The annotation of Information Structure in spoken Japanese

Emanuela Cresti, Massimo Moneglia
LABLITA, University of Florence, Italy
https://doi.org/10.36505/ExLing-2018/09/0001/000334

Abstract

This paper presents the main results of a pilot aimed at verifying the consistency of the Language into Act Theory model for the annotation of Information Structure in spoken Japanese. The segmentation of the Japanese speech flow into utterances through the detection of terminal prosodic breaks and the segmentation into information unit through non-terminal breaks works fine in Japanese. The main Information unit types characterizing the L-AcT approach (Comment, Topic, Parenthesis, Appendix) and their main properties, fit well with the Japanese data-set.
Key words: Spoken Japanese, Information Structure, Prosodic Structure.

Introduction

This paper presents a pilot aimed at verifying the consistency of the Language into Act Theory model (L-AcT, Cresti 2000; Moneglia & Raso 2014; Cresti & Moneglia 2018a) for the annotation of information structure in Japanese. The pilot is intended at grounding the development of an annotated mini-corpus to be stored in the IPIC Database (Panunzi & Gregori 2012) which is devoted to the Cross-linguistic Comparison of Information Structure. At present, IPIC stores resources of Italian, Brazilian and Spanish (Panunzi & Malvessi-Mittmman 2014; Nicolas-Martinez & Lombán fothcaming)

The Japanese data set relies on the Nagoya University Conversation Corpus - NUCC (Fujimura et al. 2012) and corresponds to approx. 80 hours of conversation for 1.5 million transcribed morphemes. Transcripts are in Japanese characters, recently automatically transliterated into Latin characters. NUCC contains 129 natural dialogues and conversations between friends, family members and colleagues, representing a large variety of contexts. For this reason, it can be the source of a selection of samples fitting with the IPIC corpus design model (Cresti & Fujimura 2018). The pilot considers around 100 excerpts derived from four recordings.[1]

L-AcT model foresee the alignment of each utterance in the corpus to its acoustic counterpart through the speech software WinPitch and the annotation of information structure according to a specific methodology and tagset (Moneglia & Raso 2014). In 2. we will briefly sketch the main assumptions for what regards the prosodic cues that are necessary to this end and we will verify the consistency of this model to the Japanese data set. More specifically, in 3 we will consider the criteria for the segmentation of the speech flow into

ExLing 2018: Proceedings of 9th Tutorial and Research Workshop on Experimental Linguistics, 28-30 August, Paris, Frannce

utterances and in 4. the segmentation of the utterance into information unit types.

The general feature of the L-AcT model

L-AcT assumes that the speech flow can be segmented into reference units by means of both pragmatic and prosodic cues. In this framework, a reference unit may belong to two types, respectively *utterance* and *stanza*. The utterance is defined as the counterpart of a Speech act (Austin 1962) and is the primary reference unit for speech (Biber et al. 1999). A stanza expresses the flow of thought (Chafe 1994) and corresponds to a sequence of week speech activities packaged together.[2] The boundaries of both reference units are marked by *prosodic breaks* ('t Hart et al. 1990; Swerts 1999) that are perceived with the quality of being *terminal* (Moneglia & Cresti 2005)

Every reference unit is composed of an Information Pattern which can be simple or complex. Each unit of the pattern necessarily corresponds to a prosodic unit. The prosodic units of a complex pattern are separated the one from the other by *non-terminal breaks*.

Given that information units match in one-to-one way to prosodic units, the prosodic annotation grounds the identification of information units in the flow of speech. Therefore, in order the model to be applied in a language, two preliminary operations are compulsory: 1) identification of terminal breaks; 2) identification of non-terminal breaks.

According to L-AcT, the core of the Information pattern is one specific information unit (Comment) devoted to the expression of illocutionary force. For this reason, a Comment unit is necessary and sufficient for a complete Information pattern. The latter may be simple, which is to say composed of only a Comment or complex. In Complex utterances other optional information unit types may support the Comment, each one corresponding to a dedicated prosodic unit and to a specific information function. Information functions are classified into two basic types, depending on whether they work in fulfilling the semantic content of the utterance or in its communicative support (Discourse markers).

Information unit types with their tags and formal definitions are detailed in Moneglia & Raso 2014. The aim of the pilot is to verify the Adequacy of L-AcT model for the segmentation of spoken Japanese according to key operational principles. We will verify breaks detection, the consistency of the Comment principle, and the consistency of the main textual Information functions; i.e Topic (TOP), Parenthesis (PAR); Appendix (APC).

Terminal breaks, non-terminal breaks and the pragmatic independence of the reference unit

Although major prosodic breaks are prominent also to non-natives, they cannot really judge their terminal or non-terminal nature. The following two examples

correspond to opposite judgements given by non-natives, both not fitting with speech act performance. The major break in figure 1, which is connected to a rising contour, is perceived as a continuation, while the major boundary in figure 2, showing a falling contour, is perceived terminal.

(1) J1:
十三？　うち 十三…
jusan ?　uchi　jusan::
thirteenth ? we　thirteenth::
'tirteenth?' 'we (are) thirteenth…'

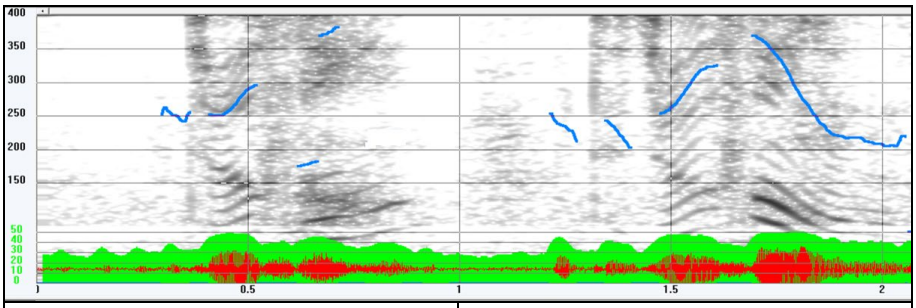

Figure 1. Terminal break with rising contour.

(2) *M3A18:
もう あんた 今ごろ 全部, 葉っぱ-が 出そろって な-あかん　よ。
mou anta　imagoro　　zenbu　/ happa-ga　desorotte　na-akan　yo //
already　you now　　every　/ leave-SUB come-out must　PR FIN //
'As a whole for now / leaves had to be already born'

As the transcript shows, competent speakers easily recognise that the first break in (2) is terminal, since it corresponds to a concluded speech activity (*request of confirmation*) that is followed by a second speech activity (*supposition*). If a stretch of speech can be interpreted in isolation as a speech act the prosodic break is judged terminal.

On the other way around, in (3) competent speaker do not assign the value of independent speech act to the first prosodic unit. The break is perceived non-terminal since it cannot be interpreted in isolation, and the prosodic unit is considered part of a sequence interpreted as *self-conclusion.*

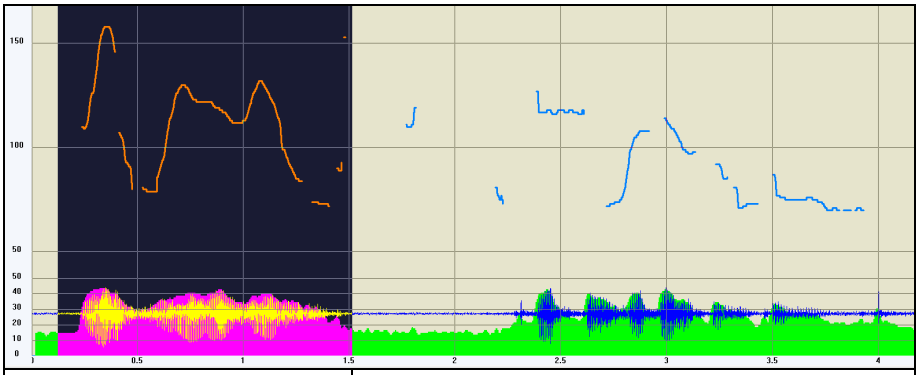

Figure 2. Terminal break with falling contour.

Therefore, the identification of the terminal quality in a major boundary does not follow from intrinsic prosodic properties (rising vs falling boundary tones), but strictly requires the access to the language competence which grounds the pragmatic interpretation. Based on this competence the linguist determines whether the prosodic unit can be interpreted or not in isolation. When it doesn't, the unit is part of a larger utterance and the perceived prosodic break is considered non-terminal. Therefore, the assignment of a value to prosodic breaks and pragmatic judgements go hand in hand.

The Comment principle and the structure of information within the reference unit

L-AcT foresees that when the utterance is segmented into information units these are marked by prosodic boundaries. (2) as well as the following examples, allow to verify: a) the segmentation of the utterance into information units according to non-terminal breaks detection; b) the correspondence of the information units to the typology of information function foreseen in L-AcT.

As we will see, the first prosodic unit in (2) corresponds to a Topic unit of a complex utterance, however, what is more interesting in (2) to our ends is the nature of the second unit. L-AcT assumes that within an utterance, characterized by an illocutionary value, one and only one unit identifies the information unit bearing the illocutionary information. We call Comment this unit.

This core assumption of the theory is confirmed in (2). Indeed, listening in isolation to its second unit, competent speakers find that it can receive a pragmatic interpretation. The Comment principle hold in all terminated sequences of the pilot, grounding the application of the L-AcT model. For instance, let's consider the following dialogue between wife and husband, where she complains about a delay in planting tulips and the husband note that indeed nothing flourished.