4 Processing 'words' in early-stage foreign language acquisition: a comparison of first exposure and low proficiency learners

S. E. Carroll

Words are not primitives of language. Words are triplets of representations, consisting of a phonological structure, a morphosyntactic structure, and a meaning. To fully know a word, a learner must represent all three types. However, even this view of words may be a simplification. Psycholinguistic research suggests that these three types of representations are computed within distinct processing modules of the language faculty (Jackendoff 1990, 2002). Consequently, a learner may encode one kind of representation before he or she encodes the other kinds. Developmental studies can shed light on the emergence of each type of representation, and the kind and amount of input each requires.

These representations may also be represented in memory in different ways. Consider sound forms. Although we often talk about the learner learning a single sound form, what is stored in memory may, in fact, be the result of abstraction over multiple sound forms segmented from the speech signal on different occasions. On being exposed to a word for the first time, a learner may segment a single phonetic *token* (or exemplar) of a word from the speech signal (Goldinger 1998). By hypothesis, episodic memory would support recognition of the exemplar on repetition of the same stimulus. However, to reliably recognize the word in other phonetic contexts, the learner might need to be exposed to multiple tokens exhibiting the different acoustic properties found in such contexts, incrementally representing in memory a cluster of distinct exemplars (Ellis 2003: 69). Research on the recognition of phoneme contrasts suggests that adult L2 learners benefit from richness in the signal such as multiple talkers and variable phonetic contexts (Lively, Pisoni & Logan 1992). However, as Cutler and Broersma (2005) emphasize, learning phoneme contrasts results from lexical processing. Infants and young children

The research was made possible through generous funding from the Social Sciences & Humanities Research Council of Canada through both its Canada Research Chairs program (950–202408) and its general grants program (410–2006–0323). I wish to thank my research assistants Lindsay Hracs, Danica MacDonald, and Silke Weber, who carried out essential tasks during various stages of this research. I am especially grateful to the reviewers for extensive commentary and suggestions for improving this chapter. Errors and remaining opaqueness are my own fault.

learn those contrasts because they permit them to understand words in the input before they have a target lexicon. Bilinguals, in contrast, can draw on several sources of information to recognize words, including abstract sound units of the L1 (Cutler 2001).

A wealth of processing studies demonstrates that listeners do compute abstract representations of sounds from speech (Cutler 2008). Such abstract representations are required not just to segment continuous speech into discrete units of sound but also to identify *lexemes*, the inflectional variants of a word recognizably still the ‘same’ word, e.g., *dog* and *dogs* or *eat*, *ate*, and *eaten* (Matthews 1974). A lexeme can be expressed by sound forms that do not mirror its boundaries. For example, many languages exhibit processes of re-syllabication where a consonant ending one word will appear as the initial sound of the following word. Giegerich (1992: 280) provides the English example *These are old eggs* where the [z] of *these* and the [d] of *old* re-syllabify onto *are* and *eggs* respectively. In this case, the word boundaries of the lexemes *egg* and *old* do not align with the edges of prosodic units, as they would when *eggs* and *old* are pronounced in isolation and bounded by pauses. It follows that hearing words in isolation, for example, in lists, will not provide the learner with representations suitable for recognizing words in continuous speech and that simply storing exemplars of concatenated strings of syllables segmented from speech is not going to provide a representation of the lexeme either (contra Ellis 2003). Other abstraction processes must intervene.

Such considerations have implications for debates about the amount of exposure required to learn a word. If words consist of multiple representations, talk of frequency effects on word learning will be unenlightening unless input counts and discussion are relativized to that aspect of knowledge of a word one chooses to focus on. Thus, if what is at stake is the learner’s ability to segment phonetic tokens spoken by a single individual on a single occasion, the answer might well be: “Not much exposure to a word is needed!” Indeed, studies of first exposure or *ab initio* learners have shown that they can rapidly segment tokens of sound forms from continuous speech. This is true whether adults are trained on synthesized strings of sounds from artificial languages (Saffran, Newport & Aslin 1996; cf. Folia et al. 2010 for review), whether they are exposed to natural language stimuli recorded by a single speaker and presented in laboratory conditions (Carroll, Jackson & MacDonald 2009; Gullberg et al. 2010; Kittleson et al. 2010), or whether they are exposed to an L2 in tutored conditions involving a single instructor (Rast 2008; Rast & Dommergues 2003; Shoemaker & Rast 2013). The segmentation of sound tokens is therefore robust and appears from the first stage of L2 development. Moreover, this ability manifests itself not only when the syllable structure of the target words is the unmarked C(onsonant)V(owel)-syllable (Saffran et al. 1996) but also when the target language has more complex syllable structures than the L1 and

Comp. by: KNarayanan Stage: Proof Chapter No.: 4 Title Name: HANandRAST
Date:2/6/14 Time:12:33:50 Page Number: 109

the repertoire of consonants and vowels in the syllable is distinct from that of the L1 (Rast 2008; Rast & Dommergues 2003). This suggests an ability to rapidly generalize beyond L1 or unmarked representations – an important result. Nevertheless, as promising as these studies are, they say nothing about the ability of first exposure learners to recognize those same words when spoken by different speakers, at different speech rates, or in noisy conditions – conditions that are known to challenge the abilities of even advanced "L2ers" (Mayo, Florentine & Buss 1997). Nor do extant studies say anything about the first exposure learner's ability to recognize words within words, e.g., *start* within *started*, *starting*, *start-pistol*, although proficient speech processing involves the activation of many possible words that compete for recognition (McQueen 2005). They say nothing about the learner's ability to recognize a lexeme under conditions of morphophonological variation, e.g., *eat*, *ate*, *eaten*. And the stimuli used in such studies present none of the problems of word identification raised whenever word boundaries fail to align with prosodic categories (under re-syllabication). In short, these studies show that first exposure learners have accomplished the first step: breaking a continuous signal into discrete units. They do not show that this ability generalizes to word recognition involving other abstract representations.

Debates about the abstractness or concreteness of representations spill over into debates about learning mechanisms. Consider, for example, that a large literature dealing with segmentation in the initial state uses a statistical learning paradigm that involves exposing people to artificial languages consisting of continuous strings of synthesized CV syllables, e.g., *tulido bagedi talodu*. Within a target 'word,' the first syllable, e.g., (tu), is a good predictor of the appearance of the following syllable, e.g., (li). The co-occurrence probabilities (called *transitional probabilities*) will be much lower, by hypothesis, between the final syllable of a word and the initial syllable of the next word, here (do) (ba). Both adults and infants can segment word tokens from continuous speech relying only on transitional probabilities (Morgan & Saffran 1995; Saffran et al. 1996). Given these results, one might conclude that statistical learning will account for all aspects of word learning (e.g., Ellis 1998, 2006). This would be a mistake. First, studies on first exposure learners that have examined transitional probabilities (TPs) calculated on controlled natural language input have found limited effects (Gullberg et al. 2010), suggesting important differences in the sound properties of the artificial languages used and natural languages. Second, it is well known that many words tend to co-occur in idioms and set phrases (*hi there*; *no problem*) which are stored in memory too (Jackendoff 2002, 2008). Thus, a low TP alone will not predict the word status of a unit since, e.g., *hi* will be a good predictor of the occurrence of *there* but they are distinct words nonetheless. That child and adult L2 learners may segment strings that involve sequences of words in the target language ('formulae') is well documented (Hakuta 1974; Wong Fillmore 1976; cf. Wray 2002 for discussion). Third, recent work has shown that statistical learning in adults is constrained by prior knowledge of the sound system of the L1 (Finn & Hudson Kam 2008). Learners trained on an artificial language successfully segmented word-length sequences inconsistent with L1 constraints on sequences of consonants and vowels within the syllable only when the 'words' presented were separated by slight pauses. Other researchers have also found that pauses change the nature of speech processing (Endress & Bonatti 2007; Gómez 2002; Mattys & Clark 2002). This suggests that a resolution of the word segmentation problem may be necessary *before* learners can extract co-occurrence dependencies between non-adjacent syllables, a capacity that some have stated lies behind our ability to abstract exponents of inflectional morphology (Perruchet et al. 2004). The ability to encode inflectional morphology is, recall, part of the solution to recognizing lexemes. Fourth, when first exposure learners have been trained on natural language stimuli that sound like L1 words, automatic activation of L1 lexical entries has been observed (Carroll 2012a; Park & Han 2008), showing effects that are well documented in more advanced learners and highly proficient bilinguals (Costa, Caramazza & Sebastián-Gallés 2000; de Groot & Keijzer 2000). Similarity in sound forms has been measured as phonemic similarity, and shown to affect word learning at the initial stage (Rast 2008[4]; Rast & Dommergues 2003). Thus, a small but diverse literature has demonstrated robust effects of L1 knowledge and top-down processing effects on segmentation of words on first exposure to an L2 (Han & Liu 2013). These facts are consistent with studies on native speakers which show that segmentation is affected by lexical and morphosyntactic factors as much as by distributional properties of the input (Mattys & Melhorn 2007; Mattys, White & Melhorn 2005) and suggest a complex interaction of the different kinds of representations of words stored in memory during speech processing.

Turning to considerations of the meanings of words, some first exposure studies show that adults are able to rapidly associate auditory stimuli to referents (Carroll 2012a; Carroll, Jackson & MacDonald 2009; Gullberg et al. 2010). However, reference is a language-to-world relation rather than a relation between linguistic units. The ability to use sentences to comprehend meaning involves being able to compute the specific semantic contribution of each word (more precisely, each morpheme) to the meaning of the whole. Consider that a learner might be shown a picture of a person and hear that the person so depicted is called *Jack*, *Blackbeard*, or *Jack the one-eyed-pirate*. Each sound string corresponds to the person's name. On first exposure to these names, the learner might segment the sound tokens, store them in memory along with a different representation, say, a 3-D representation of the picture, and a concept of the person based on identifiable traits encoded in that 3-D representation (sex, shape, hair-colour, clothing, etc.). This is all that sound-referent association requires. However, the *linguistic* status of these names is quite different. The name *Jack* is monomorphemic and means something like 'the individual x, who bears the name ['ʤæk]' (Katz 1977, 2001). The name *Blackbeard* could have a comparable semantic representation ('the individual x, who bears the name ['blækbiɚd]'), but in the target system *Blackbeard* is a complex word, a compound, that consists of the adjective *black* and the noun *beard*. *Blackbeard* can, unlike *Jack*, function as a description when it is referentially transparent, namely, whenever the bearer of the name sports a black beard. In order to be able to compute the referential transparency of the name from its component parts, i.e., to grasp that it is a description, the learner must be able to analyze the word *Blackbeard* into its grammatical parts, map those grammatical parts onto semantic constituents, namely a property BLACK and an object or thing BEARD, and then compute a predication relation such that the predicate BLACK has the object BEARD within its scope. Example (1a) provides a conceptual representation of the description, and (1b), as a first approximation of a transparent reading of *Blackbeard*, a representation of a proposition equivalent to some individual having a black beard.[1]

(1) a. [THING
BEARD
[PROPERTY
BLACK]]

b. [BE [THING
BEARD
[PROPERTY
BLACK]] AT (x)]

Encoding this kind of meaning is therefore quite a different process from encoding sound-referent associations. Is the ability to encode abstract representations something that distinguishes beginners from first exposure learners? Tracking the emergence of this kind of knowledge is just one of the tasks of first exposure research.

[1] For an account of the notation of conceptual structure semantics, cf. Jackendoff (1983, 1990).

The Calgary first exposure studies

Relativizing the analysis of frequency of inputs

Since 2006, a number of exploratory studies of first exposure learning have been carried out at the University of Calgary. This research was designed to test a novel paradigm for assessing how rapidly learners can come to segment and represent words and map them to a referent provided by a picture. The paradigm permits us to measure learning on the basis of a quantification of exposures to specific linguistic inputs, rather than assessing exposure in terms of minutes or hours (a common but inexact measure of input). The same procedures were used in every study and involve (i) a two-stage training trial that is replicated (up to a maximum of ten training trials) based on the participant's success rate, (ii) a test that confirms that the participants have indeed learned the words, and (iii) a re-test phase two weeks later. This design permits us to measure frequency of exposure in very precise ways and to measure retention over a period where there is no L2 input.[2] By measuring accuracy of word recognition after Training Trial1, we can say if learners recognize target words in two quite distinct phonetic, phonological, and morphosyntactic contexts. A low score would suggest that learners need to extract distinct phonetic tokens on the basis of multiple inputs. A high score would suggest that learners can abstract away from differences in the signal. We can also assess if hearing different words in the same abstract contexts affects performance. This is because each phase of the training trials introduces a name in a recurring prosodic frame, and we can track a participant's accuracy *within* each training trial and determine if accuracy at the end of Training Trial1, when participants have heard the same frame four or five times, is greater than at the beginning when they have heard a given frame once or twice. We can also track increases in accuracy on the lists after four exposures, six exposures, etc. This allows us to measure frequency of exposure with respect to all twenty items. We can ask: How many training trials does it take for a learner to map all twenty names to their pictures? It might turn out to be true (and, indeed, it systematically has been the case) that learners can rapidly segment sound tokens from the signal and map them to referents, even on the

[2] A reviewer has pointed out that the term 'frequency' is used in many ways in psycholinguistic research. Researchers often draw on frequency counts based on occurrence of words in corpora and then show that there are correlations between those counts and behavioural effects, such as priming. An implicit assumption behind the use of such frequency counts is that they reflect the actual experience of language users. In other words, if the word *book* has a high count and *pamphlet* a lower count, a speaker of English will have been exposed to *book* more often than to *pamphlet*. This assumption licenses comparisons to frequency of exposure in first exposure studies.

first few items of Training Trial1, but require repeated exposure to the stimuli to correctly associate a number of names to their referents. The paradigm also allows for a detailed phonetic and phonological analysis of the input. The analyses (see below), coupled with results, are consistent with data from a number of different paradigms and will lead to the formulation of precise hypotheses in future work regarding the use of L1 prosodic knowledge in early L2 word learning.

In each study, learners participated in two experiments (counterbalanced for order). These are listed in (2):

(2) a. Study 1: English first names vs. German cognate first names
b. Study 2: German cognate first names vs. German non-cognate first names
c. Study 3: German first + last name sequences vs. German cognate first names
d. Study 4: German phrasal names vs. German cognate first names

German was chosen as the target language for two principal reasons: its phonetics and phonology are well documented; comparison with English allows for an investigation of cross-linguistic influence. German and English differ at the phonetic level in a variety of ways and yet share many abstract phonological properties. By testing anglophones on German stimuli, one can investigate the extent to which learning is affected by these more abstract similarities. Given that others have deliberately chosen language pairs from distinct families with quite different typological properties (Dutch L1/Chinese L2 in the case of Gullberg et al. 2010, and French L1/Polish L2 in the case of Rast 2008, 2010 and Shoemaker & Rast 2013), this choice permits interesting cross-study comparisons. For example, learners in the first study were almost as accurate on German cognate first names, e.g., *Anita* [aˈnita], as they were on English names which were, by hypothesis, all familiar to the participants (Carroll et al. 2009). This shows that cognate words are readily segmented from continuous speech (see also Rast 2008; Rast & Dommergues 2003). The second study showed that first exposure learners were more accurate on the cognates than on the non-cognate names, although these too were acquired. Importantly, learners responded faster to the cognate names than to the non-cognates (Carroll 2012a). The cognate effect is, moreover, robust. This has methodological consequences; it shows that choosing languages from different families is no guarantee that learners are not accessing L1 lexical representations. Cognates and borrowings that arise from language contact may need to be studied separately from other words.[3]

[3] It is worth noting that in the studies from this paradigm, German *cognate* words were actually words historically related to English words. In most psycholinguistic research, the term is used

Because cognates were used repeatedly across experiments, an L2 baseline has been established for comparing learner performance on L2 names with other properties. As noted, Study 2 showed that learners had no difficulty learning non-cognate names, but they required more training trials to do so (Carroll 2012a). Study 2 also showed that once each word type was segmented and mapped to a referent, learners were able to readily recall it. Thus, differences based on the properties of the words arise at the point of encoding, not at the moment of recall. This chapter reports on the results of Study 3, henceforth *the study*, which compared the learning of cognate names with complex first and last name sequences including semantically transparent names of the *Blackbeard* type.

Questions of 'naturalness' and ecological validity

Many studies that expose learners to a language for the first time involve artificial languages and sometimes artificial tasks. An obvious question that arises is whether we should generalize from such studies to normal cases of language acquisition. Some researchers have gone to great pains to develop paradigms that intuitively reflect 'natural' uses of language (Gullberg et al. 2010). At the same time, investigative rigor demands that we control our stimuli. The present study and other studies based on this paradigm did this, as did Gullberg et al. (2010) and Shoemaker and Rast (2013). Of course, claims to 'naturalness' are only as good as the linguistic analyses of stimuli that support them. Gullberg and her colleagues created a video that presented a pseudo-weather report in Mandarin Chinese. This paradigm presents participants with extended text rather than isolated sentences (as in the present paradigm; see below). Moreover, the extended text presents a constant field (meteorology), mode (written language presented as spoken language), and functional tenor (weather reports provide descriptions or make predictions). See Gregory and Carroll (1978) on these constructs. Weather reports are highly stereotyped text types that consist of simple sentences (little or no recursion), simple tense-marking, dense use of spatial expressions (e.g., prepositional phrases or adverb phrases), place names, and field-specific meteorological

for any L2 word that activates an L1 word based on similarity of sounds or spellings and where there is shared semantic content (see Friel & Kennison 2001). Measurements of sound or graphemic 'similarity' is motivated by the assumption that lexical activation is a bottom-up data-driven process, which, we have seen, is too simple a view. In this and related studies, both *Anita* and *Georg* [geɔ:k] (compare [dʒɔrdʒ]) were classified as cognate. The production data showed that *Anita* could trigger completely English pronunciations (showing activation of the English word) but *Georg* never did. This fact does not prove that English *George* was not activated but it clearly shows that learners segmented and retained phonetic and phonological properties of the input and used these (not an English representation) for production.

expressions. See Leplus, Langlais, and Lapalme (2004) for discussion. All of these properties will facilitate interpretation for anyone familiar with the genre (adults, say, rather than children) and makes the paradigm especially suitable for examining first exposure learning of a language where the assumption is that there are few or no commonalities to facilitate cross-linguistic influence. However, the very same properties make generalizing from this text type to other text types, e.g., conversation, difficult.

As for the current paradigm, each sentence was a natural way to introduce a referent when looking at a picture. However, no attempt was made to create a paragraph text type, although the paradigm in no way precludes this. Rather, the sentences were presented as a list. See Appendix 1. What we lose in naturalness, as measured in terms of vernacular conversation, we gain in being able to carry out detailed linguistic analyses of the stimuli, something which has been done with the other paradigms in only a limited way (looking at, e.g., effects of word repetition, word length, phonemic similarity, and sentence position).

What analyses of the stimuli tell us

Examples of the presentational contexts used in this and all studies based on this paradigm are given in (3). Henceforth, this context will be referred to as a syntactic *frame*.

(3) a. *Hier ist* Name
Here is . . .
b. *Das ist* Name
That is . . .
c. *Da sehen Sie* Name
There see you . . .
'There you see . . .'
d. *Da steht* Name
There stands . . .

The word occurring in the 'Name' position is new information, as is normal for this kind of declarative sentence. In Study 1, the study that compared performance on English stimuli to performance on German stimuli, participants heard sentences such as *Here we have Anne*, *This is Brian*, *Now we see Dave*, . . . One might assume that the target name was always focally accented and the locus of the highest pitch in the utterance. Phonetic analysis of the English-language stimuli indeed showed that the English target name was invariably the focally accented word in the utterance. However, only in some cases was the German Name the locus of pitch movement and maximal amplitude in the German utterances. In the contexts provided by (3c) and (3d), for example, the deictic *da* was the locus of focal accent. This meant that in the German stimuli, learners could not rely on focal accent as the cue to where the target name was.

Because the name occurs at the end of each declarative sentence during the first phase of a training trial, it appears before a pause. The literature on segmentation of artificial languages has revealed that even short pauses cue word boundaries (Endress & Bonatti 2007; Gómez 2002). End-of-sentence pause (that is, pause under final falling or rising pitch contour) has been taken to be a strong cue to segmentation (Monaghan & Christiansen 2010). Implicit in such a claim is the assumption that the right edge of a sentence is also the right edge of a word, an assumption explicitly built into theories of metrical phonology and prosody–syntax correspondences (e.g., Nespor & Vogel 1983, 1986; Selkirk 1978). Assuming that the constraints proposed in such theories are relevant to speech processing, the presentational sentences provided participants with clear cues to the right boundary of the name as a prosodic unit. However, names in the questions to which the participants responded (see below) appeared either in sentence-medial or in sentence-final position. Phonetic analysis of the questions confirmed that the sentence-medial names were not bounded by pauses. Thus, in the questions, which directly preceded the responses and which are critical stimuli for responses, pause was a clear cue only for some of the target words. In fact, in the various studies, position of the target name in the question made no difference to accuracy of response, suggesting that cues other than pause were readily available to facilitate segmentation.

If we assume that learners are influenced by their knowledge of the L1 in segmenting words from such stimuli, as Finn and Hudson Kam (2008) have suggested, then we may ask if English speakers use their knowledge of English prosodic structure to parse the input into discrete units of sound. A prosodic word is a prosodic unit that can appear in isolation, for example, in a list. Foot structure is hypothesized to be a level of prosodic structure above the syllable but below the prosodic word (PW). The foot is the domain over which rhythm in English is organized. It consists minimally of one syllable and maximally of three syllables, e.g. $(\sigma_{\text{Strong}})_{\text{Foot}}$, $(\sigma_{\text{Strong}}\ \sigma_{\text{Weak}})_{\text{Foot}}$ or $(\sigma_{\text{Strong}}\ \sigma_{\text{Weak}}\sigma_{\text{Weak}})_{\text{Foot}}$, where the stressed syllable is perceptually salient, as shown by the diacritic label *strong*. Anglophones are especially sensitive to strong syllables when segmenting words from speech (Cutler & Norris 1988). If English speakers bring to the task of analyzing German input their knowledge of these abstract structures, then the strings in (3) could be parsed as sequences of two and three prosodic words as shown in (4):

(4) a. $[(\textit{das})(\textit{ist})]_{\text{PW1}}$ $[\text{Name}]_{\text{PW2}}$
b. $[(\textit{hier})(\textit{ist})]_{\text{PW1}}$ $[\text{Name}]_{\text{PW2}}$
c. $[\textit{da}]_{\text{PW1}}$ $[(\textit{se})(\textit{hen})(\textit{sie})]_{\text{PW2}}$ $[\text{Name}]_{\text{PW3}}$
d. $[\textit{da}]_{\text{PW1}}$ $[\textit{steht}]_{\text{PW2}}$ $[\text{Name}]_{\text{PW3}}$

If first exposure learners are able to parse input into minimal prosodic words, perhaps computing from the right edge of the utterance (from the pause), this would explain how they locate the *left* edge of the name: the target name is the (English-based) minimal prosodic word immediately preceding a pause.

As (4) shows, the prosodic parsing for (4a,b) is the same and differs from the prosodic parsing of (4c) and (4d). Examples (4c–d) also reveal how the prosodic word can differ in syllable length. Word length has been examined in first exposure studies. In the first two studies using the present paradigm, length of the name had no effect on the learners' ability to learn the words, something that Rast (2008) also found. In the Chinese weather report studies, in contrast, Gullberg et al. (2010) found that Dutch learners were biased to reject one-syllable target words as words heard in the video and performed significantly better on two-syllable words, although the one-syllable words appeared to be more frequent overall in the input. Still, in these studies, word forms tended to be relatively short (one to four syllables in length), raising the question of whether first exposure learners can segment much longer sequences.

In the current studies, stimuli were presented in two phases: declarative sentences and questions. The question types to which participants responded are shown in (5):

(5) a. *Ist hier* Name1 *oder* Name2?
Is here . . . or . . .
b. *Ist das* Name1 *oder* Name2?
Is that . . . or . . .
c. *Sehen Sie da* Name1 *oder* Name2?
See you there . . . *or* . . .
d. *Steht da* Name1 *oder* Name2?
Stands there . . . *or* . . .

Thus, each training trial consisted of twenty declarative sentences followed by twenty questions, so that each sentence-question pair over a training cycle presented a given target name twice. Repetitions of training trials increased exposure by two increments. The alternative choice (the *foil*) was always a novel name so learners were hearing it for the first time in the second half of Training Trial1.[4]

Each of the syntactic frames occurred five times during a given phase of the training trials. On the assumption that learners are computing statistical

[4] A reviewer notes that responses could be predicted given the fact that the only novel word in the questions was the foil. The logic of this claim needs to be unpacked. If the logic is that learners' attention was drawn to the novel sound form, then they ought to have selected the foil more often than the target noun, which they did not. If the logic is that it is the contrast between what is in episodic memory (all of the words of the questions except the foil) and what is not in episodic memory (the foil) that permits an accurate response, then we predict that learner performance should get worse as the training trials proceeded. This is because after Training Trial1 *all* of the words in the questions were presumably in episodic memory. This was also not the case.

properties of the input in real time, participants may have computed the transitional probabilities of the different frames in the declarative sentences even before they were exposed to the questions, that is to say, after having heard each target name once. This is because repetition of the frames ought to lead the learners to treat the recurrent frames as distinct 'words' from the target names. As they hear the presentational sentences, learners will compute that each of the sequences *dasist*, *hierist dasehenSie*, and *dasteht* can be followed by a distinct set of syllables. The TP between *das* and *ist* was 1.0 while the TP between *ist* and the first syllable of the following name was 0.1. In the declarative sentences, the TP between the frame and the following name was never higher than 0.2, meaning that there was a clear contrast between the frames and the names. However, the computation of these statistics should require exposure to several sentences. Since the questions involve different linear orders, participants would have to re-compute the transitional probabilities as they listen to the questions. Accordingly, one would expect performance to be better at the end of the question phase of Training Trial 1 than at the beginning, something that was not found in any of the studies.

The rhythm of the questions was slightly different from the rhythm of the declarative sentences. The one dramatic difference between the declarative sentences and the questions was, of course, the intonation. In each question, Name1 was pronounced under a rising pitch with a falling pitch occurring over *oder* Name2. Whether this intonation might have facilitated segmentation is difficult to say without a number of explicit assumptions about how a continuous variable like pitch might be aligned to discrete prosodic units. This issue was set aside.

Studies of the learning of artificial grammars have shown that adults are sensitive to the position of 'markers,' recurrent words that behave like grammatical words (Braine 1986; Valian & Coulson 1988). It can be argued that the recurrent appearance of the word *oder* with different phonetic and phonological content on its left and right may serve the same function. In other words, the conjunct *oder* is a bit of stability in a sequence of changing sound forms. Regardless of whether learners are rapidly computing statistical properties of the input or analyzing the input prosodically (or both), they should rapidly segment *oder* as a word, which can then mark the boundaries of what appears to its left and right. Thus, the recurrent contexts (*istdas*) . . . *oder* . . . or (*isthier*) . . . *oder* . . . may also provide learners with clear cues that what occurs in the position . . . is a prosodic word. The stimuli may therefore clearly signal not only the left and right edges of Name2 but also the left and right edges of Name1. If so, the prediction is once again a rapid ability to segment names despite phonetic variability in the input contexts.

Finally, Monaghan and Christiansen (2010) have argued that proper names are themselves highly salient, readily segmented and can serve as the basis for

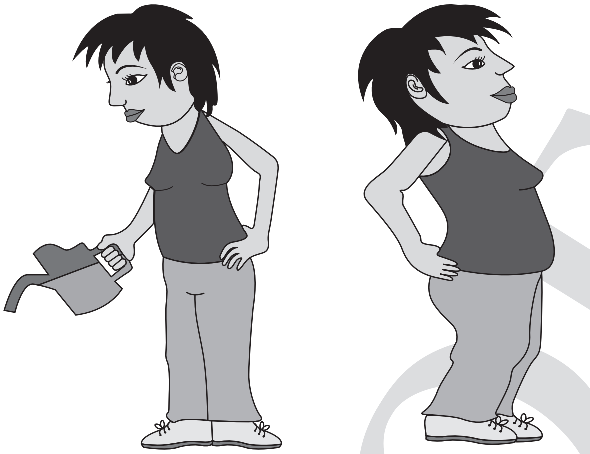

Figure 4.1 Examples of pictures used

carving up the signal between their own location and the right edge of an utterance. If this is true, sentence-medial proper names might be easily segmented regardless of their prosodic properties.

To sum up, the stimuli discussed here present a number of different cues that learners might rely on to segment the sound forms: transitional probabilities, rhythm, and an overt marker. This is also true of natural speech. What makes these stimuli less like natural speech is the frequency of the recurrent patterns.

The study

Visual and auditory stimuli

An artist was instructed to create two digitalized pictures each of a variety of male and female, adult and child persons from which twenty were selected for the study. Pictures of the same individual differed in profile, pose (seated vs. standing), and the objects they might hold in their hands. See Figure 4.1. One picture was used in training and the re-tests; the other was used during the test phase. The point of using two pictures was to encourage the learners to create an abstract conceptual representation of a PERSON, rather than simply mapping a sound token to a visual token.

Word lists were created consisting of cognate and non-cognate first names and last names. Names in this study ranged in length from four to seven syllables (mean=4.8).[5] These names were inserted in the same syntactic frames (declarative sentences and questions) used in all of the other studies, preserving the transitional probabilities previously described. Examples are shown in (6).

(6) a. *Hier ist Dagmar Baumgartner*
Here is Dagmar Baumgartner
b. *Das ist Heribert Hübschdorn*
That is Heribert Hübschdorn
c. *Da steht Axel Braunbart*
There stands Axel Braunbart
d. *Hier sehen Sie Gerda Rotkopf*
Here see you Gerda Rotkopf

Foils were created that had the same number of syllables and rhythmic structure as the target words. However, since at this point in the explorations there was no interest in assessing the extent to which first exposure learners can carry out fine-grained phonetic analysis of the input, no attempt was made to create minimal pairs.[6] Examples of the foils are shown in (7).

(7) a. *Ist hier Dagmar Baumgartner oder Trüdel Dieterich?*
Is here Dagmar Baumgartner or Trüdel Dieterich?
b. *Ist das Ulrich Isenmann oder Heribert Hübschdorn?*
Is that Ulrich Isenmann or Heribert Hübschdorn?
c. *Steht da Axel Braunbart oder Achim Meighörner?*
Stands there Axel Braunbart or Achim Meighörner?
d. *Sehen Sie hier Gerda Rotkopf oder Irmtraud Joppke*?
Here see you Gerda Rotkopf or Irmtraud Joppke?

Foils were repeated across training trials, tests, and re-tests. Thus, after Training Trial1, participants were repeatedly exposed to both foils and targets.

All stimuli were recorded by a native speaker of Standard German in a recording booth in the psycholinguistics laboratory. All stimuli were then assessed for naturalness and re-recorded where necessary. Subsequent phonetic analysis of the questions showed no pauses between the PWs preceding Name1 or Name2. A consistent rising intonation occurred over Name1 with a fall occurring over *oder* and Name2. A prosodic analysis of the input revealed that fifteen out of the twenty names could be prosodically parsed as two prosodic words, as in (8a), four names as three prosodic words (8b), and one name as four prosodic words (8c).

5 These words were twice as long as names used in previous studies (means: English study=1.65; German cognate first names=2.35; German non-cognate first names=2.1).

6 A study currently underway which presents first exposure learners with French stimuli and does offer minimal pairs suggests that identifying the target word under this condition is very hard.

(8) a. ([ˈak][səl])$_{PW1}$ ([ˈbʁaʊn][baːt])$_{PW2}$ *Axel Braunbart*
b. ([ˈga][bʁi])$_{PW1}$ ([ˈe][lə])$_{PW2}$ ([ˈblaʊ][hɛmt])$_{PW3}$ *Gabrielle Blauhemd*
c. ([maː])$_{PW1}$ ([ˈti])([na])$_{PW2}$ ([ˈpfa][nən])$_{PW3}$ ([ʃmɪt])$_{PW4}$ *Martina Pfannenschmidt*

If learners actually computed these analyses, they would be able to directly map a prosodic word to either a first name or to a last name in three-quarters of the cases.

Pictures and sound files were entered into a computer program using E-Prime. The computer automatically recorded responses as correct or incorrect as well as response latencies.

Readers familiar with German will have realized that some of the last names are unusual. This is because half of the last names were created to pick out a referential detail from the picture. Thus, the woman shown in Figure 4.1 (holding a watering can) was called *Dagmar Baumgartner*, literally ‘Dagmar [[tree] [gardener]],’ *Axel Braunbart*, literally ‘Axel [[brown] [beard]] had a brown beard,’ *Gabrielle Blauhut*, literally ‘Gabrielle [[blue] [hat]] was wearing a blue hat,’ and so on. Native speakers, familiar with proper names typical to their region and culture, tend not to notice that they sometimes have a literal interpretation. For second language learners, however, these interpretations are often highly salient, when they can parse the names into their component words. We ask here if and when learners will demonstrate sensitivity to a semantically transparent interpretation. By hypothesis, first exposure learners have no L2 lexical entries and hence no linguistic basis yet for computing the internal structure of a compound word. While they may segment phonetic tokens and map these to referents, they are not computing a morphosyntactic analysis of the compound names. Accordingly, they are not in a position to compute a transparent semantic representation that would draw their attention to specific details in the pictures, facilitating the sound-referent mapping. The questions of interest were: Would *beginner* learners be sensitive to the meaning of these words? Or would they, like the first exposure learners, simply segment the sound forms and map them to referents, treating them as monomorphemes? If learners are processing the semantically transparent compounds structurally and mapping the content to the pictures, they should perform differently on the transparent compounds than on names like *Hannelore Fleischhauer* ‘Hannelore meat + carver’ (associated to a picture of a woman holding or playing a guitar), *Antje Grünwiese* ‘Antje green + meadow’ (associated to a picture of female nurse), or *Ludo Hartstein* ‘Ludo hard + stone’ (associated to a picture of a man wearing a suit and holding a briefcase with or without a cellular telephone). Consequently, it is the *difference* in responses to the semantically transparent cases and the non-transparent cases that will show if beginners can access the semantic content of the compound’s constituents.

Measures

Accuracy and response latencies were measured on the first training trial, Test, Re-test1, and Re-test2. The number of training trials needed to learn all twenty names were also measured.

Participants

Participants were fifty University of Calgary students, recruited through posters seeking individuals with one to two semesters of German. Individuals were paid for their participation. Initially, twenty-six individuals with no prior knowledge of German (FExp group) and twenty-four individuals who knew some German were tested. Data from three participants were removed from the data set of the beginners (Beg) when analysis of the questionnaire results, their self-assessment, and scores on the Goethe Institute grammar and listening tests revealed that they were far more proficient than the other participants.[7] Among the remaining participants in the Beg group, ten ranked themselves as 3 on the scale (intermediate between 'very good' and 'very poor'), nine ranked themselves as 4 or 'poor,' and two participants ranked themselves as 'very poor.'[8] Mean scores for these participants on the Goethe Institute grammar test were 44.2 percent (s.e.=0.018), and mean scores on the listening test were 58.5 percent (s.e.=0.0306). In all, analysis was conducted on complete data sets from twenty-three beginners and twenty-five first exposure learners.[9]

Hypotheses

It was predicted that learners from both groups would be able to segment phonetic tokens of both first and last names from the stimuli regardless of the length of the name and map these names accurately to the pictures. It was also predicted that the Beg group would be advantaged over the FExp group; that is, they would make fewer errors in Trial 1 of the training trials, would take fewer trials to learn to criterion, would retain knowledge of the name-picture mappings better than the FExp group over two weeks, and would respond

[7] Two of these participants ranked themselves as 'good' and one as 'very good.' All three scored in the common European Union Framework level as 'C1' with mean grammar scores of 77.66 percent and mean scores on the listening test of 87.3 percent.

[8] Despite the differences in their self-assessments, performance on the language tests and the main tasks showed no significant differences.

[9] Some additional data cleaning was necessary when it was discovered that one first exposure participant failed to press any key during the test of the compound names. As a result, this participant also did not return for the re-tests. This participant's data are included in the analysis of the cognates and in the training trials of the compound names.

faster than the FExp group in keying in their responses. It was predicted that only the Beg group would be sensitive to the internal structure of the compound last names and be able to map meanings of words to visual details of the pictures. Finally, if the Beg group were analyzing the internal structure of the semantically transparent compound names and not just activating a sound token, they might respond more slowly to these words.

Procedures

On entering the laboratory, participants read a project description and signed a consent form to satisfy the ethical review process of the university. They filled out a questionnaire providing basic information on their prior exposure to German and assessed their own knowledge on a five-point scale (1=near native-like, 5=poor). Participants who claimed to know some German then did both parts of the online Goethe Institute's German-language proficiency test: a grammar test and a listening test. All participants then carried out the tasks of the main experiments, first with one set of stimuli (either the cognate names or the first + last names) and then with the other set, with order of the experiments counterbalanced across participants. Two weeks after doing the first session of an experiment, they returned and carried out the re-tests for that experiment.

In the training portion of the paradigm, participants sat at a computer console and heard over studio-quality headphones instructions in English about the main experiment. Instructions were also presented in writing. Participants were informed that they would see twenty line drawings of persons and simultaneously hear twenty sentences in German. They were asked to learn the names of the people depicted in the pictures. The instructions included three practice items. Participants proceeded through the instructions at their own pace. Timing of the rest of the experiment was experimenter-controlled. Stimuli were fully randomized on all training trials and on all tests and re-tests. Listening to the twenty declarative sentences, while viewing the twenty pictures, constitutes the first half of the initial training trial. The sentences were then followed by a sequence of questions (fully randomized) that correspond morphosyntactically and semantically to the sentences. These questions offered the participants a forced choice between two names (target and foil), one of which corresponded to the picture. Participants pressed a clearly marked key on the left-hand side of the computer keyboard if they thought the name of the depicted person was the first name offered in the question (Name1), or they pressed a clearly marked key on the right-hand side of the keyboard if they thought the person's name was the second in the question (Name2). Participants had 2500 ms to respond after which time an incorrect response was automatically recorded. At the end of the first training trial, the participant's accuracy score was automatically indicated on the computer screen. This indicated whether to proceed to the test or do another training cycle. Training trials could be repeated up to nine times. When a participant was not able to correctly match all twenty names to the twenty pictures within ten training trials, his participation in the experiment ended. Participants who correctly mapped all names to pictures submitted to a test. The test consisted of the same questions used in the second half of each training trial (maintaining the same order of target name and foil) but a different picture of the person named was used (see Figure 4.1). The test provided confirmation that correct performance on the final training trial was not an accident and that learners could respond correctly to a novel visual stimulus.

Approximately two weeks later, remaining participants returned to the laboratory and responded to the questions yet again. At re-test, the position of the target name and the foil were exchanged. If the target word was Name1 in the training trials and test, then it was Name2 in the re-test questions. If the target word was Name2 in the training trials and test, then it was Name1 in the re-test questions. The purpose of this was to see if the learners could segment names from a novel auditory stimulus. Thus, the stimuli at both test and re-test were distinct from those used during the training trials. The re-test checked to see if participants retained what they had encoded. Retention and forgetting have not been studied in other first exposure paradigms. Finally, there were two re-tests, Re-test1 and Re-test2. Participants submitted to Re-test2 immediately after completing Re-test1. The difference between them was that as participants proceeded through Re-test1 they were given (for the first time) feedback as to the correctness of their responses.

Coding and analysis

Scores on the online Goethe Institute grammar and listening tests were automatically generated and recorded along with the questionnaire data. Questionnaire data were entered by hand into Excel files. Responses on the main task were automatically coded by the computer as "correct" or "incorrect." These data were transferred by hand to Excel files and subsequently loaded to Stata 9 for statistical analysis (*t*-tests and analysis of variance [ANOVA]). Response latencies were log-transformed for analysis.

Results

Performance on the cognate items was very high at 90.3% (SD=0.1087, Min=0.25, Max=1.0). This result was significantly above chance ($t(173)$= 48.83, p=0.0000, SE=0082, 95%, conf. interval: 0.886–0.918). These results

Table 4.1 *Performance on the training trials, number of trials to criterion, test, re-test1, re-test2 by first exposure and beginner L2 learners*

Cognates						First + Compound names				
	Trial1	Test	Re-test1	Re-test2	M	Trial1	Test	Re-test1	Re-test2	M
FExp	78	95.6	87.7	96.4	89.3	77.8	87.7	82.7	86.8	83.6
Beg	88.8	95	87.9	93.9	91.4	90.6	93.9	88.9	91.9	91.3
Mean	83.4	95.3	87.8	95.2	90.4	83.9	90.7	85.9	89.3	87.4
Mean # Training Trials										
FExp		3.9					3.7			
Beg		2.8					2.6			

replicate the findings of earlier work (Study 2) where participants scored 89.1% on the cognate items versus 83.6% on non-cognate items (Carroll 2012a).

Table 4.1 presents a summary of the means for both experiments by group and by measure.

Results show that beginners (M=91.4, SD=8.31) outperformed first exposure learners (M=89.3, SD=12.6) only slightly on accuracy scores. A two-sample t-test with unequal variances (using Satterthwaite's degrees of freedom) showed that this difference was not significant, $t(162.36)=-1.36$, 0.17, p=0.17, n.s. In addition, the range of scores was larger for the FExp group (50 to 100%) than for the Beg group (71.4 to 100%). The mean number of trials to learn to criterion was 3.4 for all subjects (SD=1.75). The FExp group took almost four trials to correctly identify all twenty cognate words (M=3.9, SD=1.81, Min=2, Max=8) while the Beg group took less than three trials (M=2.8, SD=1.51, Min=1, Max=7). This difference was significant ($t(44)$= 2.246, p=0.008, two-tailed, d=0.63). Examination of individual results showed that four beginners scored 100% on Trial1 after only two exposures to each name, while no FExp learner did. This provides clear evidence that beginners did not require repeated exposure to the stimuli to segment them and map them to referents. Repeated exposure did help them in mapping lists of words to their referents. On training Trial2 (after four exposures to each name) a further five beginners scored 100%, along with seven FExp learners. The slowest beginner took seven trials (fourteen exposures) to learn all twenty names, the slowest FExp learner took eight trials (sixteen exposures).[10]

[10] As is typical with this paradigm, participants who took many trials to match all names correctly to pictures were making an error on a single item from Trial2 onwards. So the number of trials

Correlational analyses between the number of trials and the self-assessment, the number of trials and scores on the Goethe Institute grammar test, and the number of trials and scores on the listening test showed no significant correlations (−0.15, p=0.457, n.s.; −0.125, p=0.558, n.s.; 0.16, p=0.452, n.s., respectively).[11]

As Table 4.1 shows, the absence of a significant difference in the means of the groups is due to the fact that although scores differed between the groups in Trial1, at Test, Re-test1, and Re-test2, there was little difference in the scores. A one-way repeated-measures ANOVA was conducted to examine differences in the measures (Training Trial1, Test, Re-test), and groups (FExp, Beg), with the participants as between-subjects error term. It showed, not surprisingly, that accuracy did indeed depend on the measure: $F(3, 122)=18.39$, $p=0.0000$, $MS_{Phase}=0.1508$, $\eta^{2=0}0.073$, a medium effect size (Cohen 1988). There was also an interaction between the measures and the groups: $F(3, 122)=4.26$, $p=0.007$, MS=0.0349, $\eta^{2=}0.017$, a small effect size.[12] Comparison of mean scores by measure (with Bonferroni correction) revealed significant differences between Trial1 and Test ($t=-0.124$, $p=0.0000$), between Trial1 and Re-test2 (t=-0.123, $p=0.000$), between Test and Re-test1 ($t=0.075$, $p=0.002$), and between Re-test1 and Re-test2 ($t=0.074$, $p=0.003$). It is typical of this paradigm that first exposure learners score much lower on Trial1 than on Test, forget somewhat over the two-week interval between testing sessions, and then score very high on Re-test2 after yet more exposure and feedback (Carroll 2012a). This study shows that more advanced learners exhibit the same pattern.

The input was analyzed in terms of the number of syllables in a target (a measure of word length), the number of feet in a target (a measure of prosodic structure), and the position of the word in the question (medial or final), as well as by syntactic frame and by the semantic transparency of the items. Means are reported in Table 4.2 by group.

As can be seen from Table 4.2, means differed very little based on the number of syllables in the words, although it should be kept in mind that the number of words of each type in the input varied (one-syllable words, N=4,

to criterion is clearly telling us about the mappings between segmented word and pictures, not about segmentation per se.

11 There was a strong negative correlation between the self-assessment and the grammar test (−0.6687, p=0.0004). The direction of the correlation is due to the use of the standard German scoring pattern where the symbol "1"="Very Good" and the symbol "5"="Very Poor." Consequently, this result is as expected. There was an even stronger negative correlation between the self-assessment and the listening test (−0.711, p=0.0001). Finally, there was a strong positive correlation between the scores on the grammar and the listening tests (0.6816, p=0.0002).

12 Huynh-Feldt epsilon calculated on measures gives p=.019, Greenhouse-Geisser epsilon gives p=0.02.

Table 4.2 *Performance on all target items categorized by input properties of questions*

	FExp	Beg	Mean of groups
Cognates			
Syll 1	92.6	88.4	90.5
Syll 2	88.3	91.8	90.0
Syll 3	92.2	91.6	91.9
Syll 4	90.6	89.7	90.1
Foot			
Ft 1	89.4	91.5	90.4
Ft 2	90.0	91.1	90.5
QPos			
Medial	90.7	93.2	91.9
Final	88.3	89.6	88.9
Frame			
istdas	86.6	93.5	89.7
isthier	87.4	92.3	89.6
sehensiehier	91.4	87.9	89.8
stehtda	92.7	91.9	92.3
Compounds			
Syll 4	83.6	90.3	86.9
Syll 5	81.5	89.9	85.7
Syll 6	87.9	93.7	90.8
Foot			
Ft 2	88.5	93.0	90.7
Ft 3	80.5	89.3	84.9
Ft 4	93.0	94.4	93.7
QPos			
Medial	85.0	91.6	88.3
Final	82.5	89.8	86.1
Frame			
istdas	86.3	92.1	88.9
isthier	88.4	90.9	89.6
sehensiehier	79.0	90.6	84.4
stehtda	82.7	89.7	85.9
SemTrans	85.8	92.5	89.1
Sem Opaq	81.5	90.1	85.8

M=90.6, *SD*=6.83; two-syllable words, N=52, *M*=89.9, *SD*=9.11; three-syllable words, N=16, *M*=91.9, *SD*=6.80; four-syllable words, N=8, *M*=90.1, *SD*=7.74).[13] Scores were virtually identical by foot length (1 foot N=65, *M*=90.3, *SD*=8.76; 2 feet N=15, *M*=90.5, *SD*=6.72). Means were

[13] Thus, there was only 1 one-syllable name, 2 four-syllable names, 4 three-syllable names, and 13 two-syllable names.

slightly lower in sentence-final position (N=40, M=88.9, SD=8.21) than in sentence-medial position (N=40, M=91.8, SD=8.38). This difference was not significant ($t(78)=-1.59$), p=0.11). Means were also virtually identical in three of the frames (*ist das* N=20, M=89.7, SD=9.68; *ist hier* N=20, M=89.6, SD=8.15; *sehen Sie hier* N=20, M=89.8, SD=9.75), with the fourth frame (*steht da*) being slightly higher (N=20, M=92.3, SD=5.56). This difference was not significant ($F(3, 76)=0.49$, $p=0.68$).

Turning to the analysis of the compounds, learners from the Beg group (M=91.3, SD=8.2) outperformed the FExp group (M=83.6, SD=10.8) on the first + last names. A one-way ANOVA revealed this difference to be significant: $F(1, 180)=28.93$, $MS_{BetGroups}=0.269$, $\eta^{2=}0.139$, $p=0.011$. Performance on all of the measures was high for the Beg and only slightly less high for the FExp group. One-sample t-tests revealed that these scores were all significantly above chance.[14] Participants took on average 3.2 trials to learn all twenty items (SD=1.43, Min=1, Max=6). The Beg group learned in approximately one trial fewer (M=2.6, SD=1.24, Min=1, Max=5) than the FExp group (M=3.7, SD=1.4, Min=1, Max=6). This difference was significant, $t(44)=2.80$, $p<0.008$, $d=0.78$. Direct examination of individual scores showed that five Beg participants scored 100% on Trial1 (after only two exposures to the twenty items), while only one FExp participant scored 100% on Trial1. Eight Beg participants scored 100% on Training Trial2 versus four FExp participants. Another seven Beg group participants scored 100 percent by training Trial3. By training Trial3, another seven FExp participants scored 100%, meaning that fifteen had acquired all twenty items after the same amount of exposure. All the Beg group learned all twenty items within the allotted training period while one FExp participant could not.

Learners from the Beg group (M=90.4, SD=7.3) outperformed the FExp group (77.8, SD=12.1) on the first training trial as well. The range of scores for the FExp group (45–100%) was much larger than it was for the Beg group (75–100%). Table 4.1 also shows that clear differences between the two groups continued into Test: FExp group (M=87.7, SD=10.3) versus Beg group (M=93.9, SD=6.02). A one-way repeated-measures ANOVA, with subjects as between-group variable revealed that the different measures (Trial1, Test, and the Re-tests) account for a significant amount of variance, $F(3, 133)=6.75$, $p=0.0003$, $MS_{Phase}=0.0484$, $\eta^{2=}0.025$. Separate repeated-measures ANOVAs computed on the means of each group revealed that the effect is due solely to the first exposure group: FExp: $F(3, 65)=5.73$,

[14] For the FExp group: Trial1 $t(24)=11.5$, $p=0.0000$; test $t(23)=17.9$, $p=0.0000$; Re-test1 $t(21)=16.2$, $p=0.0000$; Re-test2 $t(21)=20.9$, $p=0.0000$, all two-tailed. For the Beg group: Trial1 $t(22)=33.92$, $p=0.0000$; Re-test1 $t(21)=17.82$, $p=0.0000$; Re-test2 $t(20)=22.37$, $p=0.0000$. The hypothesized mean for the binary choice=0.50 in all cases.

p=0.0015, MS_{Phase} =0.0519, η^2 =0.048; Beg: $F(3, 65)$=2.20, p=0.097. With these stimuli, the magnitude of training effects is much greater for the FExp group than for the Beg. This is because beginners were already close to ceiling on Trial1 and had almost the same score on Re-test1. A different pattern can also be seen for the Beg group on the compound names (compared to the single-word cognates) for the same reason.

The input was again analyzed in terms of the number of syllables in a target, the number of feet in a target, and the position of the word in the question (medial or final), as well as by frame and by the semantic transparency of the items. Means are reported in Table 4.2.

An ANOVA comparing number of feet, number of syllables, frame, and position in the question revealed only the foot structure to be significant: F(3,146)=4.79, p=0.0033, MS_{Foot}=0.0338, η^2 =0.027. Separate one-way ANOVAs on foot structure computed for each group showed that this finding was due to the first exposure group: FExp: $F(2, 77)$=6.01, p=0.004, $MS_{BetGroup}$=0.8411, η^2=0.674; Beg: $F(2, 77)$=2.81, p=0.06. Comparison of means with Bonferroni correction revealed that in the FExp group, the difference between words with 2 and 3 feet was significant (p=0.04), and the difference between words of 3 and 4 feet was also significant (p=0.02). This result should be interpreted with caution, however, since there were only eight instances of 4-footed words, and twenty instances of 2-footed words as compared to fifty-two instances of 3-footed words. For the Beg group, the difference in means by foot length was not significant ($F(2, 77)$=2.81, p=0.067). Word length, as measured in number of syllables, was also not a significant factor in explaining variance.[15] Table 4.1 shows, for example, that the FExp group obtained higher scores on six-syllable words than on four- and five-syllable words. Sentence position has been claimed to be an important input factor in word learning (Barcroft & VanPatten 1997; Shoemaker & Rast 2013). It is therefore a striking fact that position in the sentence was not a significant factor.[16] It may well be that prosodic structure and a recurrent marker facilitated segmentation in sentence-medial position.

Table 4.1 shows that scores of the semantically transparent names were slightly higher than scores of the semantically opaque names (89.1 vs. 85.8). A t-test on these means was significant: $t(362)$=2.58, p=0.01 (two-tailed, d=0.269). It was predicted that only the Beg group would be sensitive to this difference because of their knowledge of German vocabulary. However, this prediction was not borne out because the difference in the means of the two

[15] As a final test of word length, data from both experiments were combined, and separate one-way ANOVAs were computed for each group. Neither test was significant.

[16] A separate one-way ANOVA was computed for each group. For the FExp group: $F(1, 78)$= 0.82, p=0.36, n.s. For the Beg group: $F(1, 78)$=1.10, p=0.29, n.s.

Table 4.3. *Means of response latencies on semantically transparent (T) and opaque (O) compound names of first exposure and beginner L2 learners by measure*

	Trial1		Test			
	T	O (N)	T	O (N)		
FExp	485.68	752.13 (25)	938.66	1030.81 (24)		
Beg	214.57	409.49 (20)	741.65	759.15 (21)		
	Re-test1		Re-test2		Mean	
	T	O	T	O	T	O
FExp	894.33	934.90 (21)	715.71	663.24 (22)	752.13	800.37
Beg	922.38	1007.27 (20)	684.70	635.74 (18)	633.68	697.48

types of nouns was significant for the FExp group but not the Beg group: FExp: $t(184)=2.16$, $p=0.032$ (two-tailed, $d=0.314$); Beg: $t(176)=1.63$, $p=0.10$. Given that beginners were close to ceiling, the lack of effect may be due to their high scores. However, the fact that the first exposure learners appeared to show sensitivity to the difference in structure of the words is unexpected.

Analysis of response latencies showed that both groups responded more slowly to the semantically opaque names than to the transparent ones: FExp transparent names (N=92)=752.13 ms versus opaque names (N=92)=800.37; Beg transparent names (N=77)=633.68 ms versus opaque names (N=77)= 697.48. The largest difference appeared on Training Trial1. See Table 4.3.

Analysis of individual names showed that the opaque name *Shuhmacher* had either the lowest (FExp group) or the second lowest (Beg group) response time. For the FExp group, its mean across measures was 447.50 ms versus a mean latency for all other names of 749.34. For the Beg group, its mean across measures was 443.21 versus a mean latency for all other names of 696.17. This suggests that this name was familiar to all participants. It was removed from the data before the log-transformed latencies were submitted to paired t-tests. The means of the transparent names across measures was 698.16 ms. Participants were slower on the opaque nouns across measures: mean=713.52 ms. The transformed means of the transparent names (9.38) and the opaque names (8.61) was submitted to a paired t-test; the difference was significant, $t(168)=185.44$, $p=0.0000$ (two-tailed). A paired *t*-test on the transformed means of each group was also significant: FExp group $t(91)=132.28$, $p=0.0000$; Beg group $t(76)=131.02$, $p=0.0000$. This suggests that each group was processing the names in the same way.

Discussion and conclusions

This chapter presents data from an exploratory first exposure study using a receptive task to measure word learning. The results presented here are in line with results of prior studies based on this paradigm, showing that cognate stimuli are easy to segment and learn (Carroll 2012a). In particular, participants were able to segment sound forms of words and map them to referents even on the first few items of Training Trial1. The rapid ability to segment sound tokens is now well documented for different L1–L2 pairs and different methodologies (Endress & Bonatti 2007; Gullberg et al. 2010; Rast 2008, 2010; Saffran et al. 1996; Shoemaker & Rast 2013). Although recent work in emergentist approaches to second language learning has insisted that it is slow and incremental (Ellis 1998, 2002; Mellow 2008), the claim may reflect a theoretical bias, rather than an empirical result of L2 learning (Ellis 1998, 2002; Mellow 2008). The claim, as a *general* claim about SLA, is patently false. Not all aspects of language learning are slow; not all aspects of language learning require repeated exposure to L2 input.

Results show on both types of words that once segmentation and sound form-referent mapping has occurred, learners can readily access representations of both cognate words and the first + last names. This is important because it shows that word learning is rapid even with completely novel words. In addition, while beginners were clearly advantaged on Training Trial1 in comparison to the first exposure group (the encoding phase), this advantage did not hold on subsequent measures for the cognates (recall phase).

First exposure learners can segment tokens that are almost twice as long as words used in previous studies (which have tended to use stimuli that are one syllable to three syllables long). This confirms findings by Rast (2008, 2010) that word length is not a factor in segmenting phonetic tokens and is consistent with the literature on early use of formulae in L2 acquisition. Findings presented here stand in stark contrast to the findings of Gullberg et al. (2010). Analysis of the first + last names data suggests that learners may have been sensitive to differences in the prosodic structure of the words. Future research should systematically test for preferences for language-specific minimal prosodic words, effects of lower-level L1 prosodic constituents (feet) and higher-level prosodic constituents (intonational phrases), where relevant.

Beginners required less input than first exposure learners to learn all twenty items and list learning was not affected by the type of word involved. The current study also revealed that retention of all items (either cognate or compound) was good after a delay of two weeks with no exposure to stimuli. This has been a consistent finding of studies using this paradigm and shows that retention does not depend on the cognate status of the words.

A robust finding of this study was that the position of the stimulus word in the utterance did not have an effect on the results. Both groups of learners were equally likely to segment and learn a word in sentence-medial position as in sentence-final position. These results do not show that position in the sentence never plays a role in segmentation and word learning; however, claims that there is an effect (e.g., Barcroft and VanPatten 1997) are not based on solid evidence (but see Shoemaker & Rast 2013). As Carroll (2012b) has emphasized, sentence position is not a well-defined theoretical construct. Moreover, it interacts with a number of linguistic properties such as information structure, focal accent, and prosodic constituent boundaries. When a word appears in a conjunct, as in the current stimuli, and is clearly marked on both sides by prosodic words and by changes in intonational contour, this context may provide ample cues to segmentation processes as to where the sound form begins and ends even though the word is in sentence-medial position. Thus, future studies of linear position would do well to define sentence position in prosodic and morphosyntactic terms.

One question of this study was whether more proficient learners parse compounds into their component lexical parts and are sensitive to the meanings of compounds that such a parse makes available. Such sensitivity, if found, would provide evidence that more proficient learners are analyzing novel input in terms of abstract linguistic units, not just as phonetic tokens. An effect was found, with the transparent items having higher means than the opaque items. Participants were also found to respond faster to the transparent items. However, the significant effect was seen in the data of the wrong group. If parsing the distinction requires prior knowledge of the meanings of the words, this result is odd. A reviewer points out, correctly, that since the task is too easy (beginner scores are above 90 percent on both types of compound names) and both types of compound included cognates that would have increased, in particular, the performance of the first exposure group, another study with better stimuli is warranted. This brings us back to concerns about the naturalness of input that have been raised in some of the first exposure literature (Gullberg et al. 2010). In particular, concerns have been expressed about the generalizability of results from studies with artificial languages and equally artificial tasks (Robinson 2010). Studies that replicate findings with the full complexity of natural languages and ecologically valid tasks are needed, especially given the variability observed across natural languages and the clear influence of L1 knowledge even on first exposure. However, if we are to show that learners are sensitive to particular kinds of *linguistic* distinctions, we may well need to draw on simplified and possibly even artificial input. This is because natural language simply confounds too many properties. Finally, the investigation of frequency effects on word learning needs to be relativized to the type of representation that learners are assumed to be computing. The construct 'word' is too vague to be useful. The study of frequency effects demands that we be clear about what counts as an input for the creation of a particular kind of mental representation. See Carroll (2001) for more general discussion.

To sum up, the paradigm presented here appears to facilitate segmentation, providing learners with a variety of cues that can lead to the creation of phonetic and phonological representations of words. Learners need very little exposure to actually segment sound forms. This study has shown that learners can segment quite long words and that they appear to be influenced by top-down processing effects because of the activation of L1 words by L2 words that sound similar. They may, as well, be drawing on the prosodic categories of the L1. In short, adults bring to the task of learning an L2 a rich set of tools, both in the form of existing representations and in the form of transferrable processing procedures that enable them to immediately get down to the business of creating novel representations. They are able to cognize far more on first exposure than many had perhaps thought possible.

Appendix: Question stimuli (Trials & Test)

COGNATES

1. Ist hier Agnes oder Angela?
2. Ist das Anna oder Anita?
3. Sehen Sie hier Carolina oder Carola?
4. Steht da Charlotte oder Claudia?
5. Ist hier Eva oder Edwina?
6. Ist das Julia oder Johanna?
7. Sehen Sie hier Karin oder Kirsten?
8. Steht da Lauren oder Laura?
9. Ist hier Margarete oder Martina?
10. Ist das Sandra oder Sonja?
11. Sehen Sie hier Albert oder Alexander?
12. Steht da Adolf oder Andreas?
13. Ist hier Bruno oder Bertram?
14. Ist das Erik oder Edmund?
15. Sehen Sie hier Frank oder Franz?
16. Steht da Gregor oder Georg?
17. Ist hier Harald oder Harry?
18. Ist das Hubert oder Hermann?
19. Steht da Josef oder Johannes?
20. Sehen Sie hier Oskar oder Otto?

First + Compound Names (Trials & Test)

1. Ist hier Dagmar Baumgartner oder Trüdel Dieterich?
2. Ist das Ortrud Dahlhoff-Benke oder Gabrielle Blauhemd?
3. Sehen Sie hier Gerda Rotkopf oder Irmtraud Joppke?
4. Steht da Gudrun Käppler oder Ingrid Kleinmund?
5. Ist hier Jana Langbein oder Gisa Grunow?
6. Ist das Jutta Peitz-Seifarth oder Antje Grünwiese?
7. Sehen Sie hier Dörte Kaltwasser oder Korinne Nügel?
8. Steht da Ulla Ruß-Barlösius oder Hannelore Fleischhauer?

9. Sehen Sie hier Jutta Hohlbein oder Anke Abels?
10. Ist hier Frederike Thomé oder Martina Pfannenschmidt?
11. Steht da Axel Braunbart oder Achim Meighörner?
12. Ist hier Lars Girst-Schwillus oder Torsten Blauhut?
13. Ist das Rainer Weisskopf oder Heiner Domsgen?
14. Sehen Sie hier Wilhelm Gröppel-Klein oder Malte Langnase?
15. Steht da Manfred Schnurrbart oder Karsten Hein-Stockhorst?
16. Ist das Ulrich Isenmann oder Heribert Hübschdorn?
17. Ist hier Leo Kleinberger oder Niko Refardt-Hein?
18. Sehen Sie hier Jan Helms-Engelhardt oder Oskar Schuhmacher?
19. Steht da Ludo Hartstein oder Maurus Luckow?

Sehen Sie hier Bodo Minna-Meier oder Remo Wagenbrenner?

REFERENCES

Barcroft, J. and VanPatten, B. (1997). Acoustic salience of grammatical forms: the effect of location, stress, and boundedness on Spanish L2 input processing. In W. R Glass and A. T. Pérez-Leroux (eds.), *Contemporary Perspectives on the Acquisition of Spanish,* Vol. IV: *Production Processing and Comprehension* (pp. 109–121). Somerville, MA: Cascadilla Press.

Braine, M. (1986). Learning the position of words relative to a marker element. *Journal of Experimental Psychology*, *72*, 532–540.

Carroll, S. E. (2001). *Input and Evidence: The Raw Material of Second Language Acquisition*. Amsterdam: John Benjamins.

(2010, May). Exploring the locus of transfer in the L2 pronunciation of novel words. Paper presented at the Annual Meeting of the Canadian Linguistic Association, Concordia University, Montreal, Canada.

(2012a). Segmentation on first exposure to an L2: Evidence for knowledge-driven, top-down processing. In K. Braunmüller, C. Gabriel, and B. Hänel-Faulhaber (eds.), *HSM13. Multilingual Individuals and Multilingual Societies* (pp. 23–45). Amsterdam: John Benjamins.

(2012b). When is input salient? An exploratory study of the sentence location principle and word length. *International Review of Applied Linguistics*, *50*, 1–29.

Carroll, S. E., Jackson, S., and MacDonald, D. M. (2009, May). Adult L2 learners process phonology from first exposure. Paper presented at the annual meeting of Mind-context Divide: Language Acquisition and Interfaces of Cognitive-linguistic Modules, University of Iowa.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences* (2nd edn.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Costa, A., Caramazza, A., and Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *26*, 1283–1296.

Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting*, *5*, 1–18.

(2008). The abstract representations in speech processing. *Quarterly Journal of Experimental Psychology*, *61*, 1601–1619.

Cutler, A. and Broersma, M. (2005). Phonetic precision in listening. In W. J. Hardcastle and J. M. Beck (eds.), *A Figure of Speech: A Festschrift for John Laver* (pp. 93–91). Mahwah, NJ: Lawrence Erlbaum Associates.

Cutler, A. and Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception & Performance*, 113–121.

De Groot, A. M. B. and Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, *50*, 1–56.

Ellis, N.C. (1998). Emergentism, connectionism and language learning. *Language Learning*, *48*, 631–664.

(2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143–188.

(2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty and M. H. Long (eds.), *Handbook of Second Language Acquisition* (pp. 63–103). Oxford: Blackwell.

(2006). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*, 1–24.

Endress, A. and Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream, *Cognition*, *105*, 247–299.

Finn, A. and Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, *108*, 477–499.

Folia, V., Uddén, J., De Vries, M., Forkstam, C., and Petersson, K. M. (2010). Artificial language learning in adults and children. *Language Learning*, *60*, 188–220.

Friel, B. M. and Kennsion, S. M. (2001). Identifying German–English cognates, false cognates, and non-cognates: Methodological issues and descriptive norms. *Bilingualism: Language & Cognition*, *4*, 249–274.

Giegerich, H. J. (1992). *English Phonology: An Introduction*. Cambridge University Press.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436.

Gregory, M. and Carroll, S. E. (1978). *Language and Situation: Language Varieties and Their Social Context*. London: Routledge & Kegan Paul.

Gullberg, M., Roberts, L., Dimroth, C., Veroude, K., and Indefrey, P. (2010). Adult language learning after minimal exposure to an unknown natural language. *Language Learning*, *60*(2), 5–24.

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning*, *24*, 287–298.

Han, Z-H. and Liu, Z. (2013). Input processing of Chinese by ab initio learners. *Second Language Research*, *29*, 145–164.

Jackendoff, R. (1983). *Semantics and Cognition*. Cambridge, MA: MIT Press.

(1990). *Semantic Structures*. Cambridge, MA: MIT Press.

(2002). What's in the lexicon? In S. Nooteboom, F. Weerman, and F. Wijnen (eds.), *Storage and Computation in the Language Faculty* (pp. 23–58). Dordrecht: Kluwer.

(2008). Construction after construction and its theoretical challenges. *Language*, *84*, 8–28.

Katz, J. J. (1977). A proper theory of names. *Philosophical Studies*, *31*, 1–80.

(2001). The end of Millianism: Multiple bearers, improper names and compositional meaning. *Journal of Philosophy*, *98*, 137–166.

Kittleson, M. M., Aguilar, J. M., Tokerud, G. L., Plante, E., and Asbjørnsen, A.E. (2010). Implicit language learning: Adults' ability to segment words in Norwegian. *Bilingualism: Language & Cognition*, *13*, 513–523.

Leplus, T., Langlais, P., and Lapalme, G. (2004). Weather report translation using a translation memory. In R. E. Frederking and K. B. Taylor (eds.), *Machine Translation: From Real Users to Research* (pp. 154–163). Proceedings of the Sixth Conference of the Association for Machine Translation in the Americas. Berlin: Springer Verlag.

Lively, S. E., Pisoni, D. B., and Logan, J. S. (1992). Some effects of training Japanese listeners to identify English /r/ and /l/. In Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (eds.), *Speech Perception, Production and Linguistic Structure* (pp. 175–196). Amsterdam: IOS Press.

Matthews, P. H. (1974). *Morphology: An Introduction to the Theory of Word Structure*. Cambridge University Press.

Mattys, S. L. and Clark, J. H. (2002). Lexical activity in speech processing: evidence from pause detection. *Journal of Memory & Language*, *47*, 343–359.

Mattys, S. L. and Melhorn, J. F. (2007). Sentential, lexical, and acoustic effects on the perception of word boundaries. *Journal of the Acoustical Society of America*, *122*, 554–567.

Mattys, S. L., White, L., and Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology, General*, *134*, 477–500.

McQueen, J. M. (2005). Speech perception. In K. Lamberts and R. Goldstone (eds.), *The Handbook of Cognition* (pp. 255–275). London: Sage.

Mellow, J. D. (2008). Emergentist approaches to teaching and measurement: It's about time. Paper presented at the annual meeting of the Canadian Association of Applied Linguistics, University of British Columbia, Vancouver, B.C., June 4.

Monaghan, P. and Christiansen, M. H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *Journal of Child Language*, *37*, 545–564.

Morgan, J. L. and Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, *66*, 911–936.

Mayo, L., Florentine, M., and Buus, S. (1997). Age of second language acquisition and perception of speech in noise. *Journal of Speech, Language & Hearing Research*, *40*, 686–693.

Nespor, M. and Vogel, I. (1983). Prosodic structure above the word. In A. Cutler and D. R. Ladd (eds.), *Prosody: Models and Measurements* (pp. 123–140). Berlin: Springer.

Nespor, M. and Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Park, E. S. and Han, Z.-H. (2008). Learner spontaneous attention in L2 input processing: An exploratory study. In Z.-H. Han (ed.), *Understanding Second Language Process* (pp. 106–132). Clevedon: Multilingual Matters.

Peña, M., Bonatti, L. L., Nespor, M., and Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*, 604–607.

Perruchet, P., Tyler, M. D., Galland, N., and Peereman, R. (2004). Learning non-adjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology, General*, *133*, 573–583.

Rast, R. (2008). *Foreign Language Input: Initial Processing*. Clevedon: Multilingual Matters.

(2010). The role of linguistic input in the first hours of adult language learning. *Language Learning*, *60*, 64–84.

Rast, R. and Dommergues, J.-Y. (2003). Towards a characterisation of saliency on first exposure to a second language. In S. Foster-Cohen & S. Pekarek Doehler (eds.), *EUROSLA Yearbook: Papers from the Annual Conference of the European Second Language Association*, Vol. III (pp. 131–156). Amsterdam: John Benjamins.

Robinson, P. J. (2010). Implicit artificial grammar and incidental natural second language learning: How comparable are they? *Language Learning*, *60*, Supplement 2, 245–263.

Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996). Word segmentation: the role of distributional cues. *Journal of Memory & Language*, *35*, 606–621.

Selkirk, E. O. (1978). On prosodic structure and its relation to syntactic structure. In T. Fretheim (ed.), *Nordic Prosody II* (pp. 111–140). Trondheim, Norway: Tapir.

Shoemaker, E. and Rast, R. (2013). Extracting words from the speech stream at first exposure. *Second Language Research*, *29*, 165–183.

Valian, V. V. and Coulson, S. (1988). Anchor points in language learning: the role of marker frequency. *Journal of Memory & Language*, *27*, 71–86.

Wong Fillmore, L. (1976). The second time around: Cognitive and social strategies in second language acquisition. Unpublished doctoral dissertation, Stanford University, CA.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge University Press.