Transactions of the Philological Society Volume 102:2 (2004) 227–254

SEGMENTATION: LEARNING HOW TO 'HEAR WORDS' IN THE L2 SPEECH STREAM

By Susanne E. Carroll
University of Potsdam

Abstract

We 'hear words' when we can segment prosodic units from the speech stream and activate associated lexical entries. Segmentation is sometimes regarded in SLA as a perceptual problem, not a grammatical one. I argue here that this view is wrong: segmenting formatives results when we construct prosodic units on the basis of phonetic cues to their edges. The learner's first task is to acquire the relevant cues to these edges. The problem of segmentation is discussed within the framework provided by the Autonomous Induction Theory.

1. Segmentation as a problem of second language acquisition

A competent user of an L2 is someone who shows expertise in four language abilities: reading, writing, listening, speaking (Baker 1993: 6).[1] These are cover terms for complex sets of cognitive and behavioural processes, and proficiency in one does not entail proficiency in the others. I will be concerned here only with listening ability. I will attempt a micro-analysis of what is involved in learning how to 'hear words' in an L2. Hearing words is merely a first step in a series of processes which take the speech signal as their input and

[1] I would like to thank Richard Towell and Roger Hawkins for encouraging me to present the AIT under yet another angle, audiences at the Université du Québec à Montréal, the Graduate Program of the City University of New York, and the University of Calgary, who made comments on oral presentations of this material in March and October 2003, and three anonymous reviewers whose criticisms of an earlier draft of this paper have led to substantial improvements to the text. Errors of fact and interpretation are mine.

culminate in an interpretation. Some L2 researchers appear to think that hearing words is a perceptual process which is independent of grammatical constraints. I argue here that it is the result of grammar acquisition and is constrained by universals of linguistic cognition. Moreover, it involves a kind of acquisition which is signal-dependent and 'bottom up', unaffected by the learner's beliefs or conceptual knowledge.[2] It results from the modular organisation of linguistic cognition. Since hearing words is a necessary first step in word learning, it follows that a proper account of word learning must also assume a modular theory of linguistic cognition. I will attempt to frame the problem of learning how to hear words within 7the Autonomous Induction Theory (AIT) which meets this requirement. To begin, we first need to understand how a knowledgeable user of a language comes to hear words on the basis of the properties of the speech signal. This is the topic of section 2. Section 3 conceptualises segmentation as a problem of L2 acquisition. Section 4 rephrases the acquisition issues from the perspective of the AIT.

2. How does the proficient listener hear words in the speech signal?

2.1. Hearing words depends on our phonological knowledge

Hearing words can be conceptualised as several distinct processes: 'segmentation', 'word activation' and 'word selection'. Segmentation requires identifying sound shapes ('formatives') in a continuous and highly variable speech signal. Psycholinguistic research suggests that proficient listeners are sensitive to language-specific properties which turn out to be the same sorts of properties identified by linguists in describing the phonetic and phonological systems of languages, notably, 'mora', 'syllables', 'feet' (Mehler and Christophe 1992). If we hypothesise that the speech processors of a

[2] It is not a form of conceptual learning, and, therefore, cannot result from an analysis of 'intake' as defined in the interactionist literature, see Gass (1997). If my account of segmentation is correct, then the interactionist approach cannot explain how we come to be able to segment formatives from the speech signal, nor how we acquire many kinds of phonetic knowledge.

particular user of a language actually implement processing procedures which incorporate these distinctions, it follows that they can only function efficiently after the relevant linguistic acquisition has occurred.

Once segmentation has taken place, hearing words, as opposed to hearing some arbitrary sounds such as [mimimimimi:], results from the activation of an appropriate lexical entry, which then makes available that formative's morphosyntactic and semantic properties. A word is selected when a specific lexical entry (as opposed to potential competitors) is integrated into the ongoing parse of a sentence. Thus, hearing words is no primitive operation, but rather the experience we have as a result of some rather complex phonetic and phonological parsing and word-recognition operations. There is, unfortunately, no consensus as to the definition of these processes. One paradigm hypothesises that speech processing involves operations which build structure (linear and hierarchical phonetic and phonological representations). An alternative paradigm proposes the direct activation of lexical entries on the basis of acoustic-phonetic features in so-called 'direct mappings' of the input-meaning relations (as the Competition Model hypothesises; see Bates and MacWhinney 1981; MacWhinney 1987, 1997, inter alia).

Four properties of speech make the direct-mapping model implausible. The first is the fact that the signal constitutes a continuum. Word boundaries are not part of the input to speech processors but rather result from the processing of the signal. One of the major tasks of the speech-perception system is thus to impose discrete units on the signal (segmentation). Second, our speech processors have to find the words within the units segmented from the signal. The edges of words in continuous speech are difficult to locate because they are not reliably cued by any necessary (universal) acoustic cues. A third fact making word recognition complicated is that a signal intended to communicate a given message can have highly variable acoustic properties. The fourth property at first presents a puzzle: processing is not deterministic but depends on processing mode. We are able to ignore the acoustic variation arising from co-articulatory effects when processing language in 'speech mode'. In speech mode, two acoustically quite different stimuli may be perceived by knowledgeable listeners as the same unit. This is what Fowler and Smith (1986) have referred to as the 'perceptual invariance problem'. Their solution to it is to propose that knowledgeable listeners analyse the acoustic signal using anticipatory co-articulatory information for a given phonetic segment and factor this information out of the signal.[3]

Studies of cross-linguistic perception (see Strange 1995), in particular, comparative studies of the development of the ability to detect and discriminate phonetic segments (Eimas 1990; Werker 1995; Jusczyk 1997) show quite clearly that the ability to impose perceptual invariance on the signal is based on knowledge of the grammar of the language and is not a consequence of general properties of auditory processing (Kuhl 1992). This filtering effect in speech processing is neither permanent nor rigidly present, but rather task dependent. Thus, we appear to be able to detect and discriminate many more acoustic differences when simply asked: 'Is the sound "$[x]_{syllable}$" the same as or different from the sound "$[y]_{syllable}$"?' When asked, in contrast, to discriminate sounds contained in words, our perceptual acuity is affected by the status of the acoustic distinction in the organisation of the lexical system. In short, when it is a matter of word recognition, we impose perceptual invariance on the signal based on our knowledge of the function of particular phonetic distinctions in defining formatives. This knowledge is acquired.

2.2. Hearing words is not due just to bottom-up analyses of the signal but depends also on the size and organisation of our mental lexicons

Word recognition is influenced by properties of the organisation of the listener's lexicon, in particular, by the potential 'confusability' of particular words, the size of the lexicon, and the organisation of words into 'lexical neighbourhoods' (Goldinger et al. 1989; Cluff and Luce 1990; Luce et al. 1990; Mehler et al. 1990; McQueen et al. 1995). People on the north shore of Lake Ontario systematically

[3] While the [d] sounds of *deed* and *dope* have very different acoustic properties, knowledgeable listeners of English hear them as the same sound. The acoustic variation is real but irrelevant in many speech contexts for recognising *deed* and distinguishing it from {*bead*, *feed*, ... *seed* ...} or for recognising *dope* and distinguishing it from {*hope*, *lope*, ... *grope* ...}.

mishear the speech of speakers on the south shore because those speakers tend to pronounce words like *John* with a fronted vowel and words like *Jan* with a raised vowel. Torontonians will, therefore, tend to hear a Buffalo speaker's rendition of *John saw Jan* as 'Jan saw Jen'.[4] There is not much 'space' for acoustic variability in the articulation of these syllables, as there are many words with which they can be confused: *June*, *Joan*, *Jane*, *Jen*, *gin*, *Jean*. The neighbourhood of words with similar segments is, in this case, crowded. Words that are polysyllabic have much smaller neighbourhoods so there is less chance of a longer word being confused with something else in the listener's lexicon. So when a Buffalo speaker produces *Canada* with a fronted vowel on the first syllable, the word will be heard with the intended referent – and simultaneously as an unintended 'funny' foreign pronunciation of the word. Our lexical entry of the word is activated and, simultaneously, we recognise [ˈkʰɛnərə] (or something similar) as instantiating a non-local accent. Our word-recognition systems are flexible enough to be able to adjust to a non-local pronunciation of words organised in single-item or small lexical neighbourhoods.

2.3. Modelling the addresses of lexical entries

The analysis of the speech signal provides one or more representations which can serve to make the contents of the lexical entry in long-term memory available for further processing. This moment in speech processing is referred to as lexical activation. There are various models of lexical activation. In some models, it is assumed that signal processing builds a representation which activates something comparable stored as the address to the word's lexical entry. The address might consist of an abstract phonological representation (Lahiri and Marslen-Wilson 1991, 1992), or it might consist of a fairly concrete and low-level acoustic representation (Klatt 1979, 1986) or indeed something in between. If the address consists of an abstract phonological representation, then one must postulate speech processors which will build such representations. If

[4] The fronting and raising of the relevant vowels is part of ongoing sound change affecting many varieties of North American English, see Labov et al. (2001).

the address consists of an acoustic representation, then one can dispense with the phonological processors. The trade-off comes in the recognition of the variability of the signal. The closer the address is to the properties of the signal, the more addresses will be needed in long-term memory to activate a word.

Models differ also in their assumptions about how lexical entries are selected. Some models operate 'serially', i.e., one hypothesises that speech processors first build a complete representation of the formative of the word which is then matched to an address in memory. Other models adopt the hypothesis of incremental and massive 'parallel' lexical activation, i.e., as phonological processes build a phonological representation, possible targets are activated, with subsequent deactivation occurring in all addresses which at some point in the sound parse of the signal start to deviate from the input in the signal. The listener would have the impression of hearing a word when there is a point in the processing of the signal when a unique lexical entry remains active. Marslen-Wilson's (1993) Cohort Model of word recognition thus views word recognition as a two-step process, consisting first of the activation of many addresses of lexical entries, and then the selection of a unique address.

3. Learning to hear words as a part of L2 word learning

Our notion of word learning must be enlarged to include the acquisition of precisely the kinds of phonetic and phonological knowledge which permit us to segment a prosodic unit from the speech signal. To date, the interests of SLA phonologists have focused more on the acquisition of knowledge of phonemes or of phonological generalisations (like stress shifts, vowel shortenings, vowel lengthening or consonantal feature changes). It is entirely possible, however, that this knowledge is acquired only after the learner has extracted and stored a formative in long-term memory. Space limitations preclude detailed discussion of this issue but one view of phonemic and phonological acquisition would entail encoding in long-term memory multiple formatives which are analysed and re-analysed offline, with relevant contrasts being slowly extracted. What is relevant at this point is the hypothesis that the acquisition which permits segmentation may constitute the very first stage of SLA. The learner must acquire enough phonetic knowledge to be able to do this before many other forms of phonological acquisition can take place.[5]

3.1. Do acquisition questions help us choose a model of word segmentation?

The plethora of processing models proposed to explain word recognition in knowledgeable and proficient language users is entirely unsatisfactory from an acquisition perspective. We would like to characterise language acquisition as both the encoding of grammatical representations and of those processing procedures which build structure (Carroll 2001). It is an explicit hypothesis of only two SLA theories, namely, the Competition Model and the AIT, that our models of language acquisition are dependent on assumptions about processing.[6] It follows that we need to make those assumptions explicit in order to be more precise about the nature and time course of acquisition. Surely there must be some reason for preferring a particular processing model over others.

Cutler (1996) has noted that models of word recognition have typically taken two quite distinct approaches to the problem of locating word boundaries in the signal, namely, those which postulate mechanisms whose function it is to impose word boundaries on the signal (Explicit Segmentation Models), versus those which propose that word boundaries implicitly emerge when lexical entries are selected (Serendipitous Segmentation Models). Explicit Segmentation Models hypothesise explicit procedures for identifying word boundaries. For example, Cutler and Norris (1988) and Cutler and Butterfield (1992) have postulated for

[5] I want to make it clear that I am *not* asserting that SLA learners first learn the sound system of a language before they learn everything else. That claim must surely be false since it requires learning the contrasts among distinctive units and the constraints on the combinability of units. This presumably occurs at the same time that the learner is acquiring semantic, pragmatic, morphological and syntactic information.

[6] Several recent proposals present models of language processing in the context of a theory of SLA, e.g., Gass (1988, 1997) and Towell and Hawkins (1994). However, the connection between processing and acquisition remains implicit in these theories.

English listeners a Strong Syllable Strategy which imposes a left word boundary immediately to the left of a strong syllable.[7] Mehler et al. (1981) have proposed that French listeners have acquired a Syllabic Segmentation Strategy. Otake et al. (1993) have proposed that Japanese listeners have acquired a Moraic Segmentation Strategy. Such Explicit Segmentation Models regard hearing words as a language-specific phenomenon because the strategies directly reflect phonological properties of words in the language, as well as the relative frequency of specific types of patterns in the input. In the case of Serendipitous Segmentation Models, the word boundaries are obtained 'for free' when lexical entries are activated and selected. One might think that this is an advantage in that it eliminates a particular type of processing but Cutler observes (1996: 88) that only the Explicit Segmentation Models can explain how words are initially segmented in L1 acquisition because the infant cannot rely on lexical knowledge to find word boundaries before it has acquired words. Mehler and Christophe (1992) make the same point. Similar arguments can be extended to SLA. Since Serendipitous Segmentation Models cannot explain how language acquisition might occur, we should prefer Explicit Segmentation Models of word recognition. For SLA, this has the consequence that we must postulate language-specific segmentation strategies induced on the basis of the phonological properties of the L1 lexicon and the relative frequency of particular prosodic patterns of the L1. In what follows I will simply assume that such strategies exist. In addition, I re-conceptualise SLA as involving both the acquisition of L2 knowledge (representations) as well as the acquisition of L2-appropriate segmentation strategies.

3.2. Learning to hear new formatives in an L2

It could be objected that the problem of learning to segment formatives from the speech stream for the pre-language infant and the L2 learner are not equivalent. Thus, the L2 learner does have an

[7] Strong syllables may bear primary stress, have full rather than reduced vowels, are longer than weak syllables and can be the locus of a tone alignment.

extant lexicon and may have specific expectations about the structure of words. At the very least, the L2 learner might expect formatives to exist and to map onto morphological and syntactic classes and conceptual representations. The banal observation that L2 learners regularly ask: 'What is the word for ... [insert any favourite translation equivalent here] in language X?' provides concrete evidence for the hypothesis that they do. Why, then, could not prior knowledge, including specific L1-grammatical knowledge, guide the learner to hear words in the L2? It does. Cutler et al. (1992) have shown that English–French bilinguals may adopt either the Strong Syllable Strategy (the same strategy as monolingual Anglophones) or the Syllable Segmentation Strategy (the same strategy as monolingual Francophones) when listening to *both* English and French stimuli. In other words, a single strategy is applied to speech stimuli regardless of the source. Although more studies are required, we may tentatively conclude that bilinguals deploy one segmentation strategy, based on exposure to the prosodic properties of their dominant language, and transfer it to their weaker one. It stands to reason that L2 learners, listeners who are in the process of acquiring knowledge of the target language, will transfer their L1 segmentation strategy to the target language input as well.

In a certain sense, this is exactly what we would expect given the observation that learners often cannot parse the L2 on first exposure, for that result also derives from the transfer of L1 parsing procedures. In many learning situations, the L2 learner first hears the signal as a continuous stream of noise. Only subsequently (sometimes weeks or months later) does the learner hear the L2 signal as bits of recognisable sounds. This, again, banal observation is readily explained by the hypothesis that segmentation strategies automatically transfer. What the work by Cutler et al. (1992) suggests, however, is that the segmentation strategies apply long after the learner has acquired the necessary phonetic and prosodic knowledge to hear syllables and other prosodic units in the speech stream. In other words, acoustic-phonetic acquisition can take place and still the learner might continue to transfer segmentation strategies appropriate to the L1. If substantiated by additional research, such studies will show that the acquisition of the knowledge leading to segmentation might be modular in the sense of Fodor (1983): segmentation strategies apply automatically, independently of the beliefs or attitudes of the listener, are acquired unconsciously, and may not be alterable. It even suggests that segmentation strategies apply independently of the bilingual's phonological knowledge. While Cutler and her colleagues were not interested in collecting relevant data from their bilingual subjects, we know from SLA research that L2 users can acquire subtle and sophisticated kinds of knowledge about the phonology of the L2, including knowledge of possible and impossible syllables, patterns of stress placement, and so on. Since Cutler's subjects were described as highly sophisticated L2 users with pronunciation hardly distinguishable from that of native speakers, it seems reasonable to suppose that they would also have possessed this kind of phonological knowledge. If so, some of them possessed rich knowledge of the L2 phonology and still parsed the L2 with L1-appropriate segmentation strategies. This could only happen if the phonetic parsing which results in segmentation can occur uninfluenced by higher-order phonological knowledge.

How does the learner make the transition from the stage of hearing incomprehensible noise to the stage when she can hear some sequence of syllables? There is precious little relevant research to draw on.[8] Some psycholinguists, e.g., Morgan and Demuth (1996), hypothesise that all learners are sensitive to pause as a cue to a word boundary. If true, we might hypothesise that words which are located either immediately following or immediately preceding a pause would be precisely the words whose formatives would be segmented from the speech stream first, permitting the creation of an initial L2 vocabulary (see Hatch 1983 for similar claims). We could then predict that new formatives would first be learned in utterance-initial or utterance-final position. Rast (2003: 260–272) and Rast and Dommergues (2003: 145) in an input study with beginner learners using a word repetition task showed that words in sentence-medial position were far less likely to be repeated than

[8] Rast's doctoral dissertation (2003) and Rast and Dommergues (2003) are worth mentioning here. They deal with word learning (recognition and production) in French learners of Polish during the first eight hours of exposure.

those in sentence-initial or sentence-final position. Although not directly related to word-learning, two additional studies have investigated the perception of words, stress perception and sentence position among L2 learners using Spanish stimuli. Barcroft and VanPatten (1997) and Rosa and O'Neill (1998) show positional sensitivities in processing stressed words and argue that sentence-initial position is more salient than either sentence-final position or sentence-medial position. VanPatten couches their results as the Sentence Location Principle (VanPatten 2003: 10). There are, however, reasons to question these conclusions, when couched as an initial and universal sensitivity to stress.

3.3. Focus and segmentation: three empirical problems

In current work with my colleague Ruben van de Vijver, I am exploring the idea that focus, especially as instantiated by prosodic prominence, provides the initial context for segmenting formatives from the signal.[9] We see the problem of learning how to segment words as having several parts: determining (i) the domain in which segmentation begins, (ii) the phonetic properties learners are initially sensitive to, (iii) how learners impose the 'edges' of phonological units onto the signal, (iv) how they map syllables or syllable sequences onto morphemes, and (v) what properties they transfer from the L1 at particular developmental phases.

3.3.1. Do learners exhibit an initial preference for a particular domain in which to begin segmenting formatives?

We hypothesise that the learner must first begin to detect novel acoustic properties relevant to segmenting a linguistic unit (by hypothesis, a syllable or mora) in some relatively small domain of the utterance, given well known limitations on working memory.

[9] This research is being conducted within Project C4 'Prosody and Information Structure as forms of 'input' in second language acquisition', part of the Collaborative Research Centre 632 – *Information Structure: Linguistic markers of the organisation of utterance, sentence and text*, which began on 1 July 2003. The Collaborative Research Centre is funded jointly by the German Research Foundation, the University of Potsdam and the Humboldt University in Berlin.

We hypothesise that there must be an initial preference for some particular part of the utterance, as the logic of the Sentence Location Principle makes clear. In contrast to VanPatten, however, we see no reason to assume that learners will prefer the left over the right end of the sentence as the initial domain of attention. Rast's results suggest that learners may not exhibit such a preference in segmentation, either. Indeed, we think the notion of 'sentence end' may be formally undefinable. Our hypothesis, rather, is that speakers will be sensitive to domains defined by information structure, particularly focus marking. Focus is a distinction of conceptual structure but interacts with prosodic structure and syntactic structure in interesting ways. The markers of focus will define the relevant domains in which word segmentation will begin and learners will be sensitive to the realisation of focus markers.

We are most interested in the domains defined by prosodic markers of focus in a language like English. As for syntax, in English focus tends to occur at the right end of declaratives and at the left edge of wh questions. If, in the initial stage of learning, the learner receives more exposure to wh questions than to declaratives, then he or she might show a preference for words occurring at the left edge of sentences. If, on the other hand, the learner receives more exposure to declaratives, then he or she might show a preference for words at the right edge of sentences. Both possibilities need to be carefully explored, but we anticipate input studies to reveal that L2 learners get a lot of exposure to both questions and declaratives.

Focus marking also interacts with phrase-final lengthening. Numerous studies on markers of focus show that the ends of phrases of simple declarative sentences in English are marked by syllable lengthening (Cooper and Paccia-Cooper 1980; Warren 1985).[10] This literature has traditionally been concerned with figuring out how phonetic cues might permit a direct mapping to

[10] Not all simple declarative sentences exhibit phrase-final lengthening on the final stressed syllable of the sentence. Tonic location itself depends on such things as the presence or absence of temporal and locative adjuncts and the use of lexemes versus pronouns. There is, consequently, no reason to think that the right edge of a declarative sentence will uniformly present the learner with an accented word to be segmented from the signal.

syntactic structure. Phrase-final lengthening is interesting for other reasons. Words which are lengthened contain syllables which are lengthened. Lengthened syllables are less likely to vary from a prototypical articulation of consonants and vowels. Lengthened syllables might offer learners a better and more stable input. Focus, the location of the tonic accent and lengthening of the accented syllable, will potentially interact with word structure. Monosyllables would presumably be segmentable earlier than polysyllabic words because their formative would consist of a single syllable lengthened under focus in phrase-final position. Interestingly, Rast (2003: 251, 279) and Rast and Dommergues (2003: 141, 148) report no effect of the variable 'length of word' (measured on words of one, two and three syllables) but an interaction of length of word and sentence position such that words of one or two syllables were much easier to repeat in sentence-initial and sentence-final position.

3.3.2. What phonetic features is the learner initially sensitive to?

We are working with the hypothesis that acoustic features are used to construct pre-lexical representations which consist of (at least) strings of phonological units such as mora or syllables. Once the L2 learner can start localising cues to the edges of syllables, he may begin transferring structural units of the L2 phonology to the task of analysing segmented syllables into hierarchical structures, i.e., deploying L1 segmentation strategies. Before learners can transfer L1 segmentation strategies, however, they must construct representations of the syllable. The second and third parts of the segmentation problem involve explaining exactly what phonetic features are being extracted as the cues to a syllable (in the preferred domain) and how these features align to syllable boundaries. If learners are constructing syllables on the basis of the acoustic properties of the signal, then they must first identify the edges of syllables. If learners are identifying syllables on the basis of the cues to syllable nuclei, then they will show an initial preference for cues to the periodicity and duration of vowels (the unmarked syllable nucleus). How might this work?

We hypothesise that users of languages with intonation are sensitive right from the start of L2 acquisition to the prosodic realisations of focus. In English this includes, pitch movement, loudness peak, crescendo, decrescendo and various combinations of these features (Wells 1986).[11] Anglophones are sensitive to these phonetic properties when processing English. We hypothesise, following the assumptions of the AIT, that learners will automatically transfer this sensitivity in the form of L1 processing strategies. These cues to tonic accent are not, however, universal. Speakers of even a closely related language like German or Dutch would have to acquire these cues or new relevant configurations of cues to locate focus in English utterances (Terken and Nooteboom 1987). In our research we are working with German learners of English precisely because these languages are phonetically and phonologically well studied, making it possible for us to formulate precise hypotheses about transfer. Since it takes prior learning of individual cues to learn cue configurations, we hypothesise that learners will be sensitive first to pitch movement and only later to cue configurations. Whether this is the right hypothesis and how the acquisition of the full configuration of cues to focus in English proceeds is an open question at the moment.

Hatch (1983) and others have speculated that all L2 learners are sensitive to accent and stress. This is unlikely in that stress is not a universal property. However, there are some interesting things to note about accent and stress. Accented words in languages like English, Dutch and German are stressed words. Locating an accented word would automatically force the learner to process a stressed word. These languages are also stress-timed languages, meaning that sequences of stressed and unstressed syllables are hierarchically structured into larger rhythmic constituents. A given syllable will be longer when it is stressed than when it is unstressed. Stressed words will be longer than unstressed words because the

[11] What the phonetic correlates of focus in other languages are is not well known. It is also important to note that prosodic prominence does not appear to be one of the markers of focus in languages with lexical tone. Such languages appear to use syntactic or lexical markers of focus. Whether L1 speakers of Chinese would be equally sensitive to prosodically prominent focus domains for the initial segmentation of syllables is an interesting question.

stressed syllable is longer than the unstressed syllable. Stressed syllables, like accented syllables, appear to show less variability in their pronunciation than unstressed syllables, which might speed up the creation of a stable representation of the formative in long-term memory relative to unstressed syllables. There is already anecdotal evidence that L2 learners learn stressed syllables before they learn unstressed ones (including the fact that functional categories, which are typically monosyllabic and stressless, are learned after lexical categories). Rast (2003: 252) and Rast and Dommergues (2003: 142) show that their subjects repeated stressed words more accurately than unstressed words. It might ultimately turn out to be the case that the stressed syllables which learners produce first were first acquired in accented positions. In that case, accent would facilitate the learning of stressed syllables and stressed (monosyllabic) words. What the role of focus marking would be in languages which are not stress-timed is a question for further research.

3.3.3. What properties do learners transfer?

The final parts of the problem are, How are segmented formatives preferentially mapped onto morphemes? How does transferred knowledge interact with newly acquired knowledge? Do learners, regardless of the L1–L2 pair, initially show a preference for segmenting disyllabic words, meaning that a morpheme is preferentially a formative with two syllables (Broselow et al. 1998)? Or should we expect learners of L1s with agglutinative morphologies to show different initial preferences given the complex correspondences between their prosodic words and the morphemes of the L1? Cutler's research and the Broselow et al. study suggest that knowledge of preferred L1 prosodic word–morpheme correspondences may guide early word learning.

What is the interaction between acquisition and transfer in shaping the course of development? Once segmentation is possible, do the scales immediately fall from our ears (as it were)? Or do we continue to hear much speech as noise? Little research exists on the exact time course of increased proficiency in segmentation. However, we do know that we can expect a differentiated picture of formative learning. Research on cognates suggests that properties of phonological representations play an important role in explaining how L2 stimuli can activate L1 lexical entries in reading (Dijkstra et al. 1999). This hypothesis is confirmed in Rast's study (Rast 2003: 255–60; Rast and Dommergues 2003: 143). Phonetic information from the signal, although quite different from the pronunciation of the L1 word, may be sufficient to activate the L1 lexical entry and trigger word recognition. What counts as phonetic similarity in cognate pairing is still not clear and this too is a question requiring empirical study. I anticipate that the structure of the L1 lexicon and typological properties of language pairs will have an independent influence on the course of word learning subsequent to the emergence of segmentation capacities based on phonetic learning.

A distinct question is, Are some languages just easier to segment than others because they have 'unmarked' (consonant–vowel) syllables? Or is segmentation always a matter of the particular phonetics of L1 and L2 pairs? The issue of marked-versus-unmarked syllable structures has received a good airing in the L2 phonological literature (Flege and Wang 1989; Eckman 1991; Edge 1991; Major 1996; Major and Faudree 1996; Broselow et al. 1998). What needs to be sorted out is the connection between L2 perceptual learning, preferences for unmarked syllables and the acoustic cues to syllables.

4. The Autonomous Induction Theory (AIT)

4.1. SLA in the context of speech processing

The discussion above has made reference to the Autonomous Induction Theory (Carroll 1999b, 2001, 2002a, b). This is a theory of SLA which incorporates explicit assumptions about Universal Grammar as a set of constraints on the Language Acquisition Device (LAD) but puts the emphasis on acquisition, not universals.[12] The AIT integrates acquisition within a processing framework

[12] In this, it is quite distinct from proposals in Herschensohn (2000) or White (2003). In particular, the AIT rejects the claim that there are such things as parameters functioning as mechanisms of acquisition. If parameters can be construed as constraints on acquisition, it is at a much more abstract level of generalisation than that which the AIT is operating at.

but, unlike the Competition Model, eschews direct mapping. Rather, the structure of representations plays a crucial role in explaining the grammatical development. It is a mixed model in that it hypothesises that both structure building and activation levels are necessary components of language acquisition and language processing. Structure building is required to explain novel content in linguistic representations. Changing activation thresholds is necessary to explain why L1 and novel processing procedures can compete in various psycholinguistic tasks. Thus, the AIT is designed to explain not only the content of linguistic representations but the fact that there are preferred parsing strategies and differences in activation levels of particular sorts of representations. This second property sets it apart from all other SLA theories to emerge from within a generative paradigm.

The AIT states that learners create novel linguistic representations from existing ingredients. These ingredients may be primitives provided by our linguistic endowment (UG) or they may be complex entities which are part of an extant grammatical representation (the transferred contents of an L1 lexical entry). The trigger for the LAD is parsing failure. The LAD functions to repair breakdowns in the analysis of the speech signal when existing parsing procedures are inadequate for the real-time analysis of a novel stimulus. Repairs will take various forms depending on the nature of the parse failure. What counts as a parse failure is dependent on the parsing model adopted. All learning problems are conceptualised in terms of specific levels of analysis (phonetic, phonological, morphosyntactic or semantic). More specifically, all learning problems are conceptualised in terms of structure building in a particular internal code (for ease of exposition, let us assume that this amounts to tree building) or in stating equivalences ('correspondences') across internal codes. In short, certain parsers must be able to attach a novel unit to a tree structure being built; other parsers must put a unit in one code in correspondence with a unit in another code. The LAD is activated when a parser cannot do its job because relevant representations or procedures are lacking. Input to the LAD can come from a lower level of analysis of a stimulus or from a higher level. Input to the LAD can also come from representations stored in long-term memory. The AIT is thus quite different from other proposals which attempt to provide a unified approach to perception, parsing and learning. The AIT treats them as distinct cognitive problems.

4.2. The processing assumptions of the AIT

The AIT builds directly on proposals by Jackendoff (1983, 1987, 1997) regarding the architecture of the language faculty. Jackendoff models the language faculty as a variety of modular processors, corresponding more or less to the traditional levels of linguistic analysis, and the links between them. The theory postulates two basic types of processors, 'integrative processors', which build linguistic representations of a given type, and 'correspondence processors', which equate a unit of one linguistic representational format with a unit of another. Thus, processing procedures within an integrative parser build level-appropriate mental representations — acoustic, phonological, morphosyntactic or conceptual representations, depending on the module. The task of integrative processors is to integrate features and constituents into larger hierarchical representations. Inputs to the integrative processors must, therefore, be expressed in the right representational format. A morphosyntactic processor cannot analyse patterns of periodic energy, only the acoustic processor can 'read' such inputs. Conversely, an acoustic processor cannot analyse configurations like [+N, –V, +feminine]; only a morphosyntactic processor can do that. Each processor analyses inputs appropriate to it and builds a representation given processor-appropriate processing strategies able to analyse the input. The resulting representation is the output of the processor.

The hierarchical structures which the integrative processors put out are language specific. They make use of universal constraints in their operations but the structures they build are unique to the language in question. This must be true because languages vary in the word classes they deploy, the syllable types they exhibit, or the sets of syntactic constructions they possess. We saw in section 2 that the language-specific units of linguistic description are directly relevant to a characterisation of speech processing. While the linguist–reader might take this as self-evident, the point bears emphasis because the superiority of speech-processing models which incorporate linguistic representations over direct-mapping models such as the Competition Model hinges precisely on this evidence.

Languages also vary in the ways in which units of one level of representation map onto or correspond to units of a level higher up or lower down. In Jackendoff's theory, making links across levels is the job of the correspondence processors. Thus, the correspondence processor mapping phonological representations of English onto morphosyntactic ones will include a correspondence schema like that shown in (1):

(1) The prosodic-word – morphosyntactic-word correspondence

$$(\quad)_{\text{Prosodic Foot}} \Rightarrow [\qquad]_{[\alpha N, \beta V]}$$

This correspondence rule incorporates part of the information expressed as the Strong Syllable Segmentation Strategy insofar as the Prosodic Foot will be a constituent defined as a sequence of strong and weak syllables, e.g., $(\sigma_s)_{\text{Prosodic Foot}}$, $(\sigma_s\sigma_w)_{\text{Prosodic Foot}}$ or $(\sigma_s\sigma_w\sigma_w)_{\text{Prosodic Foot}}$. It states that strong syllables mark the left edges of words. The word here is more precisely defined as a morphosyntactic entity. A separate correspondence rule is needed to map a morphosyntactic lexical unit onto a conceptual unit of a particular sort, i.e., onto word meaning. The correspondence rule in (1) takes the output of the phonological parser and provides the morphosyntactic integrative processor with a unit (analysed in terms of the relevant set of primitives) for integration into a morphosyntactic parse. Correspondence processors thus move an analysis of a given stimulus up or down the levels of analysis (from signal to semantics or from semantics to the phonology).

The content of (1) is acquired. Anglophones clearly have internalised this information. Francophones and Japanese listeners have quite different correspondence procedures, tuned to the properties of the phonology of French and Japanese. When they begin to learn English as an L2, they will have to develop a new correspondence procedure, namely (1), if they are learn to segment the speech stream as native speakers do. Cutler and her colleagues have suggested that even fluent bilinguals do not make use of (1) when their dominant language utilises another correspondence. It remains to be seen, however, whether French or Japanese L2 users of English have no correspondence rule like (1) at all at their disposal, or whether (1) is, for some language users, simply more speedily activated and deployed.

One of my objectives in formulating the AIT was to develop a constrained theory of SLA. It is important to understand, therefore, that some of the constraints arise from the adoption of the hypothesis of Representational Modularity (Jackendoff 1987). Since a parsing problem will arise within a given parser (either one of the integrative processors or one of the correspondence processors), a solution can only be found within this processor. Within the integrative processors, this will mean that a feature or constituent cannot be integrated into a hierarchical structure because no procedure can be transferred from the L1 for doing this. Similarly, within a correspondence procedure, a parse will fail because no appropriate correspondence mapping exists. Solutions will always depend on which processor has the analytical problem: the LAD will have to integrate a novel constituent into a structure, perhaps adding a new processing procedure to the repertoire of a given integrative processor to do so; it may have to add additional correspondence rules to the repertoire of a given correspondence processor. Additionally, the solution to the parsing problem will be constrained by the very specific properties of the problematic input string. In the case of creating pre-lexical strings to hear words, the solution will involve representing the appropriate phonetic features needed to encode syllables. This might involve encoding novel features from the repertoire UG makes available into novel complexes of features (learning novel cues to structure) or it might involve re-weighting a set of combinations of universal features.

The theory is so conceptualised that initial solutions can be input specific and generalise over time. Whether this degree of input sensitivity is actually warranted is an empirical question. For certain learning problems, learners at the initial stage of exposure are quite able to formulate general solutions on the basis of minimal evidence (Carroll 1999a), but I have seen no empirical data to date which can answer the question directly for the segmentation problem.[13] However, as long as the AIT permits such input-specific solutions, the LAD must construct multiple solutions to a given parse problem which then compete with one another. Successful solutions are those which are deployed over and over again for the same parse problem. Successful solutions will also generalise when they can make use of shared parts of the 'If X, then Y' procedures which are part of the LAD.[14] Other constraints have been postulated as well. See Carroll (2001, 2002a, b) for more details.

4.3. Learning how to hear words in the signal

To return to the issue with which we began, the AIT will treat learning how to hear words as a series of distinct acquisition tasks: (i) learning the distribution of sets of acoustic properties reflecting edges of syllables in some initially privileged domain (focus? post-pause? pre-pause? locus of a kinetic tone?) if transferred L1 procedures for segmenting minimal prosodic units (mora, syllable or feet) fail; (ii) once syllables can be imposed on the signal, transferring L1-based prosodic procedures for the construction of a minimal prosodic word if possible, learning new prosodic procedures if not; (iii) mapping these initial minimal prosodic words onto minimal morphosyntactic words (morphemes) under conditions which may well reflect the learner's communicative needs; (iv) elaborating at each level of modular analysis the distribution of the features and units to meet the complexity of the input to the processor; (v) elaborating the sets of correspondences between

[13] This applies even to Rast's (2003) work which is formulated in terms of fairly crude parameters of phonetic and phonological analysis.

[14] The formal assumptions involve an adaptation of important work on induction by Holland et al. (1986), a symbolic (non-connectionist) approach to learning. Some readers will find my adoption of this approach as hopelessly unfashionable. So be it, there are numerous empirical problems with spreading activation networks which have yet to be addressed. The central problem they have been designed to deal with, namely, the so-called 'brittleness' of symbolic computational systems, can be dealt with in other ways.

the levels of analysis so that the cues for units at a higher level up come to match the complexity of the input to the processors; (vi) automating the procedures through successful competition in repeated processing of L2 speech.

With respect to the problem of learning new phonetic cues to minimal prosodic units (mora or syllable), the AIT makes no precise predictions as to the details of potential sensitivities. In other words, it does not predict that learners will be more sensitive to pitch movement than pause even if this is currently our working hypothesis. The AIT will survive should it turn out to be the case that learners are more sensitive to pause or even to amplitude. It does, however, predict that regardless of the L1–L2 pairs, we ought to find certain phonetic primitives of just this sort exhibited universally in segmentation processes. Moreover, some of these primitives should be specific to linguistic cognition and not result from properties of auditory processing in general (contra, I think, the requirements of General Nativism, on which see O'Grady 1999, 2003). In addition, the theory predicts that learners will exhibit preferences for primitives as markers of grammatical properties of languages and only later (if at all) acquire the pragmatic and cultural meanings of pause, or tones. The AIT also predicts that these pragmatic and cultural meanings will have no influence on the ways in which the linguistic cues are deployed in real-time parsing.

Since the acquisition tasks are conceptualised as involving distinct parts, mastery will take time and considerable amounts of relevant stimuli. Acquisition is construed here as a matter of representation but proficiency is a matter of automaticity and control. In neither case is the success of the outcome guaranteed. SLA, in this theory, is neither automatic nor inevitable even when constrained by UG. The theory thus predicts variability in the knowledge of individual L2 learners. Studies of the syntax of interlanguage grammars have shown that representations can be incomplete or different (Sorace 1993a, b). I see no reason not to assume that the same will hold true of phonetic and phonological knowledge. The theory also predicts individual differences in the ability to activate novel representations on precise linguistic and psycholinguistic tasks dependent on the degree of interference (competition) from existing L1 parsing procedures.

5. Conclusions

In this paper I have attempted to micro-analyse learning how to hear words in the speech stream, drawing on relevant psycholinguistic studies on the segmentation of words by monolinguals and highly proficient L2 users, the nature of lexical organisation, word activation and word recognition, and the few relevant L2 acquisition studies which exist. The phenomenology of SLA tells us that there are various stages in the learning process. In many cases, on first exposure, we hear the L2 as a continuous stream of unidentifiable noises. This must be, for many learners, the absolute first stage of acquisition. To get from this stage to the next, where the learner has the impression of hearing identifiable sounds, the learner has to learn how to represent L2 phonetic cues to phonological representations (formatives) given novel acoustic information. In particular, I hypothesise that the learner has to come to represent a stimulus as a string of syllables (or mora). Once the learner has acquired phonetic knowledge relevant to the location of syllable edges, she will transfer L1 segmentation strategies. If these are useful for parsing, they will persist. Should they be inadequate for analysing L2 stimuli, parsing will fail and the LAD must create alternative solutions. Once the learner has segmented some pre-lexical representations, he will be able to hear identifiable sounds (rather than noise), but these will only be 'words' when the segmented sound sequences are stored in long-term memory as the formative of a word, and put into correspondence with morphosyntactic and semantic information. Segmentation is thus the first step in acquiring L2 vocabulary.

These microprocesses are framed within the Autonomous Induction Theory, a novel SLA theory rooted in Jackendoff's tripartite modular functional architecture of linguistic competence. The AIT is entirely novel, however, in that it goes well beyond this architecture to propose an account of how novel grammatical knowledge and novel parsing abilities can arise, consistent with the empirical studies cited. It is, moreover, highly constrained and limits alterations to extant representations in the grammar to the locus of parse failure and to specific sets of operations made available by UG. My current research on the acquisition of segmentation construes the problem of learning how to segment formatives as the specification of domains of initial sensitivity to phonetic properties in the signal, the specification of the phonetic features or configurations of features to which learners are initially sensitive, preferred correspondences between phonetic features and the edges of prosodic units (mora, syllable), preferred correspondences between segmented formatives and morphemes, and the amount and kind of transfer from the L1 which occurs. The AIT makes no specific predictions as to the learners' initial sensitivity to particular phonetic features, although we hypothesise that kinetic tone will be important. More importantly, we hypothesise that learners are sensitive to the prosodic exponents of focus, which will, in turn, delimit the domain in which first formatives are segmented. Learners will be sensitive to focus and focus will define the locus where the learner's initial processing of the signal will begin.[15] Focus has various linguistic markers. In a language like English, these include tonic accent, realised by a rich complex of phonetic properties which will, by hypothesis, facilitate the encoding of a stable phonetic and phonological representation of a syllable. Both the formal task of spelling out how words are segmented from the speech stream over the time course of learning an L2, and documenting what occurs empirically is a formidable research task. My colleague Ruben van de Vijver and I have just begun the task of sorting out the learnability issues, discussed here in some detail. The presentation of the facts of the matter must await another occasion.

Institute for English and American Studies
University of Potsdam
P.F. 60 15 53, 14415 Potsdam
Germany
Email: carroll@rz.uni-potsdam.de

[15] We note that Anne Cutler's research group on multilingual processing is now investigating similar questions (Akker and Cutler 2003).

References

Akker, Evelien & Cutler, Anne, 2003. 'Prosodic cues to semantic structure in native and non-native listening', *Bilingualism: Language and Cognition* 6, 81–96.

Baker, Colin, 1993. *Foundations of Bilingual Education and Bilingualism*, Clevedon: Multilingual Matters.

Barcroft, John & VanPatten, Bill, 1997. 'Acoustic salience of grammatical forms: the effect of location, stress and boundedness on Spanish L2 input processing', in Ana Perez-Leroux and William R. Glass (eds.), *Contemporary Perspectives on the Acquisition of Spanish*, Vol. 1, Somerville, MA: Cascadilla Press, 109–121.

Bates, Elizabeth & MacWhinney, Brian, 1981. 'Second language acquisition from a functionalist perspectives: pragmatic, semantic and perceptual strategies', in Henry Winitz (ed.), *Native and Foreign Language Acquisition. Annals of the New York Academy of Sciences*, Vol. 379, New York: New York Academy of Sciences, 190–214.

Broselow, E., Chen, S.-L. & Wang, C., 1998. 'The emergence of the unmarked in second language phonology', *Studies in Second Language Acquisition* 20, 261–280.

Carroll, Susanne, 1999a. 'Adults' sensitivity to different sorts of input', *Language Learning* 49, 37–92.

Carroll, Susanne, 1999b. 'Putting "input" in its proper place', *Second Language Research* 15, 337–388.

Carroll, Susanne, 2001. *Input and Evidence: the Raw Material of Second Language Acquisition*, Amsterdam: Benjamins.

Carroll, Susanne, 2002a. 'I-learning', in Susan Foster-Cohen, Tanja Ruthenberg and Marie-Louise Poschen (eds.), *Eurosla Yearbook*, Vol. 2, Amsterdam: Benjamins, 7–28.

Carroll, Susanne, 2002b. 'Induction in a modular learner', *Second Language Research* 18, 224–249.

Cluff, M. S. & Luce, P. A., 1990. 'Similarity neighborhoods of spoken two-syllable words: retroactive effects on multiple activation', *Journal of Experimental Psychology: Human Perception & Performance* 16, 551–563.

Cooper, W. E. & Paccia-Cooper, J., 1980. *Syntax and Speech*, Cambridge, MA: Harvard University Press.

Cutler, Anne, 1996. 'Prosody and the word boundary problem', in James Morgan and Katherine Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, Mahwah, NJ: Erlbaum, 87–99.

Cutler, A. & Butterfield, S., 1992. 'Rhythmic cues to speech segmentation: evidence from juncture misperception', *Journal of Memory & Language* 31, 218–236.

Cutler, Anne, Mehler, Jacques, Norris, Dennis G. & Segui, Juan, 1992. 'The monolingual nature of speech segmentation by bilinguals', *Cognitive Psychology* 24, 381–410.

Cutler, Anne & Norris, Dennis, 1988. 'The role of strong syllables in segmentation for lexical access', *Journal of Experimental Psychology: human perception and performance* 14, 113–121.

Dijkstra, T., Grainger, J. & van Heuven, W. J. B., 1999. 'Recognition of cognates and interlingual homographs: the neglected role of phonology', *Journal of Memory & Language* 41, 496–518.

Eckman, Fred, 1991. 'The Structural Conformity Hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners', *Studies in Second Language Acquisition* 13, 23–41.

Edge, B. A., 1991. 'The production of word-final voiced obstruents in English by L1 speakers of Japanese and Cantonese', *Studies in Second Language Acquisition* 13, 23–41.

Eimas, Peter, 1990. 'Comment: some effects of language acquisition on speech perception', in I.G. Mattingly and M. Studdert-Kennedy (eds.), *Modularity and the Motor Theory of Speech Perception*, Hillsdale, NJ: Erlbaum, 111–116.

Flege, J. E. & Wang, C., 1989. 'Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/–/d/ contrast', *Journal of Phonetics* 17, 299–315.

Fodor, Jerry, 1983. *Modularity of Mind*, Cambridge, MA: MIT Press.

Fowler, C. A. & Smith, M., 1986. 'Speech perception as 'vector analysis': an approach to the problem of invariance and variability', in Samuel Perkell and Dennis Klatt (eds.), *Invariance and Variability in Speech Processes*, Hillsdale, NJ: Erlbaum, 123–139.

Gass, Susan, 1988. 'Integrating research areas: a framework for second language studies', *Applied Linguistics* 9, 198–217.

Gass, Susan, 1997. *Input, Interaction and the Second Language Learner*, Mahwah, NJ: Erlbaum.

Goldinger, S. D., Pisoni, P. A. & Luce, D. B., 1989. 'Priming lexical neighbors of spoken words: effects of competition and inhibition', *Journal of Memory & Language* 28, 501–518.

Hatch, Evelyn, 1983. 'Simplified input and second language acquisition', in Roger Andersen (ed.), *Pidginisation and Creolisation as Language Acquisition*, Rowley, MA: Newbury House, 64–88.

Herschensohn, Julia, 2000. *The Second Time Around: Minimalism and L2 Acquisition*, Amsterdam: Benjamins.

Holland, John, Holyoak, Keith, Nesbitt, Richard & Thagard, Paul, 1986. *Induction: Processes of Inferencing, Learning and Discovery*, Cambridge, MA: MIT Press.

Jackendoff, Ray, 1983. *Semantics and Cognition*, Cambridge, MA: MIT Press.

Jackendoff, Ray, 1987. *Consciousness and the Computational Mind*, Cambridge, MA: MIT Press.

Jackendoff, Ray, 1997. *The Architecture of the Language Faculty*, Cambridge, MA: MIT Press.

Jusczyk, Peter, 1997. *The Discovery of Spoken Language*, Cambridge, MA: MIT Press.

Klatt, Dennis, 1979. 'Speech perception: the role of acoustic-phonetic analysis and lexical access', in R. A. Cole (ed.), *Perception and Production of Fluent Speech*, Hillsdale, NJ: Erlbaum, 243–288.

Klatt, Dennis, 1986. 'The problem of variability in speech recognition and in models of speech perception', in Samuel Perkell and Dennis Klatt (eds.), *Invariance and Variability in Speech Processes*, Hillsdale, NJ: Erlbaum, 300–319.

Kuhl, Patricia, 1992. 'Speech prototypes: studies on the nature, function, ontogeny and phylogeny of the 'centres' of speech categories', in Y. Tohkura, E. Vatikiotis-Bateson and G. Sagisaka (eds.), *Speech Perception, Production and Linguistic Structure*, Amsterdam: IOS Press, 237–264.

LABOV, W., ASH, S. & BOBERG, C., 2001. *Atlas of North American English: Phonetics, Phonology and Sound Change*, Berlin: Mouton de Gruyter.

LAHIRI, ADITI & MARSLEN-WILSON, WILLIAM, 1991. 'The mental representation of lexical form: a phonological approach to the recognition lexicon', *Cognition* 38, 245–294.

LAHIRI, ADITI & MARSLEN-WILSON, WILLIAM, 1992. 'Lexical processing and phonological representation', in D.R. Ladd and G.J. Docherty (eds.), *Papers in Laboratory Phonology: Gesture, Segment, Prosody*, Cambridge: Cambridge University Press, 229–260.

LUCE, P. A., PISONI, D. B. & GOLDINGER, S. D., 1990. 'Similarity neighborhoods of spoken words', in Gerry Altmann (ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, Cambridge, MA: MIT Press, 122–147.

MACWHINNEY, BRIAN, 1987. 'Applying the Competition Model to bilingualism', *Applied Psycholinguistics* 8, 315–327.

MACWHINNEY, BRIAN, 1997. 'Second language acquisition and the Competition Model', in A. M. B. de Groot and J. F. Kroll (eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives*, Mahwah, NJ: Erlbaum, 113–442.

MAJOR, ROY, 1996. 'Markedness in second language acquisition of consonant clusters', in Robert Bayley and Dennis Preston (eds.), *Second Language Acquisition and Linguistic Variation*, Amsterdam: Benjamins, 75–96.

MAJOR, R. C. & FAUDREE, M. C., 1996. 'Markedness universals and the acquisition of voicing contrasts by Korean speakers of English', *Studies in Second Language Acquisition* 18, 69–90.

MARSLEN-WILSON, WILLIAM, 1993. 'Issues of process and representation in lexical access', in Gerry Altmann and Richard Shillcock (eds.), *Cognitive Models of Speech Processing: The Sperlonga Meeting 11*, Cambridge, MA: MIT Press, 187–210.

MCQUEEN, J. M., CUTLER, A., BRISCOE, E. J. & NORRIS, D. G., 1995. 'Models of continuous speech recognition and the structure of vocabulary', *Language & Cognitive Processes* 10, 309–333.

MEHLER, JACQUES & CHRISTOPHE, ANNE, 1992. 'Speech processing and segmentation in Romance languages', in Y. Tokhura, E. Vatikiotis-Bateson and Y. Sagisaka (eds.), *Speech Perception, Production and Linguistic structure*, Amsterdam: IOS Press, 221–238.

MEHLER, JACQUES, DOMMERGUES, JEAN-YVES, FRAUENFELDER, ULI & SEGUI, JUAN, 1981. 'The syllable's role in speech segmentation', *Journal of Verbal Learning & Verbal Behaviour* 20, 298–305.

MEHLER, JACQUES, DUPOUX, EMANUEL & SEGUI, JUAN, 1990. 'Constraining models of lexical access: the onset of word recognition', in Gerry Altmann (ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, Cambridge, MA: MIT Press, 236–262.

MORGAN, JAMES & DEMUTH, KATHERINE, 1996. 'Signal to syntax: an overview', in James Morgan and Katherine Demuth (eds.), *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*, Mahwah, NJ: Erlbaum, 1–22.

O'GRADY, WILLIAM, 1999. 'Towards a new nativism', *Studies in Second Language Acquisition* 21, 621–633.

O'GRADY, WILLIAM, 2003. 'Language without grammar', Plenary to the *Annual Meeting of the American Association of Applied Linguistics*, Arlington, VA, March 2003.

Otake, T., Hatano, G., Cutler, A. & Mehler, J., 1993. 'Mora or syllable? speech segmentation in Japanese', *Journal of Memory & Language* 32, 358–378.

Rast, Rebekah, 2003. *Le Tout Début de l'acquisition: le traitement initial d'une langue non maternelle par l'apprenant adulte*, Vol. 1, *L'étude*, Doctoral dissertation, Université de Paris VIII.

Rast, Rebekah & Dommergues, Jean-Yves, 2003. 'Towards a characterisation of saliency on first exposure to a second language', *Eurosla Yearbook*, Vol. 3, Amsterdam: Benjamins, 131–156.

Rosa, E. & O'Neill, M., 1998. 'Effects of stress and location on acoustic salience at the initial stages of Spanish L2 input processing', *Spanish Applied Linguistics* 2, 24–52.

Sorace, Antonella, 1993a. 'Incomplete and divergent representations of unaccusativity in non-native grammars', *Second Language Research* 9, 22–47.

Sorace, Antonella, 1993b. 'Unaccusativity and auxiliary choice in non-native grammars of Italian and French: asymmetries and predictable indeterminacy', *Journal of French Language Studies* 3, 71–93.

Strange, W. (ed.), 1995. *Speech Perception and Linguistic Experience*. *Issues in Cross-linguistic Research*, Baltimore, MD: York Press.

Terken, J. & Nooteboom, S. G., 1987. 'Opposite effects of accentuation and deaccentuation on verification of latencies for given and new information', *Language & Cognitive Processes* 2, 145–164.

Towell, Richard & Hawkins, Roger, 1994. *Approaches to Second Language Acquisition*, Clevedon: Multilingual Matters.

VanPatten, Bill, 2003. *Processing Instruction: Theory, Research, and Commentary*, Mahwah, NJ: Erlbaum.

Warren, P., 1985. *The Temporal Organisation and Perception of Speech*, Doctoral dissertation, University of Cambridge.

Wells, W. H. G., 1986. 'An experimental approach to the interpretation of focus in spoken English', in C. Johns-Lewis (ed.), *Intonation in Discourse*, London: Croom Helm, 53–75.

Werker, Janet., 1995. 'Age-related changes in cross-language speech perception: standing at the crossroads', in W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-linguistic Research*, Baltimore, MD: York Press, 155–169.

White, L., 2003. *Second Language Acquisition*, Cambridge: Cambridge University Press.