ChunkitApp

Vetchinnikova, Svetlana

2017

Vetchinnikova , S , Mauranen , A K & Mikusova , N 2017 , ChunkitApp : Investigating the relevant units of online speech processing . in INTERSPEECH 2017 : 18th Annual Conference of the International Speech Communication Association . pp. 811-812 , Interspeech 2017 , Stockholm , Sweden , 20/08/2017 . https://doi.org/10.21437/Interspeech.2017

http://hdl.handle.net/10138/229889
https://doi.org/10.21437/Interspeech.2017

INTERSPEECH 2017: Show & Tell Contribution
August 20–24, 2017, Stockholm, Sweden

ChunkitApp: Investigating the relevant units of online speech processing

Svetlana Vetchinnikova, Anna Mauranen, Nina Mikušová

University of Helsinki, Finland

{svetlana.vetchinnikova, anna.mauranen, nina.mikusova}@helsinki.fi

Abstract

This paper presents a web-based application for tablets 'ChunkitApp' developed to investigate chunking in online speech processing. The design of the app is based on recent theoretical developments in linguistics and cognitive science, and in particular on the suggestions of Linear Unit Grammar [1]. The data collected using the app provides evidence for the reality of online chunking in language processing and the validity of the construct. In addition to experimental uses, the app has potential applications in language education and speech recognition.

Index Terms: web application, tablets, online chunking, speech comprehension, Linear Unit Grammar

1. Introduction

Real-time processing of language is rapid, and relies on chunking up the information flow as it comes in. Current understanding in both linguistics and cognitive science recognizes that humans process language in chunks of some kind. ChunkitApp is a web-based application for tablets developed in project CLUMP (*Chunking in language: units of meaning and processing*) for data collecting. The project explores properties of online chunking while listening to natural speech, as a cross-disciplinary collaboration of linguists and cognitive neuroscientists. It sets out to test the cognitive reality of chunking and to identify its neuronal correlates.

The property of chunking in language processing manifests itself in pervasive patterning observed at all levels of language organization, especially lexis and grammar. Large digital databases of written and spoken text have enabled linguists to identify a large number of conventionalized lexico-grammatical combinations of various kinds [2, 3]. Arguably, such 'chunked' organization emerges over time as a product of recurrent real-time chunking processes. Cognitively, incremental language processing can be put down to limitations of working memory capacity and the need to integrate incoming information in larger units [4].

Capturing the real-time process of chunking requires a new methodological apparatus. In this project, we adopt the model of Linear Unit Grammar (LUG, [1]). It contrasts sharply with most grammatical models, where units of analysis are pre-defined, e.g. sentences, clauses, or constructions. LUG, in turn, postulates an intuitive capacity to process speech incrementally in a way that is crucial for processing meaning. It is a dynamic grammar: it traces chunks as they emerge in real time from the interaction of cognitive, linguistic and physical factors, and imposes an analytical framework on the chunks only after they have been identified. A complex interplay of the factors determining chunk boundaries include memory capacity, predictive processing, previous language exposure, and linearity of text. We hypothesize that the ability to identify relevant chunk boundaries reflects understanding.

ChunkitApp serves to collect data on linear, temporally sequential chunk boundaries. The boundaries, once identified, feed into a subsequent MEG experiment for exploring the neuronal correlates of chunking. This presentation demonstrates the app, with its underlying theoretical framework, and presents experimental findings to test the validity of the construct. We also discuss future potential of the app for educational and speech recognition purposes.

2. ChunkitApp

ChunkitApp makes use of audio files of recorded speech with their transcripts. The transcript of an audio file appears on the screen when it is playing. Each space in the transcript is marked with the symbol '〜' and is clickable. A user is asked to listen to the recordings while following transcripts and "mark boundaries between chunks by clicking '〜' symbols" (Figure 1) Users are to rely on their intuition in chunking; the term 'chunk' is not explained.

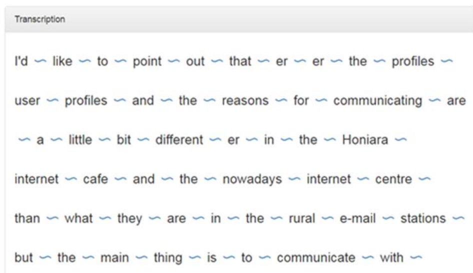

Figure 1: *User interface, ChunkitApp.*

Boundary markings collected lend themselves to both individual and aggregate analysis.

3. Chunking experiment

3.1. Materials

Participants worked with ChunkitApp. They listened to 66/100 short (ca. 30 sec) audio clips of speech. The speech extracts were retrieved from the Corpus of English as a Lingua Franca in Academic Settings (ELFA) [5]. The extracts were short enough to fit the screen of a tablet, avoiding the need to scroll down. Each extract was followed by a self-assessment question: "Did you understand what the speaker was saying?" with three response alternatives: 'yes', 'no', or 'roughly'. The experiment also contained a background questionnaire, a quick proficiency test and a feedback form.

3.2. Participants

Experimental participants were 45 students of the University of Helsinki. None were over the age of 40 (range 20-39) or had any background in linguistics. All were multilingual speakers. In this way, the participants and the speech extracts matched in terms of their likely exposure to English and familiarity with the type of language. As a pilot comparison group, two participants were included who spoke English and were multilingual, but in contrast to the main group, were secondary school students and therefore not familiar with academic English at university level. They were thus less likely to understand the extracts than the main group.

3.3. Procedures

The participants took part in the experiment in small groups. Each worked with an iPad individually, using headphones. The experiment lasted approx. 1.5 hours and included a coffee break. The participants received a movie ticket as a reward for their participation.

3.4. Results

The app received positive feedback from the participants. The majority felt that the task was simple (88%), clear (88%) and for many, fun to do (42%), despite the large number of speech extracts included in the experiment. Also, 65% thought the task reflects in some way what they naturally do when they listen to speech.

3.4.1. Distribution of boundary markings

In total, the 66 extracts included in the analysis here contain 4,799 potential boundaries since any space is open to marking. After removing 4 participants as outliers (see below), the frequencies of boundary markings are distributed as follows: 53% of possible boundaries were left unmarked, (i.e. signalling total agreement on these as no-boundary places); 13% of potential boundaries were marked by just one participant (no agreement); the rest of boundary marking frequencies divide into quartiles as shown in Table 1.

Table 1: *Boundaries in quartiles by frequency of marking.*

Quartiles	Boundary frequency
Insignificant boundary	2-3
Weak boundary	4-6
Medium boundary	7-18
Strong boundary	19-43

3.4.2. Consistency in boundary marking across participants

Individual variation in boundary marking was in evidence. A clear case are the four outliers: the two secondary school pupils and two other participants exhibited chunking behaviour clearly divergent from the rest. They marked boundaries nobody else did. In fact, these four participants together generated 65% of boundaries which were marked just once. It is likely that the secondary school students were not able to follow the meaning.

Among the rest, some participants can be regarded as 'frequent chunkers' who mark boundaries at short intervals (11-13 boundaries per extract on average), others are 'infrequent chunkers', marking only 2-3 boundaries per extract. For example, 31% of weak boundary markings (see Table 1) are generated by the same 5 participants. These 5 chunkers are also the top 5 with the highest average number of boundaries per extract. That is, weak boundaries are marked by frequent chunkers.

In principle, complete agreement on boundary/no boundary would render all frequencies as either 0 or 43. Following this, we can regard all '0' markings as equal to 43, and assign them the value 43 for computing the proportion of total agreement. Multiplying this by the number of possible boundaries (4,799) gives us the theoretical value that could result from total agreement (206,357). The observed boundary markings amount to 128,364, which is 62% of potential total agreement. This value varies from 46% to 74% across different extracts, suggesting some extract-specific variability. In all, agreement on boundaries is high if not complete. This gives strong support to the hypothesis that chunking is intuitive and part of our process of managing incoming speech input.

4. Conclusions

The results support the reality of online chunking in language processing, and the central role of meaning in the process. In line with the hypotheses that (1) chunking is directly related to understanding and (2) chunk boundaries result from the interaction of several factors (cognitive, linguistic, and physical, see Section 1), the data demonstrates a high level of agreement, and also individual differences in boundary marking.

Applicational benefits fall into three main types: (1) The experimental findings together with the positive feedback suggest the app has the potential for further development for educational purposes, particularly in language training and assessment. The potential lies in the ability of the task to differentiate between degrees of understanding talk, specifically in a target register: our findings would help develop fast and focused language tests. (2) Speech recognition models at present are based mainly on acoustic and stochastic information; a typology of chunks and chunk boundaries would help refine current tools by specifying rules of chunking. (3) Finally, the benefit for artificial intelligence and robotics would accrue from identifying the basic units of processing: this supports robots' learning of language comprehension and production in a relevant way. If this is successful, it feeds back into improved understanding of human processing.

5. Acknowledgements

We would like to thank Michal Josífko for programming the app. Our research is funded by the Finnish Cultural Foundation.

6. References

[1] J. Sinclair and A. Mauranen, *Linear unit grammar*. New York: John Benjamins, 2006.

[2] J. Sinclair, "The search for units of meaning," *Textus, vol.* 9, no. 1, pp. 75-106, 1996.

[3] S. Hunston and G. Francis, *Pattern grammar.* Amsterdam; Philadelphia: John Benjamins, 2000.

[4] M. H. Christiansen and N. Chater, "The Now-or-Never Bottleneck: A Fundamental Constraint on Language," *Behavioral and Brain Sciences*, vol. FirstView, pp. 1–52, 2016.