, MIT Press.

SCHADE, Ulrich (1992), *Konnektionismus. Zur Modellierung der Sprachproduktion*, Opladen, Westdeutscher Verlag.

SELKIRK, Elisabeth O. (1984), "On the major class features and syllable theory", in: ARONOFF & OEHRLE (1984:107-136).

SHATTUCK-HUFNAGEL, Stefanie (1979), "Speech errors as evidence for a serial-ordering mechanism in sentence production", in: COOPER & WALKER (1979:295-342).

SIEVERS, Eduard (1885/1901), *Grundzüge der Phonetik,* Leipzig, Breitkopf und Härtel.

STAMPE, David (1969), "The acquisition of phonetic representation", *Papers from the regional meeting of the Chicago Linguistic Society* 5:443-454.

STEMBERGER, Joseph P. (1985), *The lexicon in a model of language production*, New York, Garland.