Cognition 147 (2016) 70–74

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Short Communication

# Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech

Rebecca L.A. Frost *, Padraic Monaghan

*Department of Psychology, Lancaster University, Lancaster LA1 4YF, UK*

ARTICLE INFO

*Article history:*
Received 8 November 2014
Revised 23 October 2015
Accepted 20 November 2015
Available online 27 November 2015

*Keywords:*
Language acquisition
Artificial grammar learning
Speech segmentation
Grammatical processing
Statistical learning

ABSTRACT

Language learning requires mastering multiple tasks, including segmenting speech to identify words, and learning the syntactic role of these words within sentences. A key question in language acquisition research is the extent to which these tasks are sequential or successive, and consequently whether they may be driven by distinct or similar computations. We explored a classic artificial language learning paradigm, where the language structure is defined in terms of non-adjacent dependencies. We show that participants are able to use the same statistical information at the same time to segment continuous speech to both identify words and to generalise over the structure, when the generalisations were over novel speech that the participants had not previously experienced. We suggest that, in the absence of evidence to the contrary, the most economical explanation for the effects is that speech segmentation and grammatical generalisation are dependent on similar statistical processing mechanisms.



## 1. Introduction

In order to achieve linguistic proficiency, language learners must identify words from continuous speech, and work out the relations between those words, in terms of determining grammatical categories and syntactic structures. However, there are no definitive acoustic cues for word boundaries (Aslin, Woodward, LaMendola, & Bever, 1996), nor of grammatical categories of words that can help determine the syntactic dependencies between words (Monaghan, Christiansen, & Chater, 2007). Thus, learning must operate by somehow determining the regularities that are evident within the language, and how these regularities relate to meaning in terms of defining the relations between words and their mapping to intended referents in the environment (Cunillera, Laine, Camara, & Rodriguez-Fornells, 2010; Monaghan & Mattock, 2012).

There are two views about how these learning tasks proceed in language acquisition. One perspective is that similar statistical mechanisms may apply to speech segmentation and to grammatical processing (Perruchet, Tyler, Galland, & Peereman, 2004; Romberg & Saffran, 2010). An alternative view, deriving from classical cognitive psychology approaches to learning (Chomsky, 1957; Pinker, 1997), is that while speech segmentation is likely to depend on processing statistical dependencies, learning grammar relies on rather different algebraic processes that operate between symbolic representations of elements of language. Previous studies of word identification and grammatical processing have tended to be tested by distinct stimuli, and so comparison across tasks is difficult. However, assessing with the same stimuli word identification and abstraction over these sequences for grammatical processing enables a test of whether processing of these tasks proceeds in tandem or is separated in learning. Though it is not possible to establish for certain whether the same or different processes apply to these tasks, it becomes more challenging to contend that the same statistical process applies to both word identification and grammatical processing if they can be shown to be temporally distinct.

There is good reason to suspect that learning may operate in tandem, because similar sources of information appear to be useful for both segmentation and determining dependencies between words in language acquisition. Monaghan and Christiansen (2010) demonstrated in corpus analyses of child-directed speech that identifying boundaries in speech could usefully rely on determining high-frequency function words that separate other words, forming points of very low transitional probabilities in the speech stream (Ordin & Nespor, 2013). Similarly, they found that these high frequency function words also provide useful markers to the phrase structure of the utterance (Cunillera, Camara, Laine, & Rodriguez-Fornells, 2010), for instance, determiner "the" tends to reliably precede nouns, and pronoun "you" precedes verbs. It is possible that the same sources of information are consulted twice to address these tasks in sequence, but a more economical

* Corresponding author.
*E-mail address:* r.frost1@lancaster.ac.uk (R.L.A. Frost).

http://dx.doi.org/10.1016/j.cognition.2015.11.010
0010-0277/© 2015 Elsevier B.V. All rights reserved.

explanation would be that the same source of information gradually builds up the learners' understanding of what the words are and how they operate in the grammar of the language.

Statistical learning has been proposed as the principle by which speech segmentation and learning grammatical structure may be accomplished (Conway, Bauernschmidt, Huang, & Pisoni, 2010; Lashley, 1951; Redington & Chater, 1997; Rubenstein, 1973). Indeed, transitional probabilities have been found to be effective indicators of word boundaries in both artificial (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997) and natural languages (Pelucchi, Hay, & Saffran, 2009), and can be used to assist learning from infancy onward, even before learners know the meaning of a single word in the language (Saffran et al., 1996; Teinonen, Fellman, Naatanen, Alku, & Huotilainen, 2009). In addition, statistical learning has been shown to be sufficient to account for language learners' acquisition of dependencies between words in sequences (Gerken, 2010; Gómez, 2002; Lany & Gómez, 2008; Lany, Gómez, & Gerken, 2007).

If such statistical processing can be demonstrated to be sufficient for word identification and grammar learning, then this weakens the requirement to posit language-specific mechanisms for language acquisition, instead, a simpler domain-general approach to language learning could be assumed, until evidence to the contrary is ascertained (Christiansen & Chater, 2008). In an ingenious set of studies, Peña, Bonatti, Nespor, and Mehler (2002) set out to show this distinction. They focused on learning of non-adjacent dependencies, which are evident in language structure at multiple levels, from orthography (e.g., final *e* changing the pronunciation of the previous vowel, *cap* and *cape*), morpho-syntax (e.g., I go, *he* goe*s*), grammatical categorisation (e.g., high-frequent non-adjacent pairs of words assist grammatical categorisation of the intervening word, "the __ is", "you __ to" (Mintz, 2003; St Clair, Monaghan, & Christiansen, 2010), and hierarchical grammatical relations (e.g., *the boy* the cats chase *runs*). As Perruchet et al. (2004) note, if statistical dependencies can be shown to be sufficient for acquiring non-adjacencies then this increases the likelihood of the role of domain general statistical processing in language acquisition.

In Peña et al.'s (2002) study, adults were presented with synthetic speech containing items defined by non-adjacent transitional probabilities (e.g. $A_1XC_1$, $A_2XC_2$), where particular A syllables were always paired with particular C syllables, but the X syllable freely varied over a set of three other syllables. To measure speech segmentation, participants were tested on their ability to identify previously occurring words that were consistent with the non-adjacencies presented in the speech, by assessing preference for words (e.g. $A_1XC_1$) over part-words (e.g. $XC_1A_2$).

Critically, Peña et al. (2002) also used these same stimuli to test the extent to which participants could manipulate non-adjacencies to generalise to new items. This involves going beyond the surface form of the sequences, by abstracting the structure to generalise to these new sequences, and is a key property of grammatical processing (Marcus, Vijayan, Rao, & Vishton, 1999). After the same training, they tested a different set of participants on their preference for "rule-words", constructed by moving an A or a C syllable from elsewhere in the speech stream (e.g., placing $A_2$ within the $A_1\_C_1$ non-adjacency: $A_1A_2C_1$), in comparison with part-words. Participants were not able to generalise. However, when the segmentation task was solved for participants, by placing a 25 ms gap between the syllable triples during training, participants did generalise to the rule-words. Peña et al. (2002) thus suggested that although adults are capable of using statistics to identify words from a continuous speech stream, they may then apply separate computations that do not depend on learning statistical dependencies between particular elements of the language, to generalise the structure to consistent forms. They suggest that this can occur only once the task of identifying the words in the stimuli has been solved (Chomsky, 1957; Endress & Bonatti, 2007; Marchetto & Bonatti, in press; Marcus et al., 1999; Miller & Chomsky, 1963).

The interpretation of these results has been hotly debated, but previously the focus of disagreement has been on whether non-adjacencies were learned at all, or rather whether participants instead remembered particular items from the speech (Perruchet et al., 2004), or whether participants learned only the general position of syllables in the sequences rather than the dependencies between them (Endress & Bonatti, 2007; Endress & Mehler, 2009; Mueller, Bahlmann, & Friederici, 2008, 2010; Perruchet et al., 2004). However, there has been substantially less focus on the extent to which segmentation and generalisation of structure co-occur, or are temporally distinct processes.

The Peña et al. (2002) rule-word generalisation stimuli were constructed by moving an A or a C syllable to a new position in the sequence. An advantage of this is that the frequencies of individual syllables were controlled across the target and the part-word stimuli in forced choice tests, so any observed preferences must then be due to syllable co-occurrences, either of adjacent or non-adjacent elements in speech. However, this design may have made generalisation performance harder to detect because it requires not only generalisation of the non-adjacency but also unlearning of the dependency relations for the moved syllable. For instance, the moved-syllable test of Peña et al.'s (2002) study would be analogous to training participants on "the boy the cats chase runs" and "the girl the dog nuzzles smiles", and then testing whether they can flexibly apply the non-adjacency to "the boy smiles runs". Participants may reject these items because they are not able to generalise the non-adjacent structure, or because they fail to accept a violation of relational structure. The observed importance of the pause between syllable triples may then be required not to solve the segmentation task, but rather to increase the salience of syllables with regard to their position (Endress & Mehler, 2009; Perruchet et al., 2004), thus providing an additional cue to relative positions of elements of the language in the speech.

In the current study we tested whether participants are able to simultaneously segment and generalise structure of a non-adjacent dependency language if new, rather than moved, syllables comprise the sequences to be generalised. A novel syllable intervening between an $A_i\_C_i$ dependency is a stronger test of generalisation, but without interference from previous learning of relative syllable positions. Participants listened to a continuous speech stream, and then completed either a test of segmentation, or of generalisation to rule words containing a moved syllable as in Peña et al. (2002). An additional condition tested generalisation to rule words containing novel syllables. If participants are able to use the same information for segmentation and generalisation simultaneously, but were affected by having to unlearn positional information in Peña et al.'s (2002) test of rule-word generalisation, then we expect learning for the novel syllable rule-words in addition to learning for the segmentation task. However, if segmentation and structural generalisation are separable processes, then we expect to see a null effect for the novel syllable generalisation task, with similar performance to that seen in Peña et al.'s (2002) original study.

## 2. Method

### 2.1. Participants

The experiment was completed by 54 adults (8 males, 46 females) with a mean age of 18.52 years (range = 18–24 years). All participants were native-English-speakers, with no known