The Development of a Speech Segmentation Strategy for English: An Example of an Emergent Critical Period Effect

Richard Shillcock*, Paul Cairns*, Nick Chater** & Joseph P. Levy†

*Centre for Cognitive Science, University of Edinburgh
**Department of Psychology, University of Warwick
†Department of Psychology, Birkbeck College, University of London
rcs@cogsci.ed.ac.uk, Nick.Chater@warwick.ac.uk, j.levy@psychology.bbk.ac.uk

Abstract

We report an example of how different types of information can become available to the developing infant, in a temporal profile that resembles a critical period effect. We present a statistical analysis of the problem of speech segmentation in which we show that effective segmentation criteria can be discovered bottom-up using increasingly sophisticated representations to achieve a complete competence. We show that a Metrical Segmentation Strategy (Cutler & Norris, 1988) can be discovered in this way. We also show that the utility of this strategy is only apparent for a limited period. When more sophisticated representations are developed the utility of the MSS is no longer visible. This pattern is consonant with the human data. Critical period effects can arise through a process of developing increasingly useful representations.

1. Introduction

Language acquisition seems to exhibit a "critical period" during which normal development may occur, but after which development is incomplete and/or impaired (Lenneberg, 1967). The critical period was initially seen as evidence of two qualitatively different stages of processing, separated by a sharp cut-off, but has latterly been construed as a more gradual, albeit nonlinear, decline in the ability to achieve complete mastery of a first language (cf. Johnson & Newport, 1989; Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1996). Subsequent research has also revealed similar critical periods in the acquisition of sign language (Newport, 1984), in second language learning (Johnson & Newport, 1989), in song learning in birds (Kroodsma, 1981), and in imprinting in ducks (Hess, 1973).

One possible interpretation is that critical period behaviour in language learning reflects underlying neurological maturation that is specific to language (Lenneberg, 1967; Chomsky, 1981); if the brain's resources are not utilised for language learning, then they may not develop further in that respect or they may be used for other ends. This interpretation of the critical period has been the dominant one. It has, for instance, allowed a computational approach to theorising about the critical period, in which evolutionary pressures over numbers of generations affect quantifiable resources available for language acquisition that are directly related to "how much" language can be acquired (Hurford, 1991).

Latterly, there have been suggestions that the underlying maturation may not be language specific but might have general cognitive implications (Newport, 1984; Elman, 1991). Thus, for instance, Elman demonstrates how an initial limitation on the resources mediating memory and attention might facilitate the learning of local syntactic relations over more distant ones and thereby aid the acquisition of syntax overall.

In this study we explore a non-nativist account in which changes in the information available to the developing language processor are sufficient to produce a critical period effect with respect to one particular aspect of linguistic competence – the ability to segment the speech stream. The critical period effect that we discuss will not be the result of maturation alone, nor will it be directly due to the availability of processing resources; instead, the generalisations discovered by the processor will themselves give rise to the relevant effect.

Speech perception and lexical access involve segmentation of the speech stream, and this segmentation competence must be acquired. Cutler and Norris (1988, and in a series of papers with others) advanced the Metrical Segmentation Strategy (MSS) as a crucial algorithm for English: the processor segments the speech at the onset of "strong" syllables, defined as ones not containing /ə/ or a reduced /ɪ/. Such segmentation preferences may be investigated in adults by means of an auditory word-spotting task in which, for instance, *mint* is contained in the items *mintayve* and *mintesh*; subjects respond more quickly in the latter case, as the weak final syllable does not precipitate a boundary within the target substring *mint*. The MSS reflects the fact that the majority of content words in English begin with a strong syllable. The strategy contrasts with a simple syllabic strategy for French (Mehler, Dommergues, Frauenfelder & Segui, 1981), for instance, or a moraic strategy for Japanese (Otake, Hatano, Cutler & Mehler, 1993). In more recent work, the MSS has been incorporated into the SHORTLIST model of spoken word recognition (Norris, 1994) as a processing bias concerning possible-word strings in the speech input.

An individual's segmentation strategy is learned early in life but does not seem to be a simple generalisation across the lexicon of the language that is being acquired. Cutler, Mehler, Norris and Segui (1992) show that English-French bilinguals with equal assurance in both languages possess only one segmentation strategy, which they apply to both languages. Bilinguals have access to the lexica of both languages; the fact that they employ just one segmentation strategy means that such a strategy is not simply an epiphenomenon that emerges from lexical competition within the relevant lexicon. Jusczyk, Cutler and Redanz (1993) have shown that infants in an English-speaking milieu acquire their sensitivity to strong syllables between 6 and 9 months of age. These two sets of data are consonant with a processor that is able to discover an effective segmentation strategy in the first year of exposure to one language, French for instance, and is then unable to revise this knowledge as a result of subsequent exposure to conflicting data (to English, for instance).

An innatist interpretation of these data might be that the early exposure to English sets some parameter in a highly constrained set of given segmentation strategies, which cannot then be reversed, perhaps because of neurological development. Below, we present an account in which the relevant generalisation concerning a segmentation strategy only becomes visible in the statistical analysis of the input to the processor for a limited period during phonological development.

2. A corpus-based analysis

We have analysed an idealised phonetic transcription of the London-Lund Corpus (LLC) (Svartvik & Quirk, 1980) in terms of the contribution that transitional probabilities between segments can make to the development of speech segmentation (Cairns, Shillcock, Chater & Levy, 1997). The LLC is a very large corpus of spontaneous conversational English, transcribed orthographically. In order for the distributional statistics of the speech sounds of English to be truly representative, a large corpus is required. The phonological transcriptions available in dictionaries of English do not reflect the frequencies of occurrence of the different words, and they cannot provide between-word distributional statistics – segment pairs drawn from the end of one word and the start of the next. Only a phonologically transcribed corpus of spoken English can provide the relevant data. Automatic phonological transcription is the only realistic option for such a large corpus. Accordingly we replaced each orthographic word in the LLC with its citation form transcription, using phonologically reduced transcriptions for the function words. We also instantiated a limited degree of phonological assimilation between adjacent segments. We than changed each segment into the corresponding subsegmental representation. Finally, we removed the existing marking of word boundaries and pauses to reduce the corpus to one long stream of bundles of subsegmental phonological features.

The goal of the larger study was to analyse this low-level representation of speech, in which only the subsegmental representations were taken as given, and to investigate how much higher-level structure, principally word boundaries, would emerge simply from "bottom-up" statistical bootstrapping (for further details, see Cairns et al., 1997). We tested whether the distributional statistics were sufficient to indicate where the boundaries might fall, on the simple assumption that a low transitional probability between two adjacent "timeslices" (segments, bundles of features) indicates a boundary between two entities; transitions within words or syllables are more constrained than transitions between words or syllables. We do not need to assume that the processor is predisposed to look for words; the processor's attempts to store, reconstruct and predict speech will all be more successful at points of high transitional probability compared with points of low transitional probability.

3. Discovering the MSS

We tested the hypothesised relevance of distributional statistics to speech segmentation using a recurrent neural network as a convenient means of calculating the statistics of a representative fraction of the corpus. Input to the network was the subsegmental representation of the corpus, one segment at a time. At its output level, the network was simultaneously required (a) to recreate the input at the current timeslice, (b) to recall the output at the last timeslice, and (c) to predict the identity of the subsegmental features at the next timeslice. The network was required to do this at each timeslice in the continuous input on which it was trained. For our purposes we were only concerned with the accuracy of its predictions. If the error associated with a particular predicted segment at the output was high, then a boundary was assumed. These hypothesised boundaries were compared with the veridical boundaries, for the whole range of different values of the error cut-off point for assuming a boundary.

The results confirmed that distributional statistics are a valuable source of segmentation information. The segmentation produced was significantly better than chance, although still modest: in its representative "best" performance some 21% of boundaries were successfully recognised, which, when combined with the pauses that provide definite evidence of word boundaries, produces a hit-rate of some 32% of boundaries, with a hits:false-alarms ratio of 2.4:1. In achieving these segmentations the network was not distinguishing between word boundaries and syllable boundaries. In summary, given only a continuous subsegmental input, rare transitions between one segment and the next tend to correspond to syllable boundaries, and therefore – typically – to word boundaries. No lexicon is necessary to begin learning about segmentation criteria. We may assume that these statistics, defined principally within pairs of adjacent segments, are available to the infant; they give rise to unbroken stretches of speech that are the precursors of lexical entries.

Crucially for our current concerns, the MSS is discovered straightaway in this approach: the network tends to discover word boundaries which are at the beginning of content words with a strong initial syllable, more often than would be expected by chance. This bias towards producing the segmentations favoured by the MSS simply emerges from the distributional statistics. This effect is surprising first because a majority of words in the

LLC are function words, which typically begin with weak syllables, and second because the network was basing its decisions largely on consonant strings and not on vowels. Thus, the MSS is not discovered simply because content words predominate; they do not. Nor is the MSS discovered through a direct consideration of the nature of the vowel in strong syllables. Discussion of prosodic aspects of speech segmentation tend to concentrate on the role of the vowel, partly because of the perceptual salience of steady-state speech sounds (see, e.g., Cutler & Mehler, 1993), but as the results from our network show, the MSS can be discovered in principle from a consideration of the relation between the first segment of the word (typically not a vowel) and the last segment of the previous word; the network's prediction task was based solely on past and current information and could not take account of information downstream, such as the nature of the vowel.

In summary, the generalisation captured by the MSS becomes apparent from the first attempts at representing the speech stream, based on subsegmental input. This fact is in accord with the MSS's appearance in the first year of life.

The network's bias towards discovering the boundaries at the beginnings of content words is explained by the fact that function words predominate in the corpus, and their constituent segments (which anyway tend to be more alike (Kelly, Shillcock & Monaghan, 1996)) are therefore the highest frequency segments, which the network tends to predict with greater readiness and reliability, and consequently less error. The initial segments – typically consonants, but also vowels – of content words are predicted with the greatest error and therefore precipitate segmentation.

4. Adding categorial knowledge

We have seen that the most basic distributional statistics of the speech stream give rise to a bias in favour of segmentation before strong syllables, thereby revealing the very useful generalisation captured by the MSS. We assume that the infant speech processor has at its disposal a certain amount of general purpose representational capacity, and that ongoing speech input configures this capacity in the relevant ways. We further assume that the developing brain is sensitive to statistical generalisation and is able to instantiate new representational levels or categories of proven utility; without the requirement of proven utility, there is no limit to the arbitrary representations that would be possible. Given the nature of learning and representation in the brain, it is perhaps truer to say that it is impossible to prevent such statistical generalisation. We have seen, then, that the MSS emerges as a useful segmentation criterion early in the development of phonological processing and we assume that it becomes instantiated in the developing segmentation competence at this point. We will now see that further phonological development can act to eclipse the utility of the MSS.

Phonological development occurs throughout the first year, and beyond (see, e.g., Werker & Tees, 1984). Categorial representations of segments are created and refined, and the infant therefore has the opportunity to represent new and better statistical generalisations at this higher, segmental level, and to construct and refine the statistics of occurrence and co-occurrence at this level. Again, configuring itself to the statistics of these higher levels of representation is the normal behaviour of the developing brain.

We simulated this development of categorial representations by normalising for segment type. In the analysis of the network's behaviour, described above, an error score was treated identically regardless of whether it belonged to a high or low frequency segment. Low frequency segments tended to be associated with higher error and to precipitate segmentation. Normalising involved dividing the relevant score by the frequency of the segment being predicted, thereby making the prediction of high and low frequency segments more nearly comparable.

Normalising for segment identity in this way improved segmentation performance somewhat, but it also caused the utility of the MSS to become invisible. There was now no significant segmentation advantage for words beginning with strong syllables, or for the boundaries occurring before content words.

Normalising for segment type had offset the network's initial facility in processing the high frequency constituent segments of function words and had augmented the error scores for these segments, increasing the chance of a boundary being postulated before these segments.

5. General discussion

We have seen that a segmentation competence can be bootstrapped from a knowledge of the distribution of subsegmental features. In Cairns et al. (1997) we show that this competence improves with the calculation and use of more sophisticated segmental distributional statistics, such as the observed distribution of segments about the veridical boundaries.

We have seen that the MSS can be discovered on the basis of the earliest available distributional statistics. The MSS is one of the most powerful generalisations about the phonological structure of English with respect to segmentation: some 50% of actual word boundaries in typical speech begin with strong syllables.

Phonological development involves the creation and fine-tuning of categorial knowledge about segments. Possessing a categorial representation of /p/, for instance, allows its distributional statistics to be calculated precisely. This increased precision improves segmentation performance, but it also eclipses the utility of the MSS. There is, therefore, only a limited window during phonological development when the processor is confronted with the apparent utility of the MSS. We claim that this discovery is necessary and sufficient to cause the MSS to be incorporated into the processor's segmentation-related behaviour.

An infant initially raised in a French-speaking milieu, having acquired the appropriate syllabic segmentation strategy and the relevant phonemic categories, and then being exposed to both English and French would be expected to bring existing segment categories and generalisations to bear on the English input. The English speech would be analysed in terms of the French segment inventory (increasingly modified to reflect the bilingual input) and the accompanying frequency statistics. Crucially, this sequence of events does not permit the generalisation underlying the MSS to be discovered. The established segment identities, and their frequencies, will be used to normalise the English input, so that segmentation will not be precipitated at the lower frequency segments that characterise the beginnings of strong-syllable-initial content words in English. (To have been in time to establish the MSS, exposure to English would have had to have occurred before the establishment and refinement of the segment categories, and their associated frequency statistics, which permit normalisation.) Because no useful generalisation has been discovered, no new processing category such as "strong syllable" can be instantiated. The result is that the French-English bilingual (in whom, presumably some greater or earlier exposure to French has caused a syllabic segmentation strategy to be adopted) brings the same (syllabic) strategy to bear on both languages. A necessary additional assumption is that the MSS cannot be retrospectively discovered (much later) as a result of observations carried out at the lexical level; this would mean reorganising lower-level processing that was already working sufficiently well to deliver the correct lexical candidates.

This whole picture is consonant with the human data, in which a segmentation strategy involving the MSS emerges in the second half of the first year of life, but the individual listener does not seem to be able later to augment this strategy with a second syllable-based strategy that would be suitable for listening to French.

6. Conclusions

It is possible for a pattern of development resembling a critical period to emerge without direct genetic supervision in terms of prespecified limits on resources or on the time during which those resources may be used. We have assumed that the processor starts with general purpose representational capacities which may be modified to accommodate the statistics of the input by creating new categories or levels of representation, such as "segment" or "strong syllable". We have said that there needs to be some criterion for the establishment of a new category or level of representation, and we have suggested that the relevant category needs to be discovered to be useful in the course of processing. We have also necessarily assumed that this discovery needs to be made in the course of an ordered progress towards increasingly more sophisticated representations, to avoid a retrospective installation of the MSS on the basis of a much later calculation across the whole lexicon.

The account we have advanced is an alternative to some innatist accounts of the critical period. It may perhaps be seen as related to the general-maturation accounts such as that advanced by Elman in his "starting small" computational experiments. Elman's claim is that the limits of infant cognitive processes may actually facilitate language acquisition because they constrain the representational and processing problems to solvable smaller problems; once this initial progress has been made, it provides a viable basis from which to approach the more complex aspects of acquisition. Acquisition based on such developing resources may be faster or more complete than acquisition in which the complete resources meet the complete problem head-on. The critical period corresponds, in this view, to the period when the resources are still sufficiently immature.

The "starting small" interpretation of the critical period is more compelling when a specific resource, such as a short-term store or an attentional window, can be shown to be involved in the processing and to increase during infancy. In very early phonological development these arguments concerning the available resources seem to apply less: in normal development, sufficient general purpose representational capacity is available to do the job, and the complexity of representation is not directly constrained by resources. In fact, all of the calculations of probability with which we have been concerned above were made between adjacent segments. Rather, the processing was constrained by the fact that the processor had not had enough exposure to the input to develop the necessary new categories and levels, and the related frequency statistics. In the course of acquiring this information in the order of increasing complexity and sophistication, the utility of the MSS was temporarily visible.

We may term these two accounts of critical period phenomena the *resource-based account* (e.g. "starting small") and the *processing-based account*. They both share the claim that the more complex representations in the input (sentences and words, respectively, in the "starting small" and MSS examples) are too complex to allow easy acquisition of their subparts (short syntactic dependencies and strong syllables, respectively) given access to the whole processor. They also share the claim that the development of simpler representations precedes that of more sophisticated ones. In the resource-based account we simply have more understanding of the role of short-term storage, and less of the relationship between short syntactic dependencies and long ones. In the processing-based account, we have less appreciation of the resources needed to calculate particular statistics, and more understanding of the relationship between different levels of phonological representation. Clearly, though, there is an area of intersection between the two accounts.

We conclude by emphasising that critical period phenomena need not necessarily reflect detailed innate specifications of processing, with built-in limits on resources or the time of their use. In principle, robust critical period phenomena can emerge irresistibly during the progressive development of representations as complex as those that found in spoken language, with only minimal assumptions made about the processor.

7. Acknowledgments

This research was supported by ESRC (UK) grants R000 23 3649 and R000 22 1435.

8. References

Cairns, P., Shillcock, R.C., Chater, N., Levy, J. (1997). Bootstrapping word boundaries: a bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, **33**, 111–153.

Cutler, A. & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics*, **21**, 103-108.

Cutler, A., Mehler, J., Norris, D. & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, **24**, 381-410.

Cutler, A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, no. 1: 113-121.

Elman, J. L. (1991). Incremental learning, or the importance of starting small. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, Chicago. Erlbaum, Hove, UK.

Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996). *Rethinking Innateness.* Cambridge MA: MIT Press.

Hess, E.H. (1973). *Imprinting*. New York: Van Nostrand.

Hurford, J. (1991). The evolution of the critical period for language. *Cognition*, **40**, 159–201.

Johnson, J.S. & Newport, E.L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, **21**, 60–99.

Jusczyk, P.W., Cutler, A. & Redanz, N. (1993). Infants' sensitivity to predominant word stress in English. *Child Development.*, **64**, 675-687.

Kelly, M.L., Shillcock, R.C. & Monaghan, P. (1996). Modelling within-category function word errors in language impairment. To appear in *Proceedings of ICPLA '96* (Eds. W. Ziegler and K,. Deger), Whurr.

Kroodsma, D.E. (1981). Ontogeny of bird song. In K. Immelman, G.W. Barlow, L. Petrinovich & M. Main (Eds.), *Behavioral development: The Bielefeld Interdisciplinary Project*, Cambridge: Cambridge University Press.

Lenneberg, E. (1967). *Biological Foundations of Language*. New York: Wiley.

Newport, E.L. (1984). Constraints on learning: Studies in the acquisition of American Sign Language. *Papers and reports on child language development*, **23**, 1–22.

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

Svartvik, J., & Quirk, R. (eds.) (1980). *A Corpus of English Conversation*. Lund Studies in English 56. Lund: Lund University Press.

Werker, J. & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, **7**, 49-63.