

Fig. 2. (A) Red and blue bars represent completedness judgments of musicians and non-musicians, respectively. B and C depict completedness judgment of both groups. There were no main effects for Group, though there were for Tonality (B) and End-type (C), revealing that both tonal melodies and tonic endings elicited a stronger sense of completeness than atonal melodies and subdominant endings did. X bars represent the proportion of responses in the forced-choice task (completed or uncompleted). Error bars reflect standard error of the mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

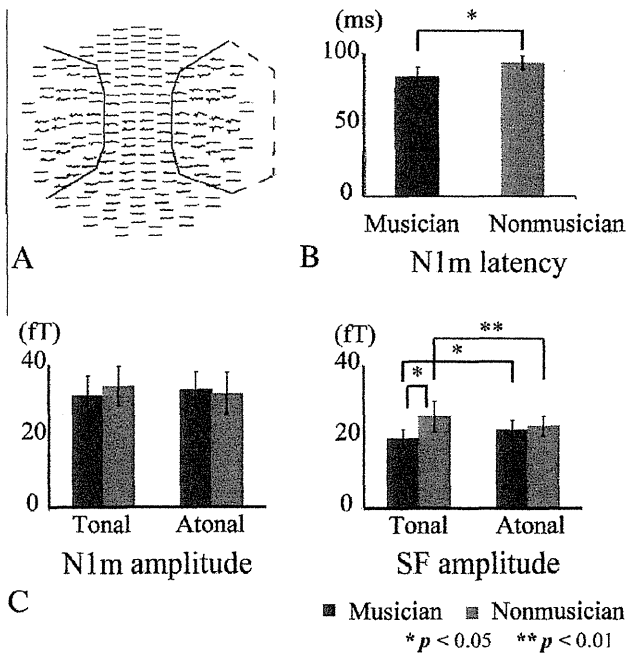


Fig. 3. MEG waveforms and the blacked line in the upper figures indicates 22 planar sensor pairs (44 channels) that were used for analysis (A). (B) ANOVA results of N1 latency. N1m latency is significantly faster in musicians than in non-musicians. (C) In SF amplitude, the responses of SF musicians were significantly larger in tonal conditions than in atonal conditions. In contrast, the opposite trend was observed in non-musicians. Error bars reflect standard error of the mean.

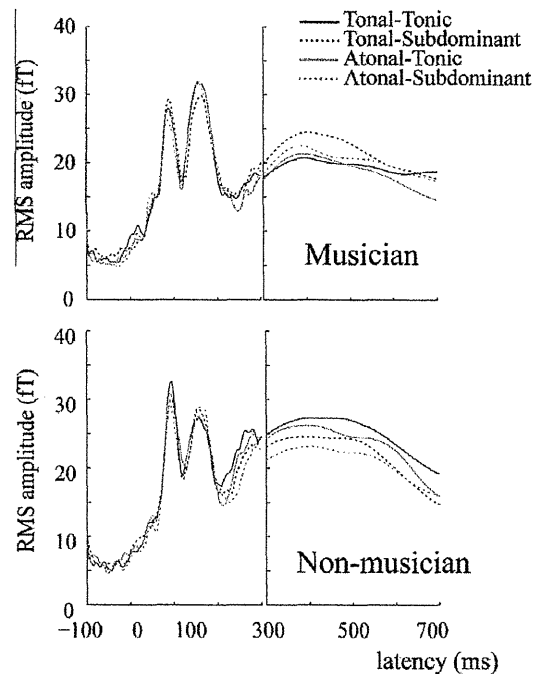


Fig. 4. Grand-averaged RMS waveforms of 44 predetermined channels (22 channels for each hemisphere). Because we applied two different low-pass filters, figures for -100–300 and 300–700 ms are shown separately.

between group and tonality ( $F(1, 16) = 17.29, p < 0.001$ ). Post-hoc analysis revealed that non-musicians had a larger peak amplitude in response to tonal melody than musicians did ( $F(1, 1) = 5.77,$

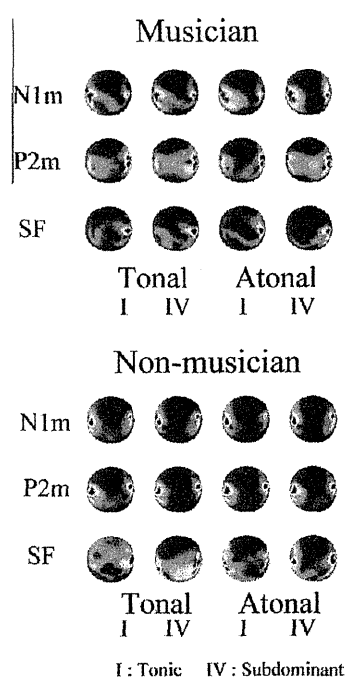
$p < 0.05$ ). The responses of musicians were larger in tonal melody than in atonal melody ( $F(1, 1) = 9.57, p < 0.001$ ). In contrast, the responses to atonal melody were larger in non-musicians than in musicians ( $F(1, 1) = 7.78, p < 0.05$ ) (Figs. 3B and 4).

**3.2.3.1. ANOVA results using averaged data of SF window.** As Fig. 3B shows, SF amplitude appears to be more sustained in non-musicians than in musicians. To verify this tendency, we also conducted repeated ANOVA for the average value of a 50-ms time window. Significant interaction between group and tonality was observed for a latency of 300–700 ms in the right hemisphere (300–350 ms: ( $F(1, 16) = 10.11, p < 0.01$ ), 350–400 ms: ( $F(1, 16) = 14.97, p < 0.005$ ), 400–450 ms: ( $F(1, 16) = 13.70, p < 0.005$ ), 450–500 ms: ( $F(1, 16) = 14.35, p < 0.005$ ), 500–550 ms: ( $F(1, 16) = 16.24, p < 0.005$ ), 550–600 ms: ( $F(1, 16) = 13.46, p < 0.005$ ), 600–650 ms: ( $F(1, 16) = 8.55, p < 0.01$ ), 650–700 ms: ( $F(1, 16) = 5.53, p < 0.05$ )). Post-hoc analysis revealed that amplitude in response to tonal melody was significantly larger in non-musicians than in musicians (300–600: 300–350 ms: ( $F(1, 1) = 6.03, p < 0.05$ ), 350–400 ms: ( $F(1, 1) = 4.96, p < 0.05$ ), 400–450 ms: ( $F(1, 1) = 4.56, p < 0.05$ ), 450–500 ms: ( $F(1, 1) = 4.57, p < 0.05$ ), 500–550 ms: ( $F(1, 1) = 4.74, p < 0.05$ ), 550–600 ms: ( $F(1, 1) = 5.17, p < 0.05$ )).

The responses of musicians were larger in tonal melody than in atonal melody (300–550 ms) while the responses of non-musicians were larger in atonal melody than in tonal melody (300–700 ms).

### 3.3. Dipole location

We calculated the  $x$  (medial–lateral),  $y$  (posterior–anterior), and  $z$  (inferior–superior) coordinates of the ECDs for the N1m, P2m and SF peaks in each group, stimulus, and hemisphere, respectively. In order to evaluate the dynamics of dipole location through time, the coordinates of each component were averaged across stimulus conditions. Fig. 5 shows the contour maps and 6A shows the calculated equivalent current dipole locations of one representative subject overlaid onto the individual MRI brain reconstruction. The group-averaged source locations of the N1m, P2m, and SF components are also plotted on an  $x$ – $y$  (axial) plane (Fig. 6B). The mean locations of ECDs across subjects with different peaks were



**Fig. 5.** Shows contour maps of the four stimulus types in representative participants. Red and blue markers indicate averaged dipole location in musicians and non-musicians, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

confined to a small volume  $\sim 8$  mm lengths in  $x$ ,  $y$ , and  $z$  coordinates, although intersubject variation existed.

For statistical evaluations of estimated locations between groups, we conducted  $t$ -tests between musicians and non-musicians across every component. There was no significant difference between groups in any component. Then, to examine the effect of statistical location shift regardless of group, we averaged the coordinates of the last note and conducted ANOVA to examine any difference in the  $x$ ,  $y$ , or  $z$  values among N1m, P2m, and SF components. Location difference was found in  $x$  and  $y$  coordinates across components in both hemispheres, indicating that the P2m ECDs had larger  $y$  coordinates than did the N1m ECDs and SF ECDs, which means that the ECD of P2m response was located anterior to the ECD of SF or N1m response in the left hemisphere. Another difference found that the N1m ECDs had larger  $x$  coordinates than did P2m ECDs in the right hemisphere, indicating that the N1m responses were located lateral to the ECDs of P2m responses (Fig. 6C).

N1m estimated sources tend to be more lateral to those of P2m and P2m estimated sources were located anterior to those of N1m. These location shifts were consistent with those seen in a previous study (Kuriki et al., 2006).

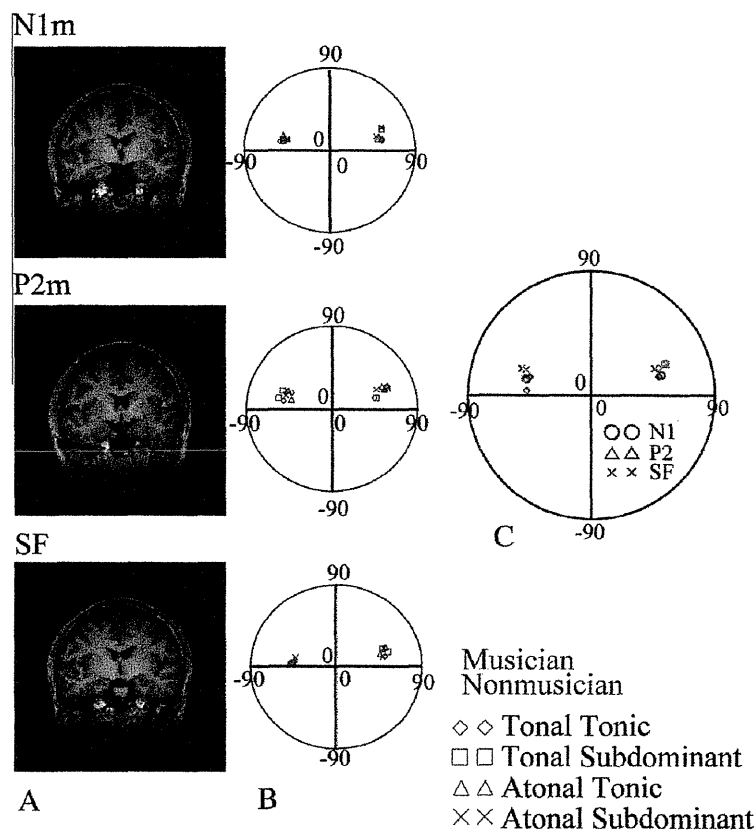
Generally, ECDs were found in the vicinity of Heschl's gyrus (HG) and the perisylvian area. Dipoles of P2m and SF were also localized in the inferior prefrontal area and the insular area. None of the variables was consistently found to influence the location of ECDs, but these dipoles were localized in the area that is responsible for tonal processing.

## 4. Discussion

The current study examined whether there was any difference in the neural responses of musicians and non-musicians to tonic and subdominant melody endings given no difference in their behavioral responses. We also examined whether the context provided by the preceding melody influenced the sense of the melody's completion. Behavioral results indicated that the judgment of a melody's completeness is similar in musicians and non-musicians. Tonal melodies with tonic endings were judged to sound the most complete, while atonal melodies without tonic endings were judged to sound the least complete. In contrast, neural responses recorded by MEG were different between musicians and non-musicians in each of two components, N1m and SF. The latency of N1m, which is sensitive to the physical aspects of tones (Näätänen and Picton, 1987), was significantly shorter in musicians than in non-musicians. On the contrary, SF, which is suggested to be relate to sustained attention and cognitive burden (Minati et al., 2008, 2009, 2010), was stronger and more sustained in non-musicians than in musicians. These findings suggest that different processing mechanisms are at work in musicians and non-musicians, even when there is no apparent difference in their behavioral responses to the melodies.

### 4.1. Differences in evoked responses between musicians and non-musicians

First, the latency of N1m was shorter in musicians than in non-musicians in the right hemisphere. The different latencies infer that the melody process in the present study is different between musicians and non-musicians. N1m is sensitive to the physical aspects of tones: its amplitude is positively correlated with the strength of a tone (Siedenberg et al., 1996) and influenced by the frequency of a tone: N1m responses are generated by tonotopically organized neuronal populations (Woods et al., 1993). In addition to being sensitive to a tone's physical aspects, N1m is also sensitive to



**Fig. 6.** (A) Dipole location on the MRI surface. (B) Calculated equivalent current dipole locations of one representative subject overlaid onto that individual's MRI brain reconstruction. (C) The group-averaged source locations of the N1m, P2m, and SF components are also plotted on an  $x$ - $y$  (axial) plane.

the cognitive factors of listeners including attention and expectation (Hillyard et al., 1973; Näätänen and Picton, 1987; Berman and Friedman, 1995; Helenius et al., 2002; Roß et al., 2002; Lange, 2009; Ohoyama et al., 2012); expectations in particular are strongly affected by the effects of training. It is also reported that the amplitude of N1m is correlated with the musicality score and with the paradigm's difficulty (Alho et al., 2012). In their study, differences between musicians and non-musicians were elicited only when both participant groups were engaged in melody oddball paradigm not but ordinary oddball paradigm, suggesting that neural differences are derived from music ability (Alho et al., 2012). It is also well known that listening strategy differs between musicians and non-musicians (Fujioka et al., 2005; Neuhaus et al., 2006). A musician relies more heavily on an expectative way of listening than a non-musician does (Pantev et al., 1998, 2001; Shahin et al., 2003; Fujioka et al., 2004, 2005; Tervaniemi et al., 2009; Strait et al., 2010). When the subject uses expectative listening, as when, for example, an identical stimulus is presented repeatedly, a strong attenuation of N1m/N1 occurs (Siedenberg et al., 1996; Roß et al., 2005). This phenomenon is due not to any adaptation of auditory cortical neurons but rather to an effect of expectation, namely, template matching. When listeners can easily predict the next stimulus, cortical activity declines.

In the present experiment, the physical aspects of the tones were completely equal between conditions, allowing us to consider other factors that may influence N1m latencies other than the physical aspects of the tone. Although we did not present the same stimulus repeatedly, we presented stimulus sequences based on musical rules, so that listeners with knowledge about musical rules could easily predict the ending tone. The music experts who participated in this study fall into this category because almost all of

them said that they noticed the presentation rule in this study and accordingly listened with expectation. The amplitude and latency of N1 is enhanced by memory-based attention (Ohoyama et al., 2012). The differences between musicians and non-musicians in N1m amplitude and its latency are possibly caused by difference of memory-based attention.

Second, the amplitude of SF was significantly larger in non-musicians than in musicians in response to tonal melodies in the right hemisphere. In non-musicians it was greater in response to tonal melodies than to atonal melodies. In contrast, the responses to tonal melodies were significantly smaller than those to atonal melodies in musicians. SF, which is essentially related to the stimulus duration, has been demonstrated to be enhanced by attention, especially continuous attention during an entire sound signal presentation (Roß et al., 2004; Okamoto et al., 2011). In addition, enhanced amplitude of SF has been also suggested to reflect stimulus complexity and task difficulty (Gunji et al., 2003). Minati et al. (2008, 2009) have demonstrated that a sustained late component whose latency window and amplitude are quite similar to those of our SF was prolonged as the structural complexity of a melody and the corresponding cognitive burden, requiring the gradual formation of a melodic gestalt, increased (Minati et al., 2008, 2009, 2010). In the field of language research, it is reported that facilitating factors like semantic or phonological priming tend to reduce the amplitude of the N400 component, whose time window and sustained waveforms are also similar to those of our SF (Helenius et al., 2002).

In the present study, again, although structural complexity was identical among all stimuli, our data on the interaction between group and tonality suggest that tonal information elicits different neural processing patterns in different groups. Possibly, the same

behavioral task constitutes a more difficult task and a greater cognitive burden for non-musicians than for musicians. For musicians, it is easier to predict the ending pattern in a tonal melody than in an atonal melody, so that this task elicits a higher SF amplitude for an atonal melody than for a tonal melody. For non-musicians, in contrast, judging the completeness of a melody is a difficult task. They might focus on the tonal melody in order to judge completeness, therefore eliciting a higher SF amplitude for a tonal melody than for an atonal melody. This intergroup difference in response to tonal melody is also explainable on the basis of differences in tonal processing between the groups.

We found a similar interaction in N1m amplitude, but it did not obtain significance in *post hoc* analysis. It might be due to the style in which we presented each melody; specifically, our end tones consisted of pairs of the same semitone. Musicians were therefore more likely to expect the final tone from the tone before the final tone.

The difference in SF between musicians and non-musicians contrasts with the shorter N1m in musicians, suggesting that different melodic processing techniques are used by musicians and non-musicians.

In the current study, there was no significant difference in the P2m component between musicians and non-musicians or between melody types. Whether P2 is modulated by training is still a matter of controversy (Tremblay et al., 2001; Shahin et al., 2003; Bosnyak et al., 2004). A large N1 amplitude is sometimes accompanied by a large P2 amplitude, but not always, especially in an MEG study. There have been many articles reporting attention modulation or other differences between musicians and non-musicians in N1m and SF without P2m, especially in MEG experiments (Roß et al., 2002, 2005; Kuriki et al., 2005, 2006; Alho et al., 2012). It might be difficult to elucidate a slight difference between musicians and non-musicians in a study like the present one because the RMS increase for P2 is less prominent in MEG than in EEG (Shahin et al., 2007).

#### 4.2. Preceding melody effect

Our behavioral results clearly showed that tonal melodies with tonic endings (Tonal-Tonic) elicited a strong sense of completeness, and that this sense of completeness was attenuated as tonality was reduced. Based on these findings, the sense of completeness is considered to be unstable and easily influenced by both the tonality of the preceding melody and that of the ending pattern. Partially tonal melodies and endings (as in Tonal-Subdominant, Atonal-Tonic) might evoke a limited sense of completeness. The sense of completeness was dramatically attenuated when an atonal melody was paired with a subdominant ending (Atonal-Subdominant). This behavioral tendency demonstrates that the degree of completion does not depend on the local harmonic and acoustic features of the final two tones, but rather on the interaction between the preceding melody and the final tones (Bigand and Pineau, 1997). Furthermore, this tendency to rate more tonal melody-ending combinations as more complete was found to occur to a similar degree in both musicians and non-musicians, without any significant difference according to our statistical analyses. This suggests that the process of judging the completeness of a melody may remain unchanged after musical training, in keeping with previous reports showing that learning musical structure does not require explicit training (Bigand and Pineau, 1997).

Because we did not require our participants to judge the completeness of the melodies as quickly as possible, behavioral indices such as response time were not available for the study of additional possible differences between musicians and non-musicians. Our results, however, are consistent with previous reports showing

that even non-musicians can process tonal musical information (Koelsch et al., 2000; Koelsch and Siebel, 2005).

#### 4.3. Deviant detection

Most of the preceding studies of event-related potentials have employed tasks in which the target stimulus was presented less frequently, in order to evaluate the ability to detect the target. In the present study, in contrast, we presented four kinds of stimuli, each with the same probability. In other words, the tonal melody with the tonic ending and the other three melodies with their gradual changes in tonality of melody and ending were each presented one-fourth of the time. Therefore, given the physical similarity of the stimulus conditions, the differences between responses to tonal melodies and responses to atonal melodies are probably derived from the listener's expectation of the ending.

Music syntactic experiments repeatedly demonstrated that unexpected harmonies may evoke right lateralized ERAN or EN even when a musical stimulus is ignored and participants had no special musical expertise (Koelsch et al., 2002; Koelsch and Siebel, 2005). In the present study, we observed decreased N1m latency in the right hemisphere and differently enhanced SF amplitude between musicians and non-musicians in the right hemisphere, suggesting right lateralized music processing.

#### 4.4. Stimulus conditions

In the present experiment, we used an identical sequence of two final tones for all combinations of melody and ending types, and changed the melody and ending types by manipulating the preceding melody. Conventional experiments on musical perception usually replace only the last cadence with an unusual cadence or tone. That experimental method, however, is not ideal for our goal of investigating whether musical context influences the sense of a melody's completeness regardless of local differences in the final chord. Neural responses elicited by a single dissonant chord or minor chord in the ending may not be the same as those elicited by a single consonant chord (Blood et al., 1999; Mizuno and Sugishita, 2007; Green et al., 2008; Virtala et al., 2010). The emotions elicited by dissonant chords are also different from those elicited by consonant chords (Blood et al., 1999; Tervaniemi and Brattico, 2004; Pallesen et al., 2005). Even a consonant stimulus can be perceived as a deviant stimulus when it follows a chord in a different key, because of the juxtaposition of different harmonies (Bergelson and Idsardi, 2009). When presented in a musical context, the sensory dissonance of a chord also depends on its harmonic pitch relationship with preceding chords (Regnault et al., 2001).

The previous experiments on this topic could not eliminate the possibility of differences in response caused by differences in the final chords themselves, because the ending tones or chords of the stimuli were always physically different from those of the control stimuli (tonic chords versus dissonant/unrelated chords). In our experimental paradigm, on the other hand, we used an acoustically identical ending pattern with each stimulus in order to eliminate the possibility of differences in neural responses due to the use of different ending tones. By this means we observed neural responses indicating processing using information from the relationship between context of the preceding melody and the end of the melody.

Harmonic expectancy is generated in both bottom-up (i.e., results in peripheral analyses of auditory information evoke expectation of following musical sequence in the central structures) and top-down (i.e., template matching) manners (Narumour, 1991; Regnault et al., 2001). The bottom-up manner of listening reflects the influence of perceptual organization based on acoustical analysis of the target stimuli, while the top-down manner of listening

reflects the importance of the listeners' knowledge of a given musical idiom such as tonal-harmonic hierarchy (Regnault et al., 2001). In the present study, we set out to assess a type of selective attention, namely, the expectation of a particular type of musical ending (top-down listening based on the knowledge of musical rule), which is not realized by the bottom-up manner of listening, and also to assess specifically the effect of training on the top-down listening process. Indeed, it has been shown that top-down listening affects bottom-up sensory input and controlled cortical re-organization in adult rats (Ohoyama et al., 2012). The effect of selective attention has also been confirmed in a human imaging study, which suggested that attention focused on one auditory component intensifies its representation (Debener et al., 2003; Okamoto et al., 2007). For this purpose we compared responses to the same physical ending tone sequence paired with different preceding melodies.

## 5. Conclusion

Musicians and non-musicians rated the sense of a melody's completeness similarly when they listened to stimuli composed of various combinations of tonal and atonal melodies and tonic and subdominant endings. Musicians and non-musicians did exhibit differences in amplitude of both the N1m and SF components of MEG waves. The present findings suggest that, in assessing a melody's completion based on the combination of the preceding melody and the ending tone, musicians and non-musicians may undergo distinct neural processes, even when the two groups assess the sense of completion similarly.

## Acknowledgments

We thank Dr. Kuniyoshi, L. Sakai and Dr. Asuka Otsuka for valuable comments on the experiment. This work was supported by the Global Center of Education and Research for Chemical Biology of the Diseases, the GCOE Program.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.clinph.2013.09.027>.

## References

- Alho K, Salonen J, Rinne T, Medvedev SV, Hugdahl K, Hämäläinen H. Attention-related modulation of auditory-cortex responses to speech sounds during dichotic listening. *Brain Res* 2012;1442:47–54.
- Baumann S, Meyer M, Jäncke L. Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *J Cogn Neurosci* 2008;20:2238–49.
- Bergelson E, Idsardi WJ. A neurophysiological study into the foundations of tonal harmony. *Neuroreport* 2009;20:239–44.
- Berman S, Friedman D. The development of selective attention as reflected by event-related brain potentials. *J Exp Child Psychol* 1995;59:1–31.
- Bigand E, Pineau M. Global context effects on musical expectancy. *Percept Psychophys* 1997;59:1098–107.
- Blood AJ, Zatorre RJ, Bermudez P, Evans AC. Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nat Neurosci* 1999;2:382–7.
- Bosnyak DJ, Eaton RA, Roberts LE. Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cereb Cortex* 2004;14:1088–99.
- Brattico E, Tervaniemi M, Näätänen R, Peretz I. Musical scale properties are automatically processed in the human auditory cortex. *Brain Res* 2006;1117:162–74.
- Debener S, Herrmann CS, Kranczoch C, Gembris D, Engel AK. Top-down attentional processing enhances auditory evoked gamma band activity. *Neuroreport* 2003;14:683–6.
- Fujioka T, Trainor LJ, Roß B, Kakigi R, Pantev C. Musical training enhances automatic encoding of melodic contour and interval structure. *J Cogn Neurosci* 2004;16:1010–21.
- Fujioka T, Trainor LJ, Roß B, Kakigi R, Pantev C. Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J Cogn Neurosci* 2005;17:1578–92.
- Green AC, Baerentsen KB, Stødkilde-Jørgensen H, Wallentin M, Roepstorff A, Vuust P. Music in minor activates limbic structures: a relationship with dissonance? *Neuroreport* 2008;19:711–5.
- Gunji A, Koyama S, Ishii R, Levy D, Okamoto H, Kakigi R, et al. Magnetoencephalographic study of the cortical activity elicited by human voice. *Neurosci Lett* 2003;348:13–6.
- Hashimoto T, Hirata Y, Kuriki S. Auditory cortex responds in 100 ms to incongruity of melody. *Neuroreport* 2000;11:2799–801.
- Helenius P, Salmelin R, Service E, Connolly JF, Leinonen S, Lyytinen H. Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. *J Neurosci* 2002;22:2936–44.
- Henson RN, Mouchlianitis E, Friston KJ. MEG and EEG data fusion: simultaneous localisation of face-evoked responses. *Neuroimage* 2009;47:581–9.
- Hillyard SA, Hink RF, Schwent VL, Picton TW. Electrical signs of selective attention in the human brain. *Science* 1973;182:177.
- Koelsch S, Grossmann T, Gunter TC, Hahne A, Schröger E, Friederici AD. Children processing music: electric brain responses reveal musical competence and gender differences. *J Cogn Neurosci* 2003;15:683–93.
- Koelsch S, Gunter T, Friederici AD, Schröger E. Brain indices of music processing: “nonmusicians” are musical. *J Cogn Neurosci* 2000;12:520–41.
- Koelsch S, Schröger E, Gunter TC. Music matters: preattentive musicality of the human brain. *Psychophysiology* 2002;39:38–48.
- Koelsch S, Siebel WA. Towards a neural basis of music perception. *Trends Cogn Sci* 2005;9:578–84.
- Kuriki S, Isahai N, Ohtsuka A. Spatiotemporal characteristics of the neural activities processing consonant/dissonant tones in melody. *Exp Brain Res* 2005;162:46–55.
- Kuriki S, Kanda S, Hirata Y. Effects of musical experience on different components of meg responses elicited by sequential piano-tones and chords. *J Neurosci* 2006;26:4046–53.
- Lange K. Brain correlates of early auditory processing are attenuated by expectations for time and pitch. *Brain Cogn* 2009;69:127–37.
- Madsen CK, Geringer JM. Differential patterns of music listening: focus of attention of musicians versus nonmusicians. *Bull Counc Res Music Educ* 1990;105:45–57.
- Minari L, Rosazza C, D'Incerti L, Pietrocini E, Valentini L, Scaioni V, et al. fMRI/ERP of musical syntax: comparison of melodies and unstructured note sequences. *Neuroreport* 2008;19:1381–5.
- Minari L, Rosazza C, D'Incerti L, Pietrocini E, Valentini L, Scaioni V, et al. Functional MRI/Event-related potential study of sensory consonance and dissonance in musicians and nonmusicians. *Neuroreport* 2009;20:87–92.
- Minari L, Salvatori L, Rosazza C, Pietrocini E, Visani E, Panzica F, et al. Event-related potential (ERP) markers of melodic processing: The N2 component is modulated by structural complexity, not by melodic ‘meaningfulness’. *Brain Res Bull* 2010;83:23–8.
- Mizuno T, Sugishita M. Neural correlates underlying perception of tonality-related emotional contents. *Neuroreport* 2007;18:1651–5.
- Näätänen R, Picton T. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 1987;24:375–425.
- Narumour E. The top-down and bottom-up systems of musical implication: Building on Meyer's theory of emotional syntax. *Music Percept* 1991;9:1–26.
- Neuhaus C, Knösche TR, Friederici AD. Effects of musical expertise and boundary markers on phrase perception in music. *J Cogn Neurosci* 2006;18:472–93.
- Oberecker R, Friedrich M, Friederici AD. Neural correlates of syntactic processing in two-year-olds. *J Cogn Neurosci* 2005;17:1667–78.
- Ohoyama K, Motomura E, Inui K, Nishihara M, Otsuru N, Oi M, et al. Memory-based pre-attentive auditory N1 elicited by sound movement. *Neurosci Res* 2012;73:248–51.
- Okamoto H, Stracke H, Bermudez P, Pantev C. Sound processing hierarchy within human auditory cortex. *J Cogn Neurosci* 2011;23:1855–63.
- Okamoto H, Stracke H, Wolters CH, Schmael F, Pantev C. Attention improves population-level frequency tuning in human auditory cortex. *J Neurosci* 2007;27:10383–90.
- Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 1971;9:97–113.
- Otsuka A, Tamaki Y, Kuriki S. Neuromagnetic responses in silence after musical chord sequences. *Neuroreport* 2008;19:1637–41.
- Pallesen KJ, Brattico E, Bailey C, Korvenoja A, Koivisto J, Gjedde A, et al. Emotion processing of major, minor, and dissonant chords. *Ann N Y Acad Sci* 2005;1060:450–3.
- Pantev C, Oostenveld R, Engelien A, Ross B, Roberts LE, Hoke M. Increased auditory cortical representation in musicians. *Nature* 1998;392:811–4.
- Pantev C, Roberts LE, Schulz M, Engelien A, Roß B. Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport* 2001;12:169–74.
- Regnault P, Bigand E, Besson M. Different brain mechanisms mediate sensitivity to sensory consonance and harmonic context: evidence from auditory event-related brain potentials. *J Cogn Neurosci* 2001;13:241–55.
- Roß B, Herdman AT, Pantev C. Right hemispheric laterality of human 40 Hz auditory steady-state responses. *Cereb Cortex* 2005;15:2029–39.
- Roß B, Picton TW, Herdman AT, Pantev C. The effect of attention on the auditory steady-state response. *Neuro Clin Neurophysiol* 2004;2004:22.
- Roß B, Picton TW, Pantev C. Temporal integration in the human auditory cortex as represented by the development of the steady-state magnetic field. *Hear Res* 2002;165:68–84.

- Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J Neurosci* 2003;23:5545–52.
- Shahin A, Roberts L, Miller L, McDonald K, Alain C. Sensitivity of EEG and MEG to the N1 and P2 auditory evoked responses modulated by spectral complexity of sounds. *Brain Topogr* 2007;20:55–61.
- Siedenberg R, Goodin DS, Aminoff MJ, Rowley HA, Roberts TPL. Comparison of late components in simultaneously recorded event-related electrical potentials and event-related magnetic fields. *Electroencephalogr Clin Neurophysiol* 1996;99:191–7.
- Strait DL, Kraus N, Parbery-Clark A, Ashley R. Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear Res* 2010;261:22–9.
- Taulu S, Hari R. Removal of magnetoencephalographic artifacts with temporal signal-space separation: demonstration with single-trial auditory-evoked responses. *Hum Brain Mapp* 2009;30:1524–34.
- Taulu S, Simola J, Kajola M. Applications of the signal space separation method. *IEEE Trans Sign Proc* 2005;53:3359–72.
- Tervaniemi M, Brattico E. From sounds to music towards understanding the neurocognition of musical sound perception. *J Conscious Stud* 2004;11:9–27.
- Tervaniemi M, Kruck S, De Baene W, Schröger E, Alter K, Friederici AD. Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *Eur J Neurosci* 2009;30:1636–42.
- Tervaniemi M, Rytönen M, Schröger E, Ilmoniemi RJ, Näätänen R. Superior formation of cortical memory traces for melodic patterns in musicians. *Learn Mem* 2001;8:295.
- Tremblay K, Kraus N, McGee T, Ponton C, Brian O. Central auditory plasticity: changes in the N1–P2 complex after speech-sound training. *Ear Hear* 2001;22:79–90.
- Virtala P, Berg V, Kivioja M, Purhonen J, Salmenkivi M, Paavilainen P, et al. The preattentive processing of major vs. minor chords in the human brain: an event-related potential study. *Neurosci Lett* 2010;487:406–10.
- Wehner DT, Hämäläinen MS, Mody M, Ahlfors SP. Head movements of children in MEG: quantification, effects on source estimation, and compensation. *Neuroimage* 2008;40:541–50.
- Woods DL, Alho K, Algazi A. Intermodal selective attention: evidence for processing in tonotopic auditory fields. *Psychophysiology* 1993;30:287–95.
- Yumoto M, Matsuda M, Itoh K, Uno A, Karino S, Saitoh O, et al. Auditory imagery mismatch negativity elicited in musicians. *Neuroreport* 2005;16:1175–8.
- Yumoto M, Ogata E, Mizuochi T, Karino S, Kuan C, Itoh K, et al. A motor-auditory cross-modal oddball paradigm revealed top-down modulation of auditory perception in a predictive context. *Int Congr Ser* 2007;1300:93–6.

# 述語と項の位置関係ごとの候補比較による日本語述語項構造解析

林部 祐太<sup>†</sup>・小町 守<sup>††</sup>・松本 裕治<sup>†</sup>

一般に、項は述語に近いところにあるという特性がある。そのため、従来の述語項構造解析の研究では、候補を述語との位置関係でグループ分けし、あらかじめ求めておいたグループ間の優先順序に従って正解項を探索してきた。しかしながら、その方法には異なるグループに属する候補同士の比較ができないという問題がある。そこで我々は、異なるグループごとに最尤候補を選出し、それらの中から最終的な出力を決めるモデルを提案する。このモデルは優先度の高いグループに属する候補以外も参照することによって最終的な決定を行うことができ、全体的な最適化が可能である。実験では、提案手法は優先順序に従う解析よりも精度が向上することを確認した。

キーワード：述語項構造解析, 項と述語の位置関係, 探索先行分類型モデル

## Japanese Predicate Argument Structure Analysis by Comparing Candidates in Different Positional Relations between Predicate and Arguments

YUTA HAYASHIBE<sup>†</sup>, MAMORU KOMACHI<sup>††</sup> and YUJI MATSUMOTO<sup>†</sup>

In general, arguments are located near the predicate. A previous study has exploited this characteristic to group candidates by positional relations between a predicate and its candidate arguments and then searched for the final candidate using a pre-determined priority list of the groups. However, in such an analysis, candidates in different groups cannot be compared. Therefore, we propose a Japanese predicate argument structure analysis model that gathers the most likely candidates from all the groups and then selects the final candidate amongst them. We can account for candidates with less priority before making a final decision to perform global optimization. Experimental results show that our model outperforms deterministic models.

**Key Words:** *predicate argument structure analysis, positional relations between predicates and arguments, selection-then-classification model*

---

<sup>†</sup> 奈良先端科学技術大学院大学, Nara Institute of Science and Technology

<sup>††</sup> 首都大学東京, Tokyo Metropolitan University

## 1 はじめに

述語項構造解析の目的は、述語とそれらの項を文の意味的な構成単位として、文章から「誰が何をどうした」という意味的な関係を抽出することである。これは、機械翻訳や自動要約などの自然言語処理の応用において重要なタスクの1つである (Surdeanu, Harabagiu, Williams, and Aarseth 2003; Wu and Fung 2009)。

述語は文の主要部で、他の要素とともに文を構成する (日本語記述文法研究会 2010)。日本語では、述語は品詞によって、形容詞述語・動詞述語・名詞述語の3種類に分けられる。述語が意味をなすためには、補語 (主語を含む) が必要であり、それらは項と呼ばれる。また、述語と項の意味的な関係を表すラベルを格と呼ぶ。

項は前後文脈から推測できるとき省略<sup>1</sup>されることがあり、省略された項をゼロ代名詞、ゼロ代名詞が指示する要素を先行詞と呼ぶ。この言語現象はゼロ照応と呼ばれ、日本語では項の省略がたびたび起きることから、述語項構造解析はゼロ照応解析としても扱われてきた (河原, 黒橋 2004; 笹野, 黒橋 2011)。

本稿では、項と述語の位置関係の種類を次の4種類に分類する。述語と同一文内にあり係り受け関係にある項<sup>2</sup>、(ゼロ代名詞の先行詞として同一文中に存在する) 文内ゼロ、(ゼロ代名詞の先行詞として述語とは異なる文中に存在する) 文間ゼロ、および (文章中には存在しない) 外界項である。本稿では、それぞれ *INTRA\_D*, *INTRA\_Z*, *INTER*, *EXO* と呼ぶ。ある述語がある格にて項を持たないときは、その述語の項は *ARG\_NULL* だとし、その述語と *ARG\_NULL* は *NULL* という位置関係にあるとして考える。本稿では、*EXO* と *NULL* を総称して *NO-ARG* と呼ぶ。例えば、例1において、「受け取った」と「食べた」のヲ格項「コロッケ」はそれぞれ *INTRA\_D*・*INTRA\_Z*、「飲んだ」のガ格項「彼女」は *INTER* で、二格項は *ARG\_NULL* である。

コロッケを受け取った彼女は、急いで食べた。  
( $\phi$ が) ジュースも飲んだ。 (1)

一般に、項は述語に近いところにあるという特性 (近距離特性) を持つ。そのため、これまでの述語項構造解析の研究では、この特性の利用を様々な形で試みてきた。

河原, 黒橋 (2004) や Taira, Fujita, and Nagata (2008) は項候補と述語の係り受け関係の種類ごとに項へのなりやすさの順序を定義し、その順序に従って項の探索を行った。また、Iida, Imui, and Matsumoto (2007) は述語と同一文内の候補を優先的に探索した。これらの先行研究ではあらかじめ定めておいた項の位置関係に基づく順序に従った探索を行い、項らしいものが見つかれば以降の探索はしない。そのため、異なる位置関係にある候補との「どちらがより項

<sup>1</sup> 本稿では、省略を項が述語と直接係り受け関係にないことと定義する。

<sup>2</sup> ここでの関係は向きを持たない。複数の項が同一の述語と関係を持つこともありうる。



らしいか」という相対的な比較は行えず、述語と項候補の情報から「どのくらい項としてふさわしいか」という絶対的な判断を行わなければならないという問題点がある。

そこで、本稿では、項の位置関係ごとに独立に最尤候補を選出した後、それらの中から最尤候補を1つ選出するというモデルを提案する。位置関係ごとに解析モデルを分けることで、柔軟に素性やモデルを設計できるようになる。また、位置関係の優先順序だけでなく、その他の情報（素性）も用いて総合的にどちらがより“項らしい”かが判断できるようになる。

本稿の実験では、まず、全ての候補を参照してから解析するモデルと、特定の候補を優先して探索するモデルを比較して、決定的な解析の良し悪しを分析する。また、陽に項の位置関係ごとの比較を行わないモデルや、優先順序に則った決定的な解析モデルと提案モデルを比較して、ガ格・ヲ格ではより高い性能を達成できたことも示す。

本稿の構成は以下のようになっている。まず2章で述語項構造解析の先行研究での位置関係と項へのなりやすさの優先順序の扱いについて紹介する。3章では提案手法について詳述し、4章では評価実験の設定について述べる。5章・6章では実験結果の分析を行い、7章でまとめを行う。

## 2 関連研究

ここでは、述語項構造解析の先行研究における、位置関係と項へのなりやすさの優先順序の扱いについて紹介する。先行研究と提案手法の概要を表1にまとめた。

### 2.1 決定的な解析を行う方法

#### 2.1.1 優先順序を統計的に求める方法

河原, 黒橋 (2004) は、解析をゼロ代名詞検出と先行詞同定の2段階に分け、統計的に求めた優先順序を先行詞同定の際に用いた。彼らの手法では、まず、格フレーム辞書に基づく格解析

表 1 先行研究と提案手法の概要

	項の位置関係の数	位置関係ごとに項同定モデルを分ける	決定的な解析
河原, 黒橋 (2004)	20		✓
Iida et al. (2007)	2	✓	✓
Taira et al. (2008)	7	✓	✓
笹野, 黒橋 (2011)	—		✓
Imamura, Saito, and Izumi (2009)	—		
吉川, 浅原, 松本 (2013) <sup>3</sup>	—		
提案手法	3	✓	

<sup>3</sup> 文間項は探索しない。

によって、ゼロ代名詞の検出を行う。そして、項が存在すると判断された場合は、あらかじめ求めておいた優先順序に従って候補を探索し、候補と格フレーム用例の類似度が閾値以上かつ分類器でも正例と分類される候補を先行詞として同定する。分類器は項の位置関係に関わらず、共通のものを作成した。素性には、格フレームとの類似度や品詞などを用いた。

彼らは、従属節、主節、埋め込み文などといった文・文章中の構造をもとに、項の位置関係(彼らは「位置カテゴリ」と呼んだ)を20種類に分類した。

彼らは、位置カテゴリごとに、先行詞の取りやすさを

$$\frac{\text{先行詞がその位置カテゴリにある回数}}{\text{その位置カテゴリにある先行詞候補の数}} \quad (1)$$

でスコア化した。そして、位置カテゴリごとに、京都大学テキストコーパス (Kawahara, Kurohashi, and Hasida 2002) からスコアを算出し、得られたスコアを降順にソートしてそれぞれの格について優先順序を得た。

### 2.1.2 文内候補を優先的に探索する方法

Iida et al. (2007) は、先行詞候補とゼロ代名詞の統語的關係をパターン化するために、木を分類するブースティングアルゴリズム BACT (工藤, 松本 2004) を用いた。BACT は木構造データを入力とし、全ての部分木の中から分類に寄与する部分木に対して大きな重みをつける。彼らは、先行詞候補とゼロ代名詞間の係り受け木や、關係を表す素性を、根ノードに子としてつなげて BACT の入力とした。

文間先行詞の同定には係り受け關係を利用できないため、彼らは先行詞の同定モデルを文内と文間に分け、文内候補を優先的に探索する以下の方法をとった。

- (1) 最尤先行詞同定モデル  $M_{10}$  で、文内最尤先行詞  $C_1^*$  を求める
- (2) 照応性判定モデル  $M_{11}$  で、 $C_1^*$  の先行詞らしさのスコア  $p_1$  を求める。あらかじめ定めておいた閾値  $\theta_{\text{intra}}$  に対して、 $p_1 \geq \theta_{\text{intra}}$  であれば、 $C_1^*$  を先行詞として決定する。そうでなければ (3) に進む。
- (3) 最尤先行詞同定モデル  $M_{20}$  で、文間最尤先行詞  $C_2^*$  を求める
- (4) 照応性判定モデル  $M_{21}$  で、 $C_2^*$  の先行詞らしさのスコア  $p_2$  を求める。あらかじめ定めておいた閾値  $\theta_{\text{inter}}$  に対して、 $p_2 \geq \theta_{\text{inter}}$  であれば、 $C_2^*$  を先行詞として決定する。そうでなければ、先行詞なしとする。

$M_{10} \dots M_{21}$  はそれぞれ BACT を使って学習・分類し、パラメータ  $\theta_{\text{intra}}$  と  $\theta_{\text{inter}}$  は、開発データを用いて最適なものを求める。この手法では、文内の最尤先行詞同定や照応性判定には文間の候補の情報は参照せずに、決定的に解析している。

### 2.1.3 優先順位を経験的に決める方法

Taira et al. (2008) は、決定リストを用いて全ての格の解析を同時に行う方法を提案した。決定リストは規則の集合に適用順位を付けたものであり、機械学習の結果を人が分析しやすいという特長がある。彼らは項の位置関係やヴォイス・機能語に加えて、単語の出現形・日本語語彙大系(池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林 1997) から得られる意味カテゴリ・品詞のいずれか1つを加えたものを組として扱い、それぞれの組を1つの素性とした。そして、述語ごとに Support Vector Machine の学習で素性の重みを得て、素性を重みでソートしたものを決定リストとした。すなわち、1つの素性を1つの決定リストのルールとして扱った。

彼らは項の単位を単語とし、項の位置関係を係り受け関係に基づいて次の7種類に定義している。なお、fw と bw は追加的な種類で、その他の種類と兼ねることができる。

- Incoming Connection Type (ic): 項を含む文節が述語を含む文節に係っている  
日米交渉<sub>ガ:進展</sub>が進展した
- Outgoing Connection Type (oc): 述語を含む文節が項を含む文節に係っている  
衝動<sub>ガ:買い</sub>買った 新刊本<sub>ガ:買い</sub>
- Within the Same Phrase Type (sc): 項が述語と同じ文節内にある  
日米交渉<sub>ガ:日</sub>が
- Connection into Other Case role Types (ga\_c, wo\_c, ni\_c): 項を含む文節が述語を含む文節に、他の格の項を介して係っている  
トム<sub>ヲ:説得,ga\_c</sub>への友人<sub>ガ:説得</sub>による説得
- Non-connection Type (nc): 項が述語とは異なる文にある
- Forward Type (fw): 文章内にて、項が述語の前方にある
- Backward Type (bw): 文章内にて、項が述語の後方にある

実際の解析は、各述語について次の手順で行った。

- (1) ic, oc, ga\_c, wo\_c, ni\_c について、決定リストを用いて項を決定する
- (2) (1) で決まらなかった格について、sc の決定リストを用いて項を決定する
- (3) 対象の述語が項を持つ確率が50%以上であれば(4)に進む
- (4) nc, fw, bw に関する決定リストを用いて項を決定する

この手法は経験的に、優先順序を

ic, oc, ga\_c, wo\_c, ni\_c > sc >> nc, fw, bw

のように定めたといえる。ic, oc, ga\_c, wo\_c, ni\_c 間での、探索の優先関係はない。

この方法は、格と項の位置関係を考慮しつつ、項になりやすいものから決めていくのが特徴である。ただし、着目している候補と述語の情報のみを用いて項らしいかどうかを判断していくため、必ずしも全ての候補を参照してから最終的な出力を決定するわけではなく、候補間でどれが項らしいかの相対的な判断は行われたい。

#### 2.1.4 述語と係り受け関係にある候補を優先的に項であるとみなす方法

笹野, 黒橋 (2011) は, 解析対象述語の格フレーム候補それぞれに対して, 次の手順で格フレームと談話要素の対応付け候補を生成した.

- 解析対象述語と直接係り受け関係にある談話要素を, 選ばれた格フレームの格スロットと対応付ける. 談話要素が係助詞をともなって出現した場合や, 被連体修飾節に出現した場合など, 複数の格スロットとの対応付けが考えられる場合は, 考えうるすべての対応付けを生成する.
- 上記の処理で生成された対応付け候補に対し, 対応付けられなかったガ格・ヲ格・ニ格と, 解析対象述語と係り受け関係にない談話要素の対応付けを行う.

そして, 対数線形モデルにて最も確率的評価が高い対応付けを解析結果として出力した. 素性には, 意味クラスや固有表現情報の他に, 出現格と出現位置に関する 85 個の 2 値素性も用いた.

この手法では, 格ごとに独立に解析を行なっているのではなく, 同時に解析を行なう. しかし, 述語と係り受け関係にある候補を優先的に項であるとみなすため, 係り受け関係にある候補と, 係り受け関係にない候補または他の文にある候補との比較は行えない.

## 2.2 優先順序を素性として表現する方法

位置関係と項へのなりやすさの関係を優先順序として利用し決定的な解析を行うのではなく, 素性として利用した研究もある.

### 2.2.1 最大エントロピー法を用いる方法

Imamura et al. (2009) は, 最大エントロピー法に基づく識別的モデルを用いた. 彼らは, 位置関係ごとにモデルを分けるのではなく, 素性として, 述語と候補の位置関係, 係り受け関係を用いた. そして候補集合に, 項を持たないことを示す特別な名詞句 NULL を加え, その中から最尤候補を同定するというモデル化を行った. なお, 候補数削減のため, 文間項候補は述語を含む文の直前の文に出現したものと, これまでの解析ですでに項として同定されたものに限定している.

この方法では格ごとにモデルは1つだけ学習すればよい. ただし, この手法では, 候補間の関係を素性として用いることはできない.

### 2.2.2 Markov Logic を用いる方法

吉川 他 (2013) は, Markov Logic を利用して, 文内の複数の述語の項構造解析を同時に行う手法を提案した. Markov Logic は一階述語論理と Markov Networks を組み合わせたもので, 一階述語論理式の矛盾をある程度のペナルティの上で認めることができる統計的關係学習の枠組

みである。彼らは項同定・項候補削減・格ラベル付与を同時に行うモデルを提案した。

彼らは、文間の項候補を加えるのは計算量の問題から困難だとしている。素性（観測述語）は述語と候補の係り受け関係などを用いた。

### 3 述語と項の位置関係ごとの候補比較による日本語述語項構造解析

先行研究では、優先順位の低い位置関係にある候補は参照されずに、解析が行われていた。この方法は、優先順位の高い位置関係にある項の同定の性能は上げることができるが、優先順位の低い位置関係にある候補の再現率は下げてしまうという問題点がある。また、優先順位の低い位置関係にある候補も参照してから最終的な決定を行った方が、全体的な解析性能が向上すると考える。

そこで我々は、探索とトーナメントの2つのフェーズからなる、位置関係ごとに最尤候補を求めてから最終的な出力を決めるモデルを提案する。これは、「探索」・「分類」という2つのフェーズを持つ探索先行分類型モデル (Iida, Inui, and Matsumoto 2005) に着想を得て、後半の分類フェーズをトーナメント式に置き換えたものである。なお、このモデルは格ごとに解析器を学習・使用する。

#### 3.1 項構造解析における探索先行トーナメントモデル

##### 3.1.1 探索

はじめのフェーズでは任意の項同定モデルを用いて INTRA\_D, INTRA\_Z, INTER の最尤候補を選出する。それぞれ異なる素性やモデルを用いてもよい。モデルには、述語と探索対象の候補を入力として与え、探索対象の候補の中の1つを出力させる。

##### 3.1.2 トーナメント

次のフェーズでは探索フェーズで得られた3つの最尤候補を入力とし、そのうちの1つか“NO-ARG”を出力する。これにより、最尤候補のうちどれが正解項であるか、もしくは項を持たないかを判断する。

このフェーズは図1に示したように (a) から (c) の3つの2値分類モデルで構成される。なお、予備実験にて異なる順序を試したが、文内最尤候補同士を (a) にて直接比較できるこの順序の性能が最も高かった。

- (a) INTRA\_D と INTRA\_Z を比較して、よりその述語の項らしい方を選ぶ
- (b) INTER と (a) で選出された候補を比較して、よりその述語の項らしい方を選ぶ
- (c) (b) で選出された候補と “NO-ARG” を比較して、よりその述語の項らしい方を選ぶ

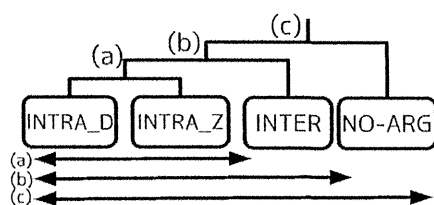


図 1 トーナメントフェーズでの、位置関係が異なる候補からの項の同定

(a) から (c) の分類器の学習事例には、Algorithm 1 で示すように探索フェーズで得られた最尤候補を用いる。

### 3.2 提案手法の関連研究

提案手法は 2 つのモデルを参考にしている。

1 つ目は名詞句の照応解析における探索先行分類型モデル (*selection-then-classification model*) (Iida et al. 2005) である。このモデルは最初に、最尤先行詞を求める (彼らはこれを “探索” と呼んだ)。次に、その最尤先行詞を用いて、名詞句が実際に照応詞であるかどうかを判定する (彼らはこれを “分類” と呼んだ)。このモデルの利点は、照応性を持たない名詞句も学習事例の生成に使えることである。彼らは実験で、最尤先行詞を用いて照応性判定を行ったほうが、最尤先行詞を用いない場合よりも高い性能が出ることを確かめた。提案手法も、位置関係ごとに最尤候補を求めた後、どの候補が実際に項であるのかを判定する。最尤候補の探索を先に行なうことで、位置関係ごとの最尤候補を学習事例の生成に用いることができる。

2 つ目はゼロ照応解析におけるトーナメントモデル (飯田, 乾, 松本 2004) である。そのモデルは、全ての先行詞候補 (実際には先行する全ての名詞句) のペアに対して、どちらがより先行詞らしいかの 2 値分類を繰り返す。トーナメントモデルの利点は候補間の関係性の素性を使うことができる点である。提案手法のトーナメントフェーズでも同様に、トーナメントモデルを用いて、位置関係ごとに選出された最尤候補のペアからどちらが正解項らしいかの 2 値分類を繰り返し、候補間の比較を行うことができる。

## 4 評価実験

### 4.1 実験データセット

評価実験には NAIST テキストコーパス 1.4β (飯田, 小町, 井之上, 乾, 松本 2010) を用いた。これは京都大学テキストコーパス 3.0<sup>4</sup> を基にしており、述語項構造, 事態性名詞の項構造,

<sup>4</sup> <http://nlp.ist.i.kyoto-u.ac.jp/nl-resource/corpus/KyotoCorpus3.0.tar.gz>

---

**Algorithm 1** 分類器 (a) classifier\_a, (b) classifier\_b, (c) classifier\_c の学習事例の作成アルゴリズム

---

```

procedure TRAIN(predicate, gold_argument, candidates)
  gold_argument_type ← getArgumentType(predicate, gold_argument)
  ▷ 正解項の位置関係を取得する

  ▷ 位置関係ごとに最尤候補を取得する
  most_likely_candidate_INTRA_D ← getMostLikelyCandidate(predicate, candidates, INTRA_D)
  most_likely_candidate_INTRA_Z ← getMostLikelyCandidate(predicate, candidates, INTRA_Z)
  most_likely_candidate_INTER ← getMostLikelyCandidate(predicate, candidates, INTER)

  if gold_argument_type = NO_ARG then
    MakeExample(classifier_c, NO_ARG, predicate, most_likely_candidate_INTRA_D)
    MakeExample(classifier_c, NO_ARG, predicate, most_likely_candidate_INTRA_Z)
    MakeExample(classifier_c, NO_ARG, predicate, most_likely_candidate_INTER)
  return
end if

  MakeExample(classifier_c, HAVE_ARG, predicate, gold_argument)
  if gold_argument_type = INTRA_D then
    MakeExample(classifier_a, INTRA_D, predicate, gold_argument,
      most_likely_candidate_INTRA_Z)
    MakeExample(classifier_b, INTRA, predicate, gold_argument, most_likely_candidate_INTER)
  else if gold_argument_type = INTRA_Z then
    MakeExample(classifier_a, INTRA_Z, predicate, gold_argument,
      most_likely_candidate_INTRA_D)
    MakeExample(classifier_b, INTRA, predicate, gold_argument, most_likely_candidate_INTER)
  else if gold_argument_type = INTER then
    MakeExample(classifier_b, INTER, predicate, gold_argument, most_likely_candidate_INTRA_D)
    MakeExample(classifier_b, INTER, predicate, gold_argument, most_likely_candidate_INTRA_Z)
  end if
  return
end procedure

procedure MAKEEXAMPLE(classifier, label, predicate, candidatel, candidate2)
  ▷ candidate2 は省略できる
  項候補 candidatel, candidate2 が照応関係にあれば事例は作成しない.
  述語 predicate と項候補 candidatel, candidate2 に対して, 素性集合  $F$  を取得する.
  学習器 classifier に対して,  $F$  を用いて, label をラベルとする学習事例を 1 つ作成する.
end procedure

```

---

共参照に関する情報が約 40,000 文の新聞記事にわたって付与されている。なお、アノテーションの誤りのため 6 記事<sup>5</sup>を除外した。このコーパスの記事を表 2 で示すように学習・開発（パラメータチューニング）・評価のために 3 分割した。これは、Taira et al. (2008) や吉川 他 (2013) と同じ分割方法である。表 3 に項の分布の統計情報を示す<sup>6</sup>。

## 4.2 実験設定

実験では、MeCab0.996・IPADIC-2.7.0-20070801 で解析して得られた形態素情報、京都大学テキストコーパス 3.0 で付与されている文節情報、CaboCha0.66 で解析して得られた係り受け関係を用いた。項の候補は文節単位で抽出した。解析は文頭から文末の順で行い、述語を含む文以降からは項候補を抽出しない。なお、ある述語の格についての解析結果は同じ述語の他の格についての解析に影響を及ぼさない。

本稿では項同定に焦点を絞るため、述語同定タスクには取り組まない。言い換えると、どれが述語であるかはあらかじめシステムに与えておく。述語には軽動詞「する」や複合動詞も含む。

最尤候補同定には、トーナメントモデル (飯田 他 2004) を用いた。その際、最尤候補の探索範囲ごとに異なるモデルを作成し、モデルの学習方法も (飯田 他 2004) に従った。例えば、提案手法は探索フェーズでは INTRA\_D, INTRA\_Z, INTER の最尤候補を同定するが、それぞれ

表 2 NAIST テキストコーパスの統計情報

	ニュース記事の範囲	社説記事の範囲	記事数	文数	述語数
学習	1月1日～11日	1月～8月	1,673	23,150	64,173
開発	1月12日～13日	9月	458	4,634	13,105
評価	1月14日～17日	10月～12月	662	8,795	24,296

表 3 NAIST テキストコーパスにおける項の分布

	格	SAME_BS	INTRA_D	INTRA_Z	INTER	EXO	NULL
学習	ガ格	128	33,775	12,057	7,436	10,529	248
	ヲ格	58	22,869	2,051	806	60	38,329
	ニ格	527	10,427	555	278	25	52,361
開発	ガ格	39	6,954	2,700	1,819	1,501	92
	ヲ格	10	4,724	445	147	18	7,761
	ニ格	51	2,673	212	80	5	10,084
評価	ガ格	40	12,805	4,829	3,241	3,140	241
	ヲ格	18	8,665	829	291	14	14,479
	ニ格	132	5,023	358	155	12	18,616

<sup>5</sup> 除外した文書 ID : 951230038, 951225057, 950106156, 950106034, 951221047, 950106211

<sup>6</sup> SAME\_BS は項と述語が同一文節であることを示す。



異なる合計3つの解析モデルを最尤候補同定に用いる。

### 4.3 分類器と素性

探索フェーズ・トーナメントフェーズで用いる分類器には, Support Vector Machine (Cortes and Vapnik 1995) を線形カーネルで用いた。具体的には LIBLINEAR1.93<sup>7</sup>の実装を用い, 開発データを用いたパラメータチューニングを行った。

素性には Imamura et al. (2009) で用いられたものとはほぼ同一の素性を用いた。

- 述語・項候補の主辞・機能語・その他の語の出現形・形態素情報
- 述語が受け身の助動詞を含むときはその原形
- 係り受け木上の述語と項候補の関係<sup>8</sup>

係り受け木上の項候補ノード  $N_a$  と述語ノード  $N_p$  からそれぞれ ROOT 方向に辿っていくときに初めて交叉するノードを  $N_c$  とし,  $N_a$  から  $N_c$  までの道のりに含むノード列を  $A_{a\dots c}$ ,  $N_p$  から  $N_c$  までの道のりに含むノード列を  $A_{p\dots c}$  とする。また,  $N_c$  から木の ROOT までの道のりに含むノード列を  $A_{c_1, c_2, \dots, c_r}$  とする。本実験では, ノード列の文字列表現として,

- 主辞の原形
- 主辞の品詞
- 機能語の原形
- 機能語の品詞
- 機能語の原形 + 機能語の品詞

の5通りを用いた。  $A_{a\dots c}$  の文字列表現を  $S_{a\dots c}$ ,  $A_{p\dots c}$  の文字列表現を  $S_{p\dots c}$  とし, それらの連結を  $S_{a\dots c} + S_{p\dots c}$  とする。

素性には,  $S_{a\dots c} + S_{p\dots c}$ ,  $S_{a\dots c} + S_{p\dots c} + S_{c_1}$ ,  $S_{a\dots c} + S_{p\dots c} + S_{c_1, c_2}$ ,  $\dots$ ,  $S_{a\dots c} + S_{p\dots c} + S_{c_1, c_2, \dots, c_r}$  の  $r+1$  個の文字列を用いた。つまり, 述語と項候補の関係を  $5(r+1)$  個の文字列で表現した。

- 係り受け木上の2つの項候補の関係  
上と同様の素性表現を行った。
- 述語と項候補・2つの項候補間の距離 (文節単位・文単位ともに)
- 「述語・項候補の主辞・助詞」のコーパス中の共起スコア<sup>9</sup>

動詞と項の共起のモデル化は (藤田, 乾, 松本 2004) に従った。名詞  $n$  が格助詞  $c$  を介し

<sup>7</sup> <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>

<sup>8</sup> Imamura et al. (2009) では, どのような素性表現に落とし込んだかは詳述されていない。

<sup>9</sup> Imamura et al. (2009) では, これに相当するものとして, Good Turing スムージングを施した共起確率を用いている。計算は NAIST テキストコーパス相当部分を除いた 1991~2002 年の毎日新聞を用いた。

て動詞  $v$  に係っているときの共起確率  $P(\langle v, c, n \rangle)$  を推定するため、 $\langle v, c, n \rangle$  を  $\langle v, c \rangle$  と  $n$  の共起とみなす。

共起尺度には自己相互情報量 (Hindle 1990) を用いた。

$$PMI(\langle v, c \rangle, n) = \log \frac{P(\langle v, c, n \rangle)}{P(\langle v, c \rangle)P(n)}$$

なお、スムージングは行わなかった。自己相互情報量の算出には次の2つのコーパスを用い、2つの値をそれぞれ二値素性として<sup>10</sup>用いた。

**NEWS:** 1995年を除く1991年から2003年までの毎日新聞約1,800万文。MeCab0.98<sup>11</sup>で形態素解析を行いCaboCha0.60pre4<sup>12</sup>で係り受け解析を行った。辞書はNAIST Japanese Dictionary 0.6.3<sup>13</sup>を用いた。約2,700万対の〈動詞, 格助詞, 名詞〉の組を抽出した<sup>14</sup>。

**WEB:** Kawahara and Kurohashi (2006) がウェブから収集した日本語約5億文。JUMAN<sup>15</sup>で形態素解析を行い、KNP<sup>16</sup>で係り受け解析を行なっている。KNPの項構造解析結果から約53億対の〈述語, 格助詞, 項〉の組を抽出した<sup>17</sup>。

- 項候補が以前の項構造解析で項となったか否かを示す2値情報
- 項候補の主辞のSalient Reference List (Nariyama 2002) における順位

## 4.4 比較対象

先行研究では、我々のものと異なる素性や機械学習の手法を使っており実験設定が異なる。そのため、ベースラインモデルとしてIIDA2005、比較対象モデルとしてIIDA2007・IIDA2007+・PPR-を実装し、位置関係ごとに最尤候補を求めてから最終的な出力を決める提案モデルPPR (Preferences based on Positional Relations) と比較する。

### 4.4.1 IIDA2005

位置関係に関わらずに、全ての候補の中から最尤の候補を探索フェーズで1つ選出した後、トーナメントフェーズでそれが項としてふさわしいか否かを判断するモデル。(Iida et al. 2005) の探索先行分類型モデルである。

全ての候補の中から1つを選ぶという点で (Imamura et al. 2009) とほぼ同等のモデルである。

<sup>10</sup> 値が  $x$  以下のときのみ発火する素性。実際には、 $x$  を  $-4$  から  $4$  まで  $0.1$  刻みで変化させた素性を用いた。

<sup>11</sup> <https://code.google.com/p/mecab/>

<sup>12</sup> <https://code.google.com/p/cabocha/>

<sup>13</sup> <http://sourceforge.jp/projects/naist-jdic/>

<sup>14</sup> 動詞が約3万種、名詞が約32万種で、ユニーク数は約700万組。

<sup>15</sup> <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

<sup>16</sup> <http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

<sup>17</sup> 動詞が約8億種、項が約2.8億種で、ユニーク数は約1.6億組。

彼らのモデルと異なる主な点は、最尤候補同定と照応性判定を異なるモデルで行う点と、最尤候補同定時に2候補間の関係性も素性として用いる点である。

このベースラインモデルとその他のモデルと比較することで、項の位置関係によって探索の優先順序をつけることの効果や、位置関係ごとに最尤候補同定モデルを作り最尤候補同士の比較を陽に行う効果を調べる。

#### 4.4.2 IIDA2007

文内最尤候補を選出した後、分類器が項としてふさわしいと判断すればそれを項として出力し、そうでなければ同様に文間候補の探索を行うモデル。2.1.2節で述べた (Iida et al. 2007) の文内候補を優先的に探索するモデルである。

彼らのモデルと異なる主な点は、最尤候補同定や候補の適格性判定を行う分類器に BACT ではなく SVM を用いる点である。

IIDA2005 と比較することで、文内候補を優先的に探索することの効果調べる。

#### 4.4.3 IIDA2007+

INTRA<sub>D</sub> の探索後、最尤候補が項としてふさわしいかどうかの判断 (適格性判定) を行う。適格であればそれを出力し終了する。非適格であれば INTRA<sub>Z</sub> の探索を行い、同様に適格性判定を行う。それも非適格であれば INTER の探索を行い、適格であればそれを出力し、非適格であれば項は無いと判断する。IIDA2005 と IIDA2007 の自然な拡張で、述語から統語的な距離の近いものを優先的に探索する。

IIDA2007 と比較することで、文内候補を細かく INTRA<sub>D</sub> と INTRA<sub>Z</sub> に分けて優先順序をつけることの効果調べる。

#### 4.4.4 PPR-

このモデルは、提案モデルとほぼ同じモデルであるが、INTRA<sub>D</sub> と INTRA<sub>Z</sub> を区別せずに、位置関係が INTRA と INTER の2グループであると仮定する。図1の (b) と (c) で示すようにトーナメントフェーズは2つの2値分類モデルからなる。分類器 (c) は INTRA と INTER の候補のどちらが最尤候補であるかを判断する。

PPR と比較することで、文内の項の位置関係を細かく INTRA<sub>D</sub> と INTRA<sub>Z</sub> に分けて最尤候補同定モデルを作り、最尤候補同士の比較を行うことの効果調べる。

#### 4.4.5 比較対象とする先行研究

NAIST テキストコーパスを使い、全ての項の位置関係で実験を行なっている (Taira et al. 2008) と (Imamura et al. 2009) との比較も行う。ただし、本実験とは微妙に実験設定が異なるため、

厳密な比較はできないことに注意してほしい。

Taira et al. (2008) の実験では 19,501 個の述語をテストに, 49,527 個を学習に, 11,023 個を開発に使っている。また学習では京都大学テキストコーパス 4.0 で付与されている係り受け情報と形態素情報を用いているが, テストでは独自の係り受け解析器を用いている。

Imamura et al. (2009) の実験では, 25,500 個の述語をテストに, 67,145 個を学習に, 13,594 個を開発に使っている。我々は京都大学テキストコーパス 3.0 を用いたが, Imamura et al. (2009) は京都大学テキストコーパス 4.0 で付与されている係り受け情報と形態素情報を学習とテストに用いている。

#### 4.4.6 その他の先行研究

笹野, 黒橋 (2011) は, 提案システムは表層格の解析を行うことから, 受け身・使役形である述語は評価から除外しており, 本稿では比較対象としない。

吉川 他 (2013) は, 文間項は解析対象としていないため, 本稿では比較対象としない。

(渡邊, 浅原, 松本 2010) は述語語義と項の意味役割の依存関係を考慮しながら, 双方を同時に学習, 解析を行う構造予測モデルを提案している。しかし, 本稿とは異なるデータセットを用いていることから, 比較対象とはしない。

## 4.5 評価尺度

Precision, Recall, F 値で位置関係ごとに評価を行う。

システムが出力した位置関係が  $T$  であるもののうち, 正しく同定できているものの数を  $tp_{(T)}$ , できていないものの数を  $fp_{(T)}$ , システムに同定されなかった項のうち位置関係が  $T$  であるものの数を  $fn_{(T)}$  とすると,

$$Precision = \frac{tp_{(T)}}{tp_{(T)} + fp_{(T)}}, \quad Recall = \frac{tp_{(T)}}{tp_{(T)} + fn_{(T)}}, \quad F = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

と定義できる。

また, システム全体 (ALL) の  $tp, fp, fn$  と Precision, Recall, F 値も, 同様に定義できる。

## 5 議論

表 4, 5, 6 にガ格・ヲ格・ニ格の実験結果を示す。  $P, R, F, A_M$  はそれぞれ Precision, Recall, F 値, F 値のマクロ平均 (INTRA\_D, INTRA\_Z, INTER の F 値の算術平均) を示す。

ALL の F 値に関して, PPR- と PPR が IIDA2007 と比較して有意差があるかどうかの検定を