Between Reading Time and Clause Boundaries in Japanese – Wrap-up Effect in a Head-Final Language

Masayuki Asahara
National Institute for Japanese Language and Linguistics, Japan
Tokyo, Japan
masayu-a@ninjal.ac.jp

Abstract

This paper presents a contrastive analysis between reading time and clause boundary categories in the Japanese language. We overlaid reading time data, made with BCCWJ EyeTrack, and clause boundary categories annotation on the Balanced Corpus of Contemporary Written Japanese. Statistical analysis based on the Bayesian linear mixed model shows that the reading time behaviours differ among the clause boundary categories. The result does not support the wrap-up effects of clause-final words. Another result we arrived at is that the predicate-argument relations facilitate the reading speed of native Japanese speakers.

1 Introduction

Readability is an important issue for evaluation of sentence generation by both humans and machines. In general, readability should be evaluated by the reading time and the comprehension question for the stimuli. The reading time analysis studies evaluated the wrap-up effect (Just and Carpenter, 1980; Rayner et al., 2000) in English, in which readers tend to spend a longer time while reading clause-end phrases than clause-internal phrases. Our motivation is to evaluate whether the wrap-up effect at clause-end phrases occurs in Japanese or not.

This paper explains how clause boundaries affect Japanese reading time using exploratory data analysis. We use reading time data from BCCWJ-EyeTrack (Asahara et al., 2016), which consists of the 'Balanced Corpus of Contemporary Written Japanese' (hereafter BCCWJ) (Maekawa et al., 2014) as the stimuli for the self-paced reading and eye tracking method. We overlaid the clause boundary category annotation data BCCWJ-ToriClause (Matsumoto et al., 2018) on the reading time data.

Statistical analysis using the Bayesian linear mixed model (Sorensen et al., 2016) shows that the clause wrap-up effect (Just and Carpenter, 1980; Rayner et al., 2000) does not appear in our experiment for Japanese.

The analysis also shows the difference among the clause boundary categories. For example, the relative clause end phrases require a shorter reading time than the apposition clause end phrases; the noun clause end phrases have a shorter reading time than the quotation clause end phrases; and there are reading time differences between causal clause end and attendant circumstance clause end.

Section 2 presents related work. Section 3 describes the data. Section 4 shows statistical analysis method. Section 5 presents results with discussions. Finally, Section 6 presents our concluding remarks.

2 Related Work

First, we present models of language analysis involving reading time or eye-tracking gaze information. (Barrett et al., 2016) presented a POS tagging model with gaze patterns. (Klerke et al., 2015) presented a grammaticality detection model for machine-processed sentences.

Second, we present related work on eye-tracking corpora. The Dundee Eyetracking Corpus (Kennedy and Pynte, 2005) contains reading times for English and French newspaper editorials collected from 10 native speakers of each language, measured using

32nd Pacific Asia Conference on Language, Information and Computation
Hong Kong, 1-3 December 2018

eye-tracking equipment. (Frank et al., 2013) developed UCL corpus which includes isolated sentences with eye-tracking data. (Futrell et al., 2018) developed Natural Stories Corpus which comprises self-paced reading data.

Eye-tracking corpora for other languages are also available, including the Potsdam Sentence Corpus (Kliegl et al., 2006), the Potsdam-Allahabad Hindi Eyetracking corpus (Husain et al., 2015) and the Beijing Sentence Corpus of Mandarin Chinese (Yan et al., 2010).

The corpus does not target a specific set of linguistic phenomena, but instead provides naturally occurring texts to test diverse hypotheses. BCCWJ-EyeTrack (Asahara et al., 2016) published reading time data for a subset of the core data of the BCCWJ (Maekawa et al., 2014), which consisted of newspaper article (PN: published newspaper) samples. The original BCCWJ-EyeTrack includes syntactic dependency information (Asahara and Matsumoto, 2016).

Third, the eye-tracking corpus-based psycholinguistic research is conducted using contrastive statistical analysis with annotations. (Asahara and Kato, 2017) overlaid the annotation of categories in a thesaurus ‘Word List by Semantic Principles’ on BCCWJ-EyeTrack, and explained the relation between reading time and syntactic/semantic categories. (Asahara, 2017) overlaid the annotation of information structure (Miyauchi et al., 2017) and explained the relation between reading time and information structure in discourse.

Fourth, there are several preceding work to evaluate wrap-up effect other than the aforementioned ones. (Hill and Murray, 2000) evaluated the reading speed around punctuations including prepositional phrases. (Hirotani et al., 2006) also evaluated clause and sentence wrap-up with punctuation and intonation. (Warren et al., 2009) evaluated intra- and inter-clause integration by eye-tracking. However, these evaluations are only for English.

3 Data

First, we explain BCCWJ-EyeTrack. Second, we describe the clause boundary annotations of BCCWJ-Toribank. We used these two as data, as given in Table 1.

Table 1: Overview of the data.

Reading Time Related (BCCWJ-EyeTrack)		
Column Name	Type	Description
surface	factor	Word surface form
time	int	Reading-time
logtime	num	Reading-time (log)
measure	factor	Reading time type
sample	factor	Sample Name
article	factor	Article information
metadata_orig	factor	Document structure tag
metadata	factor	Metadata
space	factor	Boundary with space or not
length	int	Number of characters
is_first	factor	The leftmost segment
is_last	factor	The rightmost segment
is_second_last	factor	The second rightmost segment
sessionN	int	Session order
articleN	int	Article display order
screenN	int	Screen display order
lineN	int	Line display order
segmentN	int	Segment display order
subj	factor	Participant ID
setorder	factor	Segmentation type order
dependent	int	Number of dependents
Clause Boundary Related (BCCWJ-Toribank)		
Column Name	Type	Description
HS	boolean	Nominal phrase (top level)
MS	boolean	Adnomial phrase (top level)
FU	boolean	Advervial phrase (top level)
HR	boolean	Coordinate phrase (top level)
HS*	factor	Nominal phrase (second level)
MS*	factor	Adnomial phrase (second level)
FU*	factor	Advervial phrase (second level)
HR*	factor	Coordinate phrase (second level)

3.1 BCCWJ-EyeTrack

Here, we briefly present the BCCWJ-EyeTrack data. Their specifications are described in detail in (Asahara et al., 2016). There are two types of reading time data for the newspaper texts. The first is non-cumulative self-paced moving-window presentation data (SELF) as gathered by the Linger program.[1] This type of data has been found to yield the best correlation with eye-tracking data when different styles of presentation were compared for English (Just et al., 1982). The other type of data is eye-tracking data recorded with a tower-mounted EyeLink 1000 (SR Research Ltd). We explain the two measurement methods for the estimation of reading time: eye-tracking and self-paced reading. The order of tasks was fixed using the eye-tracking method in the first session and the self-paced reading method in the second session. Each participant saw each text

[1] http://tedlab.mit.edu/~dr/Linger/

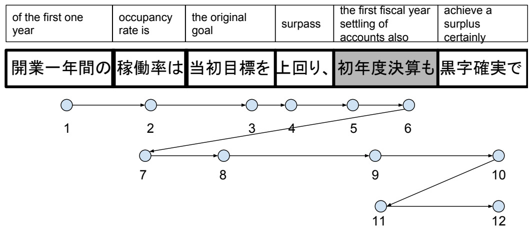
once with the task, and segmentation of the texts was counterbalanced across participants.

Eye-tracking was recorded with a tower-mounted EyeLink 1000 (SR Research Ltd). The view was binocular, but data were collected from each participant's right eye at a resolution of 1000 Hz. Participants looked at the display with the help of a half-mirror; their heads were fixed, with their chins on a chin rest. Unlike self-paced reading, during eye-tracking, all segments were shown simultaneously. This allowed more natural reading, because the participants could freely return and reread earlier parts of the text on the same screen. However, the participants were not allowed to return to previously viewed screens.

For these eye-tracking data, five types of measurements were used: first fixation time (FFT), first-pass time (FPT), regression path time (RPT), second-pass time (SPT), and total time (TOTAL).

FFT is the duration of the first fixation after the gaze first enters the area of interest. FPT is the total duration of fixation from the moment the gaze first stops within the area of interest until it leaves the focus area by moving to the right or left of this area. RPT is the total span from the moment the gaze enters the area of interest until it crosses the right boundary of this area for the first time. SPT is the total span of time the gaze spends in the area of interest, excluding the FPT. TOTAL is the total duration that the gaze spends within the area of interest. Figure 1 illustrates the measurements.

Table 1 above presents the data. `surface` is the surface form of the word. The reading time (i.e., `time`) is converted into a log scale (i.e., `logtime`). `measure` is the reading type {SELF, FFT, FPT, RPT, SPT, TOTAL}. `sample, article, metadata_orig, metadata` represent information related to the article. `space` denotes whether spaces are present between segments. `length` is the number of characters in the surface form. `is_first,is_last,is_second_first` represent the layout features on the screen. `sessionN, articleN, screenN, lineN, segmentN` represent the display order of the elements. `subj` is the participant ID, which is used as a random effect for the statistical analysis. `setorder` is the set presentation order. `dependent` shows the number of dependents for



FFT for 'the first fiscal year settling of accounts also' (hereafter 'the area of interest') is the duration of fixation 5.
FPT is the sum of the durations of fixations 5 and 6.
RPT can includes fixations to the left of the left boundary (e.g., 7 and 8) and durations of fixations when the gaze returns to the area of interest (e.g., 9).
SPT is the sum of the durations of fixations for 9 and 11.
TOTAL is the sum of the durations of fixations 5, 6, 9 and 11.

Figure 1: Five types of measurements for eye-tracking data.

the segments. The dependency relation is annotated by humans (Asahara and Matsumoto, 2016).

3.2 Clause Boundaries Annotation

The clause boundary categories are based on the Tori-Bank schema (Ikehara, 2007). Tori-Bank is a corpus developed at Tottori University in 2007 in order to compile a Japanese semantic pattern dictionary for compound and complex sentences. The clause boundary patterns are hierarchically defined, in four layers. The top level of the categories consists of Nominal Clauses (名詞節 *Hosoku-setsu*: HS), Adnominal Clauses (名詞修飾節 *Meishishushoku-setsu*: MS), Adverbial Clauses (副詞節 *Fukushi-setsu*: FU), and Coordinate Clauses (並列節 *Heiretsu-setsu*: HR). The second level of the categories is made up of 26 classes. It is described in detail on the website.[2]

(Matsumoto et al., 2018) annotated clause boundary categories on the core BCCWJ data. BCCWJ-ToriClause represents the annotation data of the clause boundary categories on BCCWJ. The data carry a three-layered (top level, second level, and third level) annotation of the clause boundary categories. The right boundaries of clauses are annotated on the data with the labels.

We converted the original annotation of the base phrase (*bunsetsu*) units of the eye-tracking data. Table 2 below presents the data. We use the top and

[2] `unicorn.ike.tottori-u.ac.jp/toribank/`

second level labels of the clause categories with their frequencies in the data.

Table 2: Frequencies of clause boundaries.

Label	Description	Count
HS: Nominal Clause		64
HSa	Noun	28
HSb	Interrogation	2
HSc	Quotation	34
MS: Adnominal Clause		94
MSa	Relative	61
MSb	Apposition	19
MSc	Other	6
MSd	Functional	7
MSe	Collocational	1
FU: Adverbial Clause		83
FUa	Temporal	5
FUb	Causal	20
FUc	Conditional, Concessive	5
FUd	Attendant Circumstances	10
FUe	Contrastive	8
FUf	Objective	3
FUh	Presuppositional	2
FUi	Means	7
FUj	Dyadic	2
FUk	Correlative	1
FUl	Conclusive	12
FUn	Restrictive	1
FUo	Absolute	8
FUp	Other	9
HR: Coordinate Clause		28
HRa	Resultative	27
HRb	Contrastive	1

4 Statistical Model

The statistical model is based on the Bayesian linear mixed model (Sorensen et al., 2016) with the R rstan package. During pre-processing, we excluded the data for {`authorsData`, `caption`, `listItem`, `profile`, `titleBlock`} from the `metadata`. We also excluded zero-millisecond data points from the eye-tracking data. We estimate `time` of the six reading time types (SELF, FFT, FPT, SPT, RPT, and TOTAL) by using `lognormal` function with the following fixed factors: layout information, number of dependencies, and clause boundary categories.

We perform the second level labels of clause boundaries for the analysis. The second level analysis is based on 3 subcategories of nominal clauses, 5 subcategories of adnominal clauses, 14 subcategories of adverbial clauses, and 2 subcategories of coordinate clauses.

We use the formulae in Figure 2 for the second level analysis.

Here, $time$ refers to the target reading time. `lognormal` is the lognormal distribution function in rstan. σ is the standard deviation of `lognormal`. μ is the mean of `lognormal` with the linear formula. α is the intercept of the linear formula. β^{length} is the slope of the variable $length(x)$, which is the length of the base phrase with fixation. β^{space} is the slope of the variable $\chi_{space}(x)$, which is with or without spaces between base phrases on the presentation.[3] $\beta^{sessionN}$, $\beta^{articleN}$, $\beta^{screenN}$, β^{lineN}, and $\beta^{segmentN}$ are the fixed effects of the presentation order indexes $sessionN(x)$, $articleN(x)$, $screenN(x)$, $lineN(x)$, and $segmentN(x)$, respectively. β^{is_first}, β^{is_last}, and $\beta^{is\_second\_last}$ are the fixed effects of the layout factors $\chi_{is\_first}(x)$, $\chi_{is\_last}(x)$ and $\chi_{is\_second\_last}(x)$, respectively. Note, the clause boundary labels are multilabel on the base phrase units, because more than one clause end boundary may appear within a base phrase. Therefore, we modelled the false class boundaries. $\sum_{a(x)\in A} \gamma^{article=a(x)}$ is the random effect for the articles, in which $a(x)$ is the article information of x. $\sum_{s(x)\in S} \gamma^{subj=s(x)}$ is the random effect for the subject, in which $s(x)$ is the subject participant ID of x.

$\sum_{HS?} \beta^{HS?} \cdot \chi_{HS?}$ are fixed effects for the second level nominal phrases. $\sum_{MS?} \beta^{MS?} \cdot \chi_{MS?}$ are fixed effects for the second level adjective phrases. $\sum_{FU?} \beta^{FU?} \cdot \chi_{FU?}$ are fixed effects for the second level adverbial phrases. $\sum_{HR?} \beta^{HR?} \cdot \chi_{HR?}$ are fixed effects for the second level coordinate phrases.

We ran four chains $\times$ 5000 post-warm up iterations, and all models were converged.

5 Results

5.1 Fixed Effects for Other than Clause Related

First, we present the confirmed results of the non-clause related terms. Figures 3 and 4 show the posterior distributions of self-paced reading (SELF) and eye tracking total time (TOTAL), respectively. We present partial results in this article due to page limi-

[3] χ_A is an indicator function

$$\chi_A(x) = \begin{cases} 1 & \text{if} \quad x \in A, \\ 0 & \text{if} \quad x \notin A \end{cases}$$

.

$$
\begin{aligned}
\mu \quad = \quad & \alpha + \beta^{length} \cdot length(x) + \beta^{space} \cdot \chi_{space}(x) + \beta^{dependent} \cdot dependent(x) \\
& + \beta^{sessionN} \cdot sessionN + \beta^{articleN} \cdot articleN(x) + \beta^{screenN} \cdot screenN(x) \\
& + \beta^{lineN} \cdot lineN(x) + \beta^{segmentN} \cdot segmentN(x) \\
& + \beta^{is\_first} \cdot \chi_{is\_first}(x) + \beta^{is\_last} \cdot \chi_{is\_last}(x) + \beta^{is\_second\_last} \cdot \chi_{is\_second\_last}(x) \\
& + \sum_{HS?} \beta^{HS?} \cdot \chi_{HS?}(x) + \sum_{MS?} \beta^{MS?} \cdot \chi_{MS?}(x) + \sum_{FU?} \beta^{FU?} \cdot \chi_{FU?}(x) + \sum_{HR?} \beta^{HR?} \cdot \chi_{HR?}(x) \\
& + \sum_{a(x) \in A} \gamma^{article=a(x)} + \sum_{s(x) \in S} \gamma^{subj=s(x)}.
\end{aligned}
$$

Figure 2: Formulae for the second level analysis.

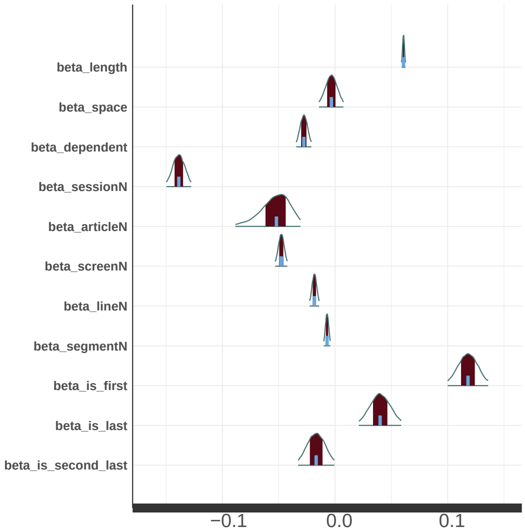


Figure 3: Fixed effects for other than clause related (SELF).

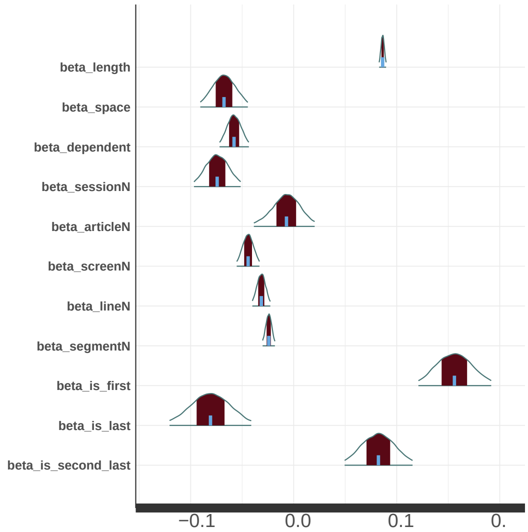


Figure 4: Fixed effects for other than clause related (TOTAL).

tation.[4] A negative value of the coefficient indicates that the factor shortens the reading time. A positive value of the coefficient indicates that the factor lengthens the reading time.

The presentation with `space`s between segments makes the reading time of TOTAL faster than the one without `space`s for the eye tracking methods. To improve the readability of texts, one should simply introduce spaces at base phrase boundaries. The longer `length` of the segment increases reading time, except for FFT, because the gazing area in this case is correlated to the probability of the fixation. More `dependency` arcs make shorter reading times possible for the segment. This fact supports *Anti-locality* (Konieczny, 2000). The layout information (`is_first`, `is_last`, `is_second_last`) is for the eye movement at the text wrap. Reading time is longer at the left most segment (`is_first`). The reading time of FPT, RPT, and Total is longer at the right most and the second right most segments (`is_last`, `is_second_last`). With regard to the presentation order (`sessionN`, `articleN`, `screenN`, `lineN`, `segmentN`), as the experiment progressed, a shorter reading time was observed. This means that the subject participants become more familiar with the experiment.

[4] All the results are presented in the supplementary materials for reviewers.

These results are nearly the same as the linear mixed model results in (Asahara et al., 2016).

5.2 Fixed Effects for the Second Level Clause Categories

Next, we investigate the fixed effects for the second level categories.

First, we present the adnominal phrases (MS). Figures 5 and 6 show the fixed effects for the second level adnominal clause categories for TOTAL and SPT.

The relative clauses (MSa) tend to require a shorter reading time than the appositional clauses (MSb) in TOTAL. This is a result of the difference in SPT. The example (1) shows a relative clause, in which the predicate in the relative clause has a predicate argument relation with the clause modified word. The example (2) shows an apposition clause, in which the predicate in the relative clause does not have a predicate argument relation with the clause modified word. The predicate-argument relation facilitates the reading process of human beings.

(1) 幼稚園から 大学まで 通った
youchienkara daigakumade kayotta
青山学院では、
aoyamagakuindeha,
'In Aoyama Gakuin, to which she went from kindergarden to university,'
(Yomiuri Newspaper 2001 [BCCWJ: 00001_A_PN1c_00001_A_1])
MSa200: Adnomial Clause: Relative Clause: non-restrictive use

(2) 支払利息や 減価償却費の
shiharairisokuya genkashoukyakuhino
計上額が 少ない 傾向が ある。
keijougakuga sukunai keikouga aru.
'The interest expense and depreciation expense tend to be less recorded.'
(Hokkaido Newspaper 2002 [BCCWJ: 00005_A_PN2e_00001_A_2])
MSb: Adnomial Clause: Apposition

We now present the nominal clauses (HS). We focus on the frequent labels of HSa (noun) and HSc (quotation). The example (3) is the noun clause with a dummy noun こと (*koto*). The example (4) is the quotation clause with a quotation particle と (*to*). The noun clause (HSa) needs a shorter reading time than the quotation clause (HSc) in self-paced reading (Figure 9) and (Figure 10).

(3) タイミングよく まぶたを 閉じてくれた
taiminguyoku mabutawo tojitekureta
ことで、独特な雰囲気の 写真に
kotode dokutokuna funikino
なりました。
shashinni narimashita
'Closing eyes timely makes uniq atomosphere on the photo.'
(Sankei Newspaper 2001 [BCCWJ: 00002_A_PN1d_00001_B_1])
HSa: Nominal Clause: Noun Clause

(4) シャープの
sha-puno
携帯情報端末「ザウルス」の
keitaijouhoutannmatsu "Zaurus" no
コンテンツを ５月中旬から
kontentsuwo gogatsuchuujunkara
販売すると 発表した。
hanbaisuruto happyoushita.
'SHARP published that they will release the contents of PDA "Zaurus"'
(Sankei Newspaper 2001 [BCCWJ: 00015_A_PN1d_00002_B_5])
HSc: Nominal Clause: Quotation Clause

Finally, we present the adverbial phrases (FU). We focus on the frequent labels of FUb (causal) and FUd (attendant circumustances). Figure 7 and 8 show the fixed effects for the two adverbial clause categories for SELF and FPT. The example (5) shows the causal relation, in which the subordinate clause is the cause of the main clause effect. The example (6) consists of the attendant circumstances, in which the subordinate clause represents a state around the event of the main clause. Interestingly, the reading time lengths between FUb and FUd are different in the presentation styles. The causal clauses require a shorter reading time than the attendant circumstances in the self-paced reading. The contrary state holds true in the eye-tracking method. The て '*te*' form has ambiguity among attendant circumstance, quotation, means, and coordi-

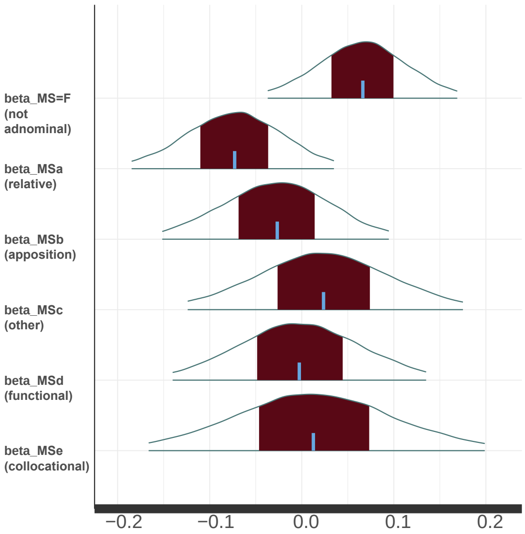


Figure 5: Fixed effects for the second level adnominal clause categories (TOTAL).

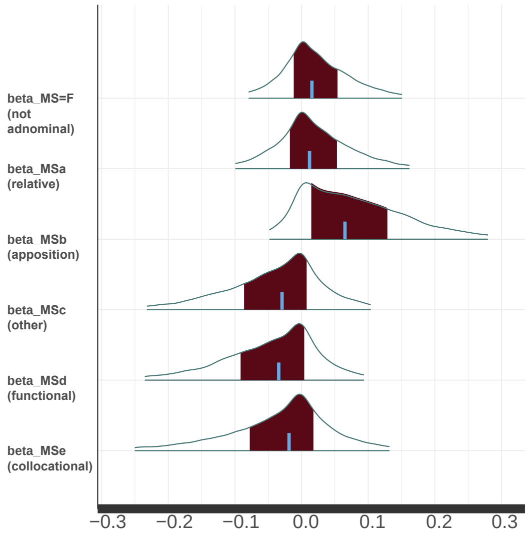


Figure 6: Fixed effects for the second level adnominal clause categories (SPT).

nation. This may come from the parafoveal preview benefit in the eye-tracking method, which is unavailable in the self-paced reading manner.

(5) 「しゃべるのが 得意なんだから、
“shaberunoga tokuinandakara,
能力を 生かしてみたら」と、
noryokuwo ikashitemitara”to,
‘(she said) that “because you are good at talking, you should use this skill.” ’
(Yomiuri Newspaper 2001 [BCCWJ: 00001_A_PN1c_00001_A_1])
FUb: Adverbial Clause: Causal

(6) もみじの 木に とまって 仲良く
momijinokini tomatte nakayoku yorisou
寄り添う 二羽の キジバト。
niwano kijibato.
‘The two eastern turtle doves are perched on the maple and cuddling close together.’
(Sankei Newspaper 2001 [BCCWJ: 00002_A_PN1d_00001_B_1])
FUd: Adverbial Clause: Attendant Circumustances

The result does not support the wrap-up effect (Just and Carpenter, 1980; Rayner et al., 2000) in English, in which readers tend to spend a longer time while reading clause-end phrases than clause-internal phrases.

6 Conclusions

We presented a contrastive analysis between reading time and clause boundary categories in Japanese.

Generally, the clause end boundaries tend to shorten the reading time compared to the other parts. Though it does not support clause wrap-up effects, as clause-end phrases were read faster than clause-internal phrases, it is compatible with the observation of a reliable anti-locality effect, as words were read faster when more dependents preceded them.

We found that the clause boundary categories affect the reading time. Relative clauses tend to have a shorter reading time than apposition clauses. This is because relative clauses and noun clauses have a predicate-argument relation with the clause modified word. The predicate-argument relation promotes the reading time by predicting the words that follow.

Noun clauses with formal nouns tend to require a shorter reading time than quotation clauses. We would like to explore the reason for the differences that occur in the reading times.

We also found incompatible results between the self-paced reading and eye-tracking methods as regards the adverbial clauses of causal and attendant circumstances. This is because the latter atten-

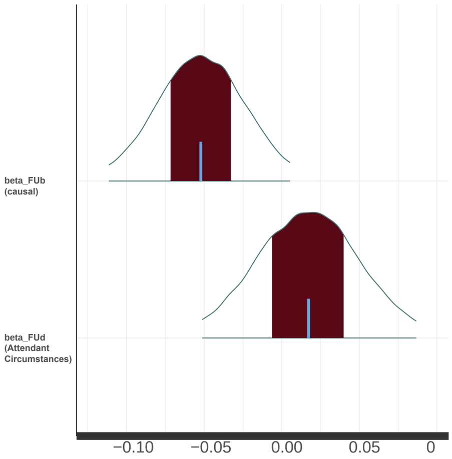


Figure 7: Fixed effects for the second level adverbial clause categories (SELF).

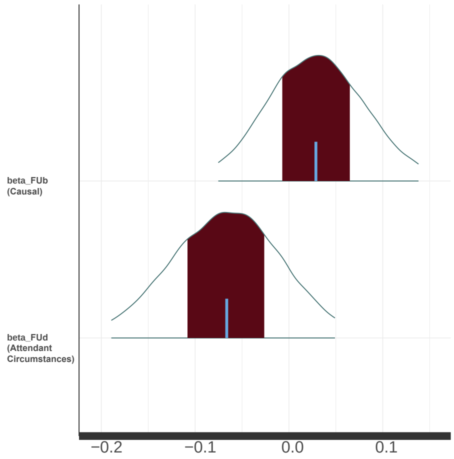


Figure 8: Fixed effects for the second level adverbial clause categories (FPT).

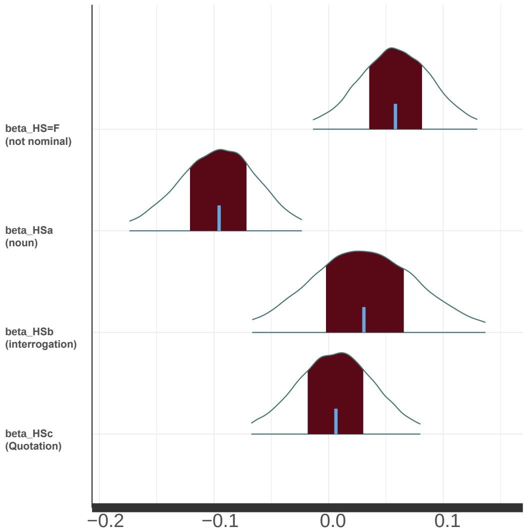


Figure 9: Fixed effects for the second level nominal clause categories (SELF).

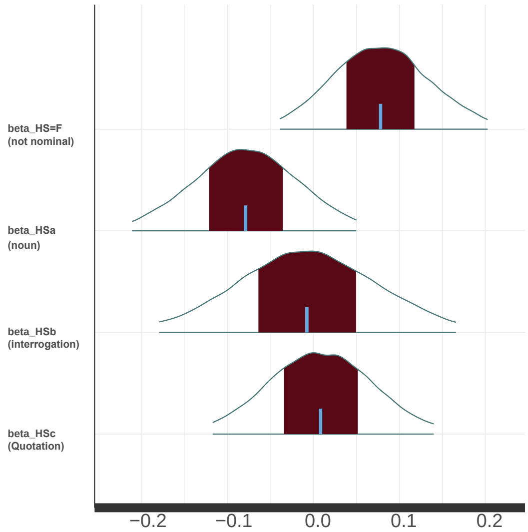


Figure 10: Fixed effects for the second level nominal clause categories (TOTAL).

dant circumstance form has ambiguity among other clause boundary categories.

In our future work, we intend to develop the readability estimation application based on the fundamental psycholinguistic research with natural language processing techniques.

Acknowledgments

This work was supported by JSPS KAKENHI Grants Number 17H00917, and 18H05521 and a project of the Center for Corpus Development, NINJAL.

References

Masayuki Asahara and Sachi Kato. 2017. Between Reading Time and Syntactic/Semantic Categories. In *Proceedings of the The 8th International Joint Conference on Natural Language Processing*, pages 404–412.

Masayuki Asahara and Yuji Matsumoto. 2016. BCCWJ-DepPara: A Syntactic Annotation Treebank on the 'Balanced Corpus of Contemporary Written Japanese'. In *Proceedings of the 12th Workshop on Asian Langauge Resources (ALR12)*, pages 49–58.

Masayuki Asahara, Hajime Ono, and Edson T. Miyamoto. 2016. Reading-Time Annotations for 'Balanced Corpus of Contemporary Written Japanese'. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 684–694.

Masayuki Asahara. 2017. Between Reading Time and Information Structure. In *Proceedings of The 31st Pacific Asia Conference on Language, Information and Computation PACLIC 31 (2017)*.

Maria Barrett, Joachim Bingel, Frank Keller, and Anders Søgaard. 2016. Weakly supervised part-of-speech tagging using eye-tracking data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 579–584.

S. L. Frank, I. F. Monsalve, R. L. Thompson, and G. Vigliocco. 2013. Reading time data for evaluating broad-coverage models of english sentence processing. *Behavior Research Methods*, 45(4):1182–1190.

R. Futrell, E. Gibson, H. J. Tily, I. Blank, A. Vishnevetsky, S. T. Piantadosi, and E. Fedorenko. 2018. The natural stories corpus. In *Proceedings of LREC-2018*.

R. Hill and W. Murray, 2000. *Reading as a perceptual process*, chapter Commas and spaces: effects of punctuation on eye movements and sentence processing, pages 565–589. Elsevier, Amsterdam.

M. Hirotani, L. Frazier, and K. Rayner. 2006. Punctuation and intonation effects on clause and sentence wrap-up: Evidence from eye movements. *Journal of Memory and Language*, 54:425–443.

S. Husain, S. Vasishth, and N. Srinivasan. 2015. Integration and prediction difficulty in hindi sentence comprehension: Evidence from an eye-tracking corpus. *Journal of Eye Movement Research*, 8(2).

Satoru Ikehara. 2007. Japanese Semantic Pattern Dictionary – Compound and Complex Sentence Eds. – . `http://unicorn.ike.tottori-u.ac.jp/toribank/`.

Marcel Adam Just and Patricia A. Carpenter. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4):329–354.

Marcel Adam Just, Patricia A. Carpenter, and Jacqueline D. Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 3:228–238.

Alan Kennedy and Joel Pynte. 2005. Parafoveal-on-foveal effects in normal reading. *Vision Research*, 45:153–168.

Sigrid Klerke, Héctor Martínez Alonso, and Anders Søgaard. 2015. Looking hard: Eye tracking for detecting grammaticality of automatically compressed sentences. In *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)*, pages 97–105.

R. Kliegl, A. Nuthmann, and R. Engbert. 2006. Tracking the mind during reading: Theinfluence of past, present, and future words on fixation durations. *Journal of Experimental Psychology*, 135(1):12–35.

Lars Konieczny. 2000. Locality and parsing complexity. *Journal of Psycholinguistic Research*, 29(6).

Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den. 2014. Balanced Corpus of Contemporary Written Japanese. *Language Resources and Evaluation*, 48:345–371.

Satomi Matsumoto, Masayuki Asahara, and Setsuko Arita. 2018. Japanese Clause Classification Annotation on the 'Balanced Corpus of Contemporary Written Japanese'. In *Proceedings of the 13th Workshop on Asian Language Resources (ALR13)*, pages 1–8.

Takuya Miyauchi, Masayuki Asahara, Natsuko Nakagawa, and Sachi Kato. 2017. Annotation of Information Structure on "The Balanced Corpus of Contemporary Written Japanese". In *Proceedings of PACLING 2017, the 15th International Conference of the Pacific Association for Computational Linguistics*, pages 166–175.

Keith Rayner, Gretchen Kambe, and Susan A. Duffy. 2000. The effect of clause wrap-up on eye movements during reading. *The quarterly journal of experimental psychology*, 53A(4):1061–1080.

Tanner Sorensen, Sven Hohenstein, and Shravan Vasishth. 2016. Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *Quantitative Methods for Psychology*, 12:175–200.

Tessa Warren, Sarah J. White, and Erik D. Reichle. 2009. Investigating the causes of wrap-up effects: Evidence from eye movements and E-Z reader. *Cognition*, 111(1):132–137.

M. Yan, R. Kliegl, E. M. Richter, and H. Shu. 2010. Flexible saccade-target selection in Chinese reading. *The Quarterly Journal of Experimental Psychology*, 63(4):705–725.