

Between Reading Time and Syntactic/Semantic Categories

Masayuki Asahara † and Sachi Kato ‡

National Institute for Japanese Language and Linguistics, Japan †, ‡

masayu-a@ninjal.ac.jp †

Abstract

This article presents a contrastive analysis between reading time and syntactic/semantic categories in Japanese. We overlaid the reading time annotation of BCCWJ-EyeTrack and a syntactic/semantic category information annotation on the ‘Balanced Corpus of Contemporary Written Japanese’. Statistical analysis based on a mixed linear model showed that verbal phrases tend to have shorter reading times than adjectives, adverbial phrases, or nominal phrases. The results suggest that the preceding phrases associated with the presenting phrases promote the reading process to shorten the gazing time.

1 Introduction

Most of the studies on sentence processing by humans are based on confirmatory data analysis. The methodology involves developing a hypothesis, constructing sample sentences, including the target language phenomena, and performing a psycholinguistic experiment, such as recording reading time or event-related potentials. In recent times, the ‘Balanced Corpus of Contemporary Written Japanese’ (hereafter ‘BCCWJ’) (Maekawa et al., 2014) was compiled and published. The reading time annotation on BCCWJ: BCCWJ-EyeTrack (Asahara et al., 2016) is available for the linguistic research community. The data in the BCCWJ enable us to perform exploratory data analysis in fair and reproducible environments.

We measured the readability of humans. More concretely, we performed a contrast comparison between reading time and syntactic/semantic categories of words. We prepared the annotation

of word senses on BCCWJ based on ‘Word List by Semantic Principles’ (国立国語研究所, 1964, 2004). The original WLSP label annotation is on both short unit words and long unit words in the BCCWJ. We then mapped these annotations into *Bunsetsu*(base phrase)-units.

The statistical analysis using a mixed linear model shows that verbal phrases tend to have shorter reading times than adjective/adverbial phrases or nominal phrases.

Section 2 presents the related research. Section 3 shows the data and methods. Section 4 presents the results, and Section 5 is the discussion. Section 6 concludes this article and presents the implications of our current work and the future work we plan to conduct.

2 Related Work

First, we present related work on eye tracking. The Dundee Eyetracking Corpus (Kennedy and Pynte, 2005) contains reading times for English and French newspaper editorials from 10 native speakers of each language that were recorded using eye-tracking equipment. The corpus does not target a specific set of linguistic phenomena but instead provides naturally occurring texts for testing diverse hypotheses. For example, (Demberg and Keller, 2008) used the corpus to test Gibson’s dependency locality theory (DLT) (Gibson, 2008) and Hale’s surprisal theory (Hale, 2001). The corpus also allows for replications to be conducted; for example, (Roland et al., 2012) concluded that previous analyses (Demberg and Keller, 2007) had been distorted by the presence of a few outlier data points.

Second, we present language analyses or models with reading time or eye tracking gaze information. (Barrett et al., 2016) presented a POS tagging model with gaze patterns. (Klerke et al.,

Table 1: Data format of BCCWJ-EyeTrack

name	type	decription
surface	factor	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
length	int	number of characters
space	factor	segment boundary with space or not
subj	factor	participant ID
setorder	factor	presentation order
dependent	int	syntactic dependency
sessionN	int	session order
articleN	int	article display order
screenN	int	screen display order
lineN	int	line display order
segmentN	int	segmentation display
is_first	factor	the left most
is_last	factor	the right most
is_second_last	factor	the second right most
WLSPLUWFALSE	factor	unknown word in WLSP
WLSPLUWA	factor	semantic category in WLSP
WLSPLUWB	factor	syntactic category in WLSP

2015) presented a grammaticality detection model for machine-processed sentences. (Iida et al., 2013) presented an analysis of eye-tracking data for the annotation of predicate-argument relations.

Our paper is slightly different from these preceding papers. We present a corpus-based psycholinguistic research on the relationship between reading time and syntactic/semantic categories.

3 Data and Method

We used the overlaid data of BCCWJ-EyeTrack and syntactic/semantic categories, as given in Table 1. We present the data below in detail.

3.1 BCCWJ and its annotation

We used BCCWJ (Maekawa et al., 2014) and its annotation data. BCCWJ is a balanced corpus of Japanese. We used newspaper articles from the core data. The data were sampled by their production. The sentences were segmented into word unit boundaries of short unit words, long unit words, and *bunsetsu*. The morphological information for the short unit words and long unit words was annotated by human annotators.

We also used *bunsetsu*-based syntactic depen-

ency annotation (Asahara and Matsumoto, 2016) for the data to investigate the correlation between syntactic dependency attachments and reading time.

3.2 Reading Time Data: BCCWJ-EyeTrack

We now explain the two methods used for measuring the reading time: eye tracking and self-paced reading. The order of tasks was fixed with eye tracking in the first session and self-paced reading in the second session. Each participant saw each text once with the task and segmentation of the texts 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 using 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 each participant could freely return and reread earlier parts of the text on the same screen. However, participants were not allowed to return to previous screens. Stimulus texts were shown in a fixed full-width font (MS Mincho 24 point) and displayed horizontally as is customary with computer displays for Japanese; there were five lines per screen on a 21.5-in display.¹ Under the segmented condition, a half-width space was used to indicate the boundary between segments. In order to improve vertical tracking accuracy, three empty lines were placed between the lines of text. A line break was inserted at the end of a sentence or when the maximum 53 full-width characters per line was attained. Moreover, line breaks were inserted at the same points in the segmented and unsegmented conditions to guarantee that the same number of non-space characters was shown under both conditions.

The same procedure was adopted for the self-paced reading presentation except that the chin rest was not used, and participants could move their heads freely while looking directly at the display. Doug Rohde’s Linger program Version 2.94² was used to record keyboard-press latencies while sentences were shown using a non-cumulative self-paced moving-window presenta-

¹EIZO FlexScan EV2116W (resolution: 1920×1080 pixels) set at 50 cm from the chin rest.

²<http://tedlab.mit.edu/~dr/Linger/>

tion. This had the best correlation with eye-tracking data when different styles of presentation were compared for English (Just et al., 1982). Sentence segments were initially shown masked with dashes. Participants pressed the space key of the keyboard to reveal each subsequent segment of the sentence, while all other segments reverted to dashes. Participants were not allowed to go back and reread earlier segments.

Twenty-four native Japanese speakers, who were 18 years or older at the time, participated in the experiment with due financial compensation. The experiments were conducted from September to December 2015. The collected profile data included the age (in 5-year brackets), gender, educational background, eyesight (all participants had uncorrected vision or vision corrected with soft contact lenses or prescription glasses), geographical linguistic background (i.e., the prefecture within Japan where they lived until the age of 15), and parents' place of birth. The vocabulary size of the participants was measured using a Japanese language vocabulary evaluation test (Amano and Kondo, 1998). Participants indicated words they knew from a list of 50 words, and scores were calculated by taking word-familiarity estimates into consideration. As a measure of the working memory capacity, the Japanese version of a reading span test was conducted (Osaka and Osaka, 1994). Each participant read sentences aloud, each of which contained an underlined content word. After each set of sentences, the participants recalled the underlined words. If they successfully recalled all the words, the set size was increased by one sentence (sets of two to five sentences were used). The final score was the largest set for which all words were correctly recalled; a half point was added if half the number of words were recalled in the last trial.

Reading times were collected for a subset of the core data of the BCCWJ (Maekawa et al., 2014), which consisted of newspaper article (PN: published newspaper) samples. Articles were chosen if they were annotated with information such as syntactic dependencies, predicative clausal structures, co-references, focus of negation, and similar details following the list of articles that were given annotation priority in the BCCWJ.

The 21 newspaper articles³ chosen were divided

³The original BCCWJ-EyeTrack paper (Asahara et al., 2016) presented 20 articles. However, there were two con-

into four data sets containing five articles each: A, B, C, and D. Table 2 presents the numbers of words, sentences, and screens (i.e., pages) for each data set. Each article was presented starting on a new screen.

Table 2: Data set sizes

Data set	Segments	Sentences	Screens
A	470	66	19
B	455	67	21
C	355	44	16
D	363	41	15

Articles were shown segmented or unsegmented (i.e., with or without a half-width space to mark the boundary between segments). Segments conformed to the definition for *bunsetsu* units (a content word followed by functional morphology, e.g., a noun with a case marker) in the BCCWJ as prescribed by the National Institute for Japanese Language and Linguistics. Each participant was assigned to one of eight groups of three participants each. Each group was subjected to one of the eight experimental conditions with varying combinations of measurement methods, and boundary marking for different data sets was presented in different orders.

During the self-paced reading session, each segment was displayed separately, and participants could not return to reread earlier parts of the text. Therefore, the latencies for the button presses are straightforward measures of the time spent on each segment.

With regard to data from eye tracking, 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). These are explained in Figure 1.

The FFT is the duration of fixation measured when the gaze first enters the area of interest. In the figure, the FFT for “the first fiscal year settling of accounts also” (hereafter “the area of interest”) is the duration of fixation 5.

The 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. In the figure, the FPT

secutive articles in data set C. These two articles were presented on separate screens. Thus, we split them into two for statistical analysis.

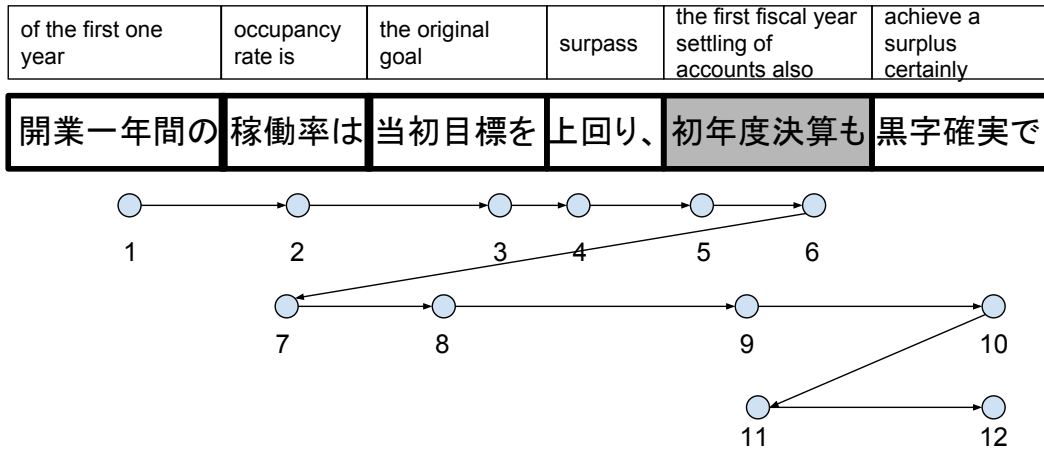


Figure 1: Example of fixations

is the sum of the durations of fixations 5 and 6.

The RPT is the total span of time from the moment the gaze enters the area of interest until it crosses the right boundary of this area for the first time. In the figure, the RPT is the sum of the durations for fixations 5–9. The RPT can include fixations to the left of the left boundary (e.g., 7 and 8) and the durations of fixations when the gaze returns to the area of interest (e.g., 9).

The SPT is the total span of time the gaze rests in the area of interest excluding the FPT. In the figure, the SPT is the sum of the durations of fixations 9 and 11.

The TOTAL is the total duration the gaze rests within the area of interest. In other words, it is the sum of SPT and FPT. In the figure, TOTAL is the sum of the durations of fixations 5, 6, 9, and 11.

Table 1 presents the data. `surface` is the surface form of the word. `reading time` (i.e., `time`) is converted into log scale (i.e., `logtime`). `measure` is the reading type {SELF, FFT, FPT, RPT, SPT, TOTAL}. `sample`, `article`, `metadata_orig`, `metadata` are information related to the article. `length` is the number of characters in the surface form. `space` denotes spaces, if they are present between segments. `subj` is the participant ID, which is used as a random effect for the statistical analysis. `dependent` is the number of dependents for the segments. The dependency relation is annotated by humans (Asahara and Matsumoto, 2016). `sessionN`, `articleN`, `screenN`, `lineN`, `segmentN` are the display order of the elements. `is_first`, `is_last`, `is_second_first`

are the layout features on the screen. `WLSPLUWFALSE`, `WLSPLUWA`, `WLSPLUWB` are described in the next subsection.

3.3 WLSP and annotation

‘Word List by Semantic Principles’ (*Bunrui Goi-hyo*) (国立国語研究所, 1964) is ‘a A collection of words classified and arranged by their meanings’. The first published version of WLSP in 1964 includes around 33,000 words. The revised and enlarged version of WLSP (国立国語研究所, 2004) was published in 2004. The data include around 79,000 word tokens with 100,000 word sense tokens.

Table 3 shows an example entry ‘この (*kono*: this)’ in WLSP. The article number ‘3.1010’ identifies a word belonging to the syntactic/semantic category. The first digit of the article number refers to ‘class’, which is a syntactic category of the entry: class ‘1’ represents a ‘体’ nominal entry; class ‘2’ represents a ‘用’ verbal entry; class ‘3’ is an ‘相’ adjective entry; and class ‘4’ is a ‘他’ other entry including conjunctive and interjection. This category classification is originally from the ‘Awakening of Faith in the Mahayana’ (大乘起信論; 大乘起信論) in Mahayana Buddhism.

The digits to the right of a period identify the semantic category. The first decimal digit represents a ‘division’, which is a major semantic category: division ‘.1’ is a ‘抽象的關係 (関係)’ relation entry; division ‘.2’ is a ‘人間活動の主体 (主体)’ subject entry; division ‘.3’ is a ‘精神および行為 (活動)’ action entry; division ‘.4’ is a ‘生産物および用具 (生産物)’ product entry; and division ‘.5’ is a ‘自然物および自然現象 (自然)’ nature entry. The

Table 3: WLSP example entry ‘この (*kono*)’ (article number 3.1010)

class	devison	section	article
相 (3)	関係 (.1)	真偽 (.10)	こそあど (.1010)
Adj/adv	Relation	Boolean	Demonstrative

first two decimal digits refer to the ‘section’. Four decimal digits refer to the ‘article’, which is article number 895, of the finest semantic categories.

We annotated the words from these WLSP article numbers based on the BCCWJ core samples. The annotation was carried out for content words for short unit words and long unit words of BCCWJ. Functional words were not annotated in the WLSP category. Now, the samples of BCCWJ-EyeTrack have already been annotated. We defined the set of right-most long unit words as the category of the *bunsetsu*. The semantic category (class) and syntactic category (division) were reassigned on segments. We called them WLSPLUWA and WLSPLUWB, respectively. We note that, there are still unassigned entries for the segment even if all the words have been manually checked. We assigned the boolean value of WLSPLUWFALSE for the unassigned words.

3.4 Statistical Analysis

We investigated the reading time (*logtime*) of NPs that were annotated with the WLSP labels. Whereas Asahara et al.’s paper was based on *time*, ours was based on *logtime* to reduce the outliers in the model. During the preprocessing, we excluded data {*authorsData*, *caption*, *listItem*, *profile*, *titleBlock*} of metadata. We also excluded zero-millisecond data points from the eye tracking data. The number of data points were 17,628 for SELF (100.0%); 13,232 for FFT, FPT, RPT, and TOTAL (75.0%); and 4,769 for SPT (27.0%). After model-based trimming was used to eliminate points beyond 3.0 standard deviations, the model was rebuilt (Baayen, 2008). *subj* and *article* were considered as random effects, as expressed in the formula in Figure 2. We used the *lme4* package on R.

4 Results

Table 4 shows the results. Each number shows the coefficient with the standard error in brackets. 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 base fixed effect of the syntactic category is the nominal phrase (WLSPLUWA1), and the base fixed effect of the semantic category is the relation (WLSPLUWB1). Note that the time is based on logarithm.

First, we confirm the results of the non-WLSP related terms. The presentation with spaces between segments makes the reading time of FPT, RPT, SPT, and TOTAL faster than the one without spaces for eye tracking methods. To improve the readability of texts, one should simply introduce spaces at *Bunsetsu* boundaries. The longer length of the segment makes reading times long 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 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. All reading times other than SPT 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, the reading time became shorter. This means that the subject participants become more familiar with the experiment.

Next, we confirm the results related to the WLSP syntactic categories. For all types of reading times, the verbal segments (WLSPLUWA2) had significantly shorter reading times than the nominal segments (WLSPLUWA1). For reading time types other than FFT, the adjective/adverbial segments (WLSPLUWA3) had significantly shorter reading times than the nominal segments (WLSPLUWA1). For reading time types other than SPT, the adjective/adverbial segments (WLSPLUWA3) had significantly longer reading times than the verbal segments (WLSPLUWA1).

```

logtime ~ space * sessionN + length + dependent
+ is_first + is_last + is_second_last
+ articleN + screenN + lineN + segmentN
+ WLSPLUWFALSE + WLSPLUWA + WLSPLUWB
+ (1 | subj) + (1 | article)

```

Figure 2: Lmer formula for the statistical analysis

Table 4: The results of statistical analysis

	<i>Dependent variable:</i>					
	logtime					
	SELF	FFT	FPT	SPT	RPT	TOTAL
space=True	-0.001 (0.002)	-0.006 (0.004)	-0.017*** (0.005)	-0.039*** (0.009)	-0.018*** (0.006)	-0.029*** (0.005)
length	0.086*** (0.001)	-0.003 (0.002)	0.135*** (0.003)	0.022*** (0.005)	0.115*** (0.003)	0.130*** (0.003)
dependent	-0.008*** (0.002)	-0.003 (0.002)	-0.016*** (0.003)	-0.016*** (0.006)	-0.012*** (0.004)	-0.018*** (0.003)
is_first	0.052*** (0.004)	0.019*** (0.006)	0.090*** (0.008)	-0.027** (0.013)	0.030*** (0.009)	0.069*** (0.008)
is_last	0.033*** (0.004)	-0.009 (0.006)	0.014* (0.008)	-0.052*** (0.016)	0.088*** (0.010)	-0.007 (0.008)
is_second_last	-0.010*** (0.004)	-0.001 (0.006)	0.034*** (0.007)	-0.005 (0.012)	0.045*** (0.008)	0.034*** (0.007)
sessionN	-0.022 (0.021)	-0.022 (0.016)	-0.041* (0.024)	-0.036** (0.018)	-0.049* (0.025)	-0.047* (0.024)
articleN	-0.028*** (0.005)	-0.004 (0.004)	-0.005 (0.007)	-0.002 (0.007)	-0.007 (0.007)	-0.001 (0.008)
screenN	-0.029*** (0.002)	-0.004 (0.003)	-0.018*** (0.003)	-0.015*** (0.006)	-0.017*** (0.004)	-0.025*** (0.003)
lineN	-0.010*** (0.001)	-0.010*** (0.002)	-0.018*** (0.003)	-0.018*** (0.005)	-0.007** (0.003)	-0.018*** (0.003)
segmentN	-0.004*** (0.001)	0.003*** (0.001)	-0.005*** (0.001)	-0.009*** (0.002)	-0.013*** (0.002)	-0.012*** (0.001)
WLSPLUWFALSE (unassigned word)	-0.030 (0.019)	0.020 (0.061)	-0.075 (0.076)	-0.031 (0.299)	-0.109 (0.092)	-0.160** (0.079)
WLSPLUWA2 (verb)	-0.047*** (0.004)	-0.038*** (0.006)	-0.096*** (0.007)	-0.029** (0.014)	-0.088*** (0.009)	-0.101*** (0.008)
WLSPLUWA3 (adj/adv)	-0.036*** (0.005)	-0.003 (0.008)	-0.056*** (0.010)	-0.034* (0.020)	-0.054*** (0.012)	-0.071*** (0.010)
WLSPLUWA4 (other)	-0.031* (0.018)	-0.020 (0.033)	-0.127*** (0.040)	-0.238** (0.100)	-0.137*** (0.049)	-0.189*** (0.042)
WLSPLUWB.2 (subject)	0.001 (0.004)	0.014** (0.006)	0.018** (0.007)	0.011 (0.013)	0.005 (0.009)	0.018** (0.008)
WLSPLUWB.3 (action)	-0.007** (0.003)	0.015*** (0.005)	0.024*** (0.006)	0.012 (0.011)	0.021*** (0.007)	0.023*** (0.006)
WLSPLUWB.4 (product)	0.017*** (0.007)	0.005 (0.010)	0.022* (0.013)	0.009 (0.021)	0.018 (0.015)	0.037*** (0.013)
WLSPLUWB.5 (nature)	0.014 (0.010)	0.034** (0.015)	0.017 (0.019)	0.054 (0.034)	0.024 (0.023)	0.040** (0.020)
space1:sessionN	-0.016 (0.042)	0.044 (0.031)	0.059 (0.049)	0.060* (0.035)	0.061 (0.050)	0.061 (0.048)
Constant	2.790*** (0.022)	2.299*** (0.017)	2.532*** (0.026)	2.456*** (0.023)	2.603*** (0.027)	2.672*** (0.026)
Observations	17,628	13,232	13,232	4,769	13,232	13,232

Note:

*p<0.1; **p<0.05; ***p<0.01

Finally, we confirm the result related to WLS semantic categories. The abstract relation (WLSPLUWB.1) shows significantly longer reading times of FFT and TOTAL than of

others such as subject (WLSPLUWB.2), action (WLSPLUWB.3), and product (WLSPLUWB.4).

5 Discussions

In this section, we discuss why reading time varies in syntactic and semantic categories.

Anti-locality is the term used to describe the phenomenon in which segments with more dependents in their preceding context have shorter reading times (Konieczny, 2000). This phenomenon was reported for German double objects (Konieczny and Döring, 2003). It was then investigated for Japanese double objects (Uchida et al., 2014). These shortened reading times cannot be explained by the predictions of the working memory models, in which segments with more dependents load for the reading (Gibson, 2008), or in which the number of dependents do not affect the reading time of the succeeding segments.

This phenomenon is compatible with surprisal theory (Hale, 2001; Levy and Gibson, 2013). It explains how double objects of head final languages, in which the predicate has both a direct and an indirect object tend to have shorter reading times than one that has only a direct object. Asahara et al. (2016) investigated the *anti-locality* phenomenon in more general settings with the dependency from BCCWJ-DepPara (Asahara and Matsumoto, 2016). The results show that the segment with higher dependency has a shorter reading time than a segment with a lower dependency.

In this research, the reading time tends to be shorter in the order of Noun (体, WLSPLUWA1) > Adjective/Adverb (相, WLSPLUWA3) > Verb (用, WLSPLUWA2) in the syntactic categories. The noun (WLSPLUWA1) tends to indicate the object and to become the argument of a predicate such as a verb or an adjective. Although the noun can also become a predicate with a copula verb, the modifier or argument for the noun is limited. The category (WLSPLUWA3) includes a predicative adjective with arguments. The verb (WLSPLUWA2) tends to be a predicate with arguments at the clause end. The tendency is reliable because the standard errors of the coefficients are very small. Though we included dependency as a fixed factor, we observed these tendencies for the reading time, in which the syntactic category with more argument tends to have a shorter reading time than the others. It indicates that arguments of a predicate in Japanese tend not to be overtly appearing in the context. The omitted arguments may help predict the upcoming predicate, although the ar-

guments tend to be omitted in the context. Therefore, the results do not support the working memory model, in which the load to memorize the preceding contexts interferes with the reading. The prediction model is a more plausible hypothesis than the working memory model.

In the semantic category, the abstract relation has a shorter reading time than others. The relation has at least two arguments. The existence of the arguments helps to promote the reading time.

6 Conclusions

This article explores the correlation between reading time and the syntactic/semantic category of the text. The reading time tends to be shorter in the order of Noun (1) > Adjective/Adverb (3) > Verb (2) in the syntactic categories. The relation (WLSPLUWB.1) tends to be the shortest in the semantic categories. The results show that the *bunsetsu* with arguments tend to have shorter reading times than the ones without arguments. This fact supports the *anti-locality* (Konieczny and Döring, 2003) and Hale’s surprisal theory (Hale, 2001).

Our current work comprises two analyses. The first one is a contrastive analysis between reading time and information structure annotation. We overlaid the annotation of information structures (Miyachi et al., 2017) on the reading time data. The result showed that reading time can reveal the difference in whether the target nominal phrase is hearer-new or bridging (Asahara, 2017). The second one is contrastive analysis between reading time and the clause boundary category annotation. The result shows that the clause end segments tend to have shorter reading times. Furthermore, the reading time of clause boundaries vary according to the classification of the clauses.

In our future work, we plan to introduce Bayesian linear mixed model (Sorensen et al., 2016) for the statistical modelling. We also hope to investigate the correlation between reading time and word familiarity rate. Word familiarity rate is the fundamental data to estimate Japanese language vocabulary evaluation test (Amano and Kondo, 1998). However, word-familiarity-rate data were constructed around 20 years ago. We now plan to reconstruct word-familiarity-rate data on WLSPL entries by crowd sourcing using a Bayesian linear mixed model.

Acknowledgments

The work reported in this article was supported by the NINJAL research project of the Center for Corpus Development. This work was also supported by JSPS KAKENHI Grant Number JP25284083 and JP17H00917.

References

- S. Amano and T. Kondo. 1998. Estimation of mental lexicon size with word familiarity database. In *Proceedings of International Conference on Spoken Language Processing*, volume 5, pages 2119–2122.
- M. Asahara. 2017. Between reading time and information structure. In *Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation (PACLIC 31)*, page (to appear).
- M. Asahara and Y. Matsumoto. 2016. BCCWJ-DepPara: A Syntactic Annotation Treebank on the ‘Balanced Corpus of Contemporary Written Japanese’. In *Proceedings of the 12th Workshop on Asian Language Resources (ALR12)*, pages 49–58.
- M. Asahara, H. Ono, and E. 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.
- R. H. Baayen. 2008. *Analyzing Linguistic Data: A practical Introduction to Statistics using R*. Cambridge University Press.
- M. Barrett, J. Bingel, F. Keller, and A. Sogaard. 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.
- V. Demberg and F. Keller. 2007. Eye-tracking evidence for integration cost effects in corpus data. In *Proceedings of the 29th Meeting of the Cognitive Science Society (CogSci-07)*, pages 947–952.
- V. Demberg and F. Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2):193–210.
- E. Gibson. 2008. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68:1–76.
- J. Hale. 2001. A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the second conference of the North American chapter of the association for computational linguistics*, volume 2, pages 159–166.
- R. Iida, K. Mitsuda, and T. Tokunaga. 2013. Investigation of annotator’s behaviour using eye-tracking data. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 214–222.
- M. A. Just, P. A. Carpenter, and J. D. Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 3:228–238.
- A. Kennedy and J. Pynte. 2005. Parafoveal-on-foveal effects in normal reading. *Vision Research* 45:153–168.
- Sigrid Klerke, Héctor Martínez Alonso, and Anders Sogaard. 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.
- L. Konieczny. 2000. Locality and parsing complexity. *Journal of Psycholinguistic Research* 29(6).
- L. Konieczny and P. Döring. 2003. Anticipation of clause-final heads. evidence from eye-tracking and srms. In *Proceedings of the 4th International Conference on Cognitive Science*.
- R. Levy and E. Gibson. 2013. Surprisal, the pdc, and the primary locus of processing difficulty in relative clauses. *Frontiers in Psychology* 4(229).
- K. Maekawa, M. Yamazaki, T. Ogiso, T. Maruyama, H. Ogura, W. Kashino, H. Koiso, M. Yamaguchi, M. Tanaka, and Y. Den. 2014. Balanced Corpus of Contemporary Written Japanese. *Language Resources and Evaluation* 48:345–371.
- T. Miyauchi, M. Asahara, N. Nakagawa, and S. Kato. 2017. Annotation of Information Structure on ‘The Balanced Corpus of Contemporary Written Japanese’. In *Proceedings of PACLING 2017*, pages 166–175.
- M. Osaka and N. Osaka. 1994. [working memory capacity related to reading: measurement with the japanese version of reading span test] (in japanese). *Shinrigaku Kenkyu: The Japanese Journal of Psychology* 65(5):339–345.
- D. Roland, G. Mauner, C. O’Meara, and H. Yun. 2012. Discourse expectations and relative clause processing. *Journal of Memory and Language* 66(3):479–508.
- T. Sorensen, S. Hohenstein, and S. Vasishth. 2016. Bayesian linear mixed models using stan: A tutorial for psychologists, linguists, and cognitive scientists. *Quantitative Methods for Psychology* 12:175–200.
- S. Uchida, E. T. Miyamoto, Y. Hirose, Y. Kobayashi, and T. Ito. 2014. An erp study of parsing and memory load in japanese sentence processing – a comparison between left-corner parsing and the dependency

locality theory -. In *Proceedings of the Thought and Language/the Mental Architecture of Processing and Learning of Language 2014*.

国立国語研究所, editor. 1964. 分類語彙表. 秀英出版.

国立国語研究所, editor. 2004. 分類語彙表 -増補改訂版-. 大日本図書.