Syntactic Constraints and Gender Congruency in Processing Chinese Compound Reflexive: Evidence from Reading Eye-Tracking

Xi Yu

School of Foreign Languages, SJTU

December 12, 2021

Abstract

In Chinese, the compound reflexive *ta-ziji* ("him/her-self") has the gender marking pronoun *ta*, hence presenting a good test case for interference effects from structurally illicit antecedents predicted by cue-based retrieval models. Using reading eye-tracking, we manipulated the gender of *ta-ziji* that (mis)matches that of two noun phrases. Results showed significance only in the main effect of the first noun phrase at the spillover region, indicating the syntactic constraints overweight the gender agreement features during anaphora resolution. Our findings are mostly consistent with the structure-favoring accounts. Besides, the significant differences in the spillover region might also reflect a lag in the syntactic processing of older adults. Thus, future work might explore the potential effects of gender or animacy information on the processing of ta-ziji.

Keywords: cue-based retrieval, gender congruency, Chinese, ta-ziji, eye-tracking

Syntactic Constraints and Gender Congruency in Processing Chinese Compound Reflexive: Evidence from Reading Eye-Tracking

In parsing sentences, one major task a reader has to accomplish is to syntactically link together two or more linguistic elements that are not adjacent to each other. For example, when a reflexive is being processed, it has to be somehow linked to its antecedent even if there is intervening material. Therefore, one central question in psycholinguistics is what mechanisms the human parser uses to identify and retrieve the previously processed part of a dependency. A widely accepted view (Lewis & Vasishth, 2005; van Dyke, 2003; van Dyke & McElree, 2011) is that a cue-based retrieval mechanism drives this dependency completion process. When a dependency needs to be completed, the cue-based retrieval account assumes that certain features (retrieval cues) are used to retrieve the co-dependent item, the retrieval target, from memory. An important consequence of such a cue-based retrieval mechanism is that whenever other items, called distractors, also match some or all of the retrieval cues, similarity-based interference can arise. A wealth of research on reflexive processing has investigated how structural constraints on antecedent retrieval interact with agreement features. Many researchers (Dillon et al., 2013; Frazier et al., 2015; Parker & Phillips, 2017) posit that structural information has priority over other cues in the retrieval process of antecedents. However, the extent to which syntactic constraints on anaphora resolution are violated during processing is contested (e.g. Cunnings & Felser, 2013; Jäger et al., 2017; Lewis & Vasishth, 2005).

To date, the majority of research that has investigated the interaction of syntactic constraints and agreement features during anaphora resolution has examined the English reflexive processing. Fewer studies have focused on reflexive processing in Chinese, and most existing research has only investigated the bare reflexives (*ziji*, “self”), not the compound reflexives (*ta-ziji*, “himself/herself”). Besides, the effects of aging on the

processing of reflexives are still unclear. The aim of the current study was to examine the interference effects in Chinese compound reflexive processing of the aging population, using reading eye-tracking. We particularly focus on the processing of gender agreement in anaphora sentences, for gender is represented at different linguistic levels (i.e., morphological, lexical, conceptual, pragmatic). Gender agreement has a major role in processing dependencies both within linguistic levels (e.g. between subject and verb) and between linguistic levels (e.g. the gender of a subject noun and the sex of its referent in a discourse model) (Kreiner et al., 2008).

Cue-based Retrieval Mechanism

Within the framework of cue-based retrieval, an important issue is the nature of the set of cues that restrict retrieval. On the one hand, it has been proposed that syntactic cues have some kind of priority over non-syntactic cues (Van Dyke & McElree, 2011). In particular, it has been proposed that for the processing of reflexive-antecedent dependencies, the set of features used for retrieving a reflexive's antecedent is limited to syntactic cues, based on Principle A of Chomsky's (1993) Binding Theory. We will refer to this proposal as a structure-favoring account. On the other hand, Lewis and Vasishth (2005) argue that all available cues are used for retrieval with equal weights being applied to all cues. We will refer to this account as the standard cue-based retrieval account.

The earliest evidence for the structure-favoring account might come from Sturt (2003), who, using materials as in (1), manipulated gender congruence between the reflexive and the matrix subject (*Jonathan/Jennifer*), with the local subject (*the surgeon*) being a stereotypical male or female.

(1) Jonathan (Jennifer) was pretty worried at the City Hospital. He/she remembered that the surgeon had pricked himself (herself) with a used syringe needle. There should be an investigation soon.

Sturt (2003) found that at the reflexive, sentences with gender-mismatching local subjects (i.e. *surgeon ... herself*) were read longer than sentences with gender-matching local subjects (i.e. *surgeon ... himself*), as reflected in first fixation duration, gaze duration, and second-pass reading time, suggesting that Principle A of Binding Theory was immediately applied in antecedent search. The interference from the gender-matching matrix subject was only observed on second-pass reading time.

In contrast, the standard cue-based retrieval account makes the following predictions: When a perfectly matched antecedent is available, its activation would be reduced in the presence of a structural illicit antecedent that matches one or more non-structural cue(s) (e.g. gender, number, or animacy) due to activation spreading from the retrieval cue to both antecedents, leading to a slowdown in retrieval latencies or increased reading times (inhibitory interference, also known as *similarity-based interference* or *cue overload*). When a perfectly matched antecedent is unavailable, that is, the structurally licit candidate mismatches a certain non-structural cue, then a candidate antecedent matching in that non-structural cue would race with the structurally licit candidate until the one with a higher activation is retrieved, resulting in a speed-up in the average retrieval latencies or decreased reading times (*facilitatory interference*) (Logačev & Vasishth, 2015).

Despite a wealth of research on English reflexive processing, few studies have focused on reflexive processing in Chinese. Reflexives in Chinese can take either bare (*ziji*, “self”) or compound (*ta-ziji*, “himself/herself”) forms, with the latter requiring its antecedent to match the pronoun ta in gender and number features (Huang et al., 2014). Previous work on reflexive processing in Chinese has mainly focused on *ziji*, with converging results verifying the theoretical claims that *ziji* can be affected by nonstructural information and be bound by a non-local antecedent. Specifically, using reading eye-tracking, Jäger et al. (2015) found an inhibitory interference from the structurally illicit antecedent (i) when *ziji* mismatched the

local antecedent in animacy (Experiment 1), and (ii) when *ziji* matched the local antecedent if an additional memory load was enforced on the participants (Experiment 2).

Given that *ta-ziji* resembles English reflexives more closely than *ziji* (Huang et al., 2014), one might assume that its online resolution process should be more likely to follow Principle A. To our knowledge, only a few experimental studies have investigated the real-time processing of *ta-ziji*. Using self-paced reading, Dillon et al. (2016) compared *ziji* with *ta-ziji* using complex sentences, where the animate feature of reflexives matched with either the local/head noun or the distant/embedded noun of a prenominal relative clause. They found that reflexives were read longer when their animacy features matched with the distant/embedded noun than with the local/head noun, and the effect size (i.e. the difference in reading times between the two types of sentences) was larger for *ziji* than *ta-ziji* in the spillover region. They took these findings as evidence for *ta-ziji* being less subject to the locality constraint than *ziji*, and attributed this difference in the parsing profile of the two reflexives to the presence of an additional retrieval cue ta in *ta-ziji* that encoded human, number and gender features.

Age-related Syntactic Processing Deficits

A great deal of the empirical research examining cognitive change with age has focused on age-related slowing, using response times (RTs) as the dependent measure, as a ubiquitous finding is that older adults on average are slower than younger adults (Baron & Cerella, 1993). Insofar as researchers have probed the consequences of aging on syntactic processing, they have concluded that as syntactic complexity increases, older adults have more difficulty recalling propositional information from sentences (Stine & Hindman, 1994) and imitating sentences (Kemper, 1986). Besides, working memory loading also represents a special problem for the elderly. Very few studies, however, have actually examined the effects of aging on the real-time processing of reflexives, which was obviously shaped by working

memory (Parker, 2019).

The Present Study

Against this background, the aim of this study was to investigate how syntactic constraints and gender congruency interact to cue antecedent retrieval for Chinese compound reflexives (*ta-ziji*, "himself/herself"). In particular, we conducted the eye-movement experiment on the aging population. Participants read texts, which manipulated gender congruence between a reflexive and two referents as in (2), while their eye movements were monitored. The noun phrase (NP) *Zhang San* was referred to as NP1, and *Li Si/Li* was referred to as NP2.

(2)

(a) *NP1 match, NP 2 match*

张三在李四的房间看到两张他自己的毕业照片。

Zhang San/ in Li Si's room/ saw/ two/ *ta-ziji*/ graduation photos.

Zhang San saw two graduation photos of himself in Li Si's room.

(b) *NP1 match, NP2 mismatch*

张三在李丽的房间看到两张他自己的毕业照片。

Zhang San/ in Li Li's room/ saw/ two/ *ta-ziji*/ graduation photos.

Zhang San saw two graduation photos of himself in Li Li's room.

(c) *NP1 mismatch, NP2 match*

张三在李丽的房间看到两张她自己的毕业照片。

Zhang San/ in Li Li's room/ saw/ two/ *ta-ziji*/ graduation photos.

Zhang San saw two graduation photos of herself in Li Li's room.

(d) *NP1 mismatch, NP2 mismatch*

张三在李四的房间看到两张她自己的毕业照片。

Zhang San/ in Li Si's room/ saw/ two/ *ta-ziji*/ graduation photos.

Zhang San saw two graduation photos of herself in Li Si's room.

In (2a-d), the nonlocal antecedent *Zhang San* is the only accessible antecedent according to binding constraints (Chomsky, 1993). In (2a,b) NP1, a typical Chinese name for males, matches with the reflexive, while it mismatches in (2c,d). The gender of NP2, a typical Chinese name for females, has also been manipulated, and matches the gender of the reflexive in (2a,c) but not (2b,d).

If syntactic constraints constitute more highly weighted cues to antecedent retrieval than gender agreement, i.e. if the structure-favoring account applies, we should observe reliable effects of the gender of NP1 only (Chow et al., 2014). In this case, reading times should be longer in (2c,d), when NP1 mismatches with the reflexive, in comparison to (2a,b), when there is a gender match. The gender of NP2 should not affect reading times. Alternatively, we might find evidence of facilitatory interference. In this case, we should observe a reliable attenuation of the NP1 gender mismatch effect when NP2 matches the gender of the reflexive, with reading times in (2c) being shorter than (2d). This would be compatible with the standard cue-based retrieval account.

Methodology

Participants and Materials

30 native Chinese speakers (aged between 55 and 65 years old) participated in the experiment. 32 sets of experimental items were constructed as in (2). All proper nouns are typical Chinese names for males (e.g. *Zhang San*) or females (e.g. *Li Li*). The critical sentence included NP1 and NP2, and the critical reflexive. In addition to the experimental items, 30 filler texts were also constructed. These included items that were structurally similar to the experimental items but did not contain reflexives.

Procedure

The experimental and filler items were pseudo-randomized such that no two experimental

items were adjacent to each other. Four presentation lists were constructed in a Latin-square design, and the experiment was divided into four blocks between which participants could take a break if required. Forward and reverse orders of items within each block were constructed, and the ordering of blocks was different for each participant.

Eye movements were recorded at a rate of 1000 Hz using the SR EyeLink 1000. Although viewing was binocular, eye movements were recorded from the right eye only. Sentence stimuli were presented at a font size of Songti 26 on a 19-inch ViewSonic G225f monitor with a display resolution of 1024x768 pixels. Subjects were seated at a comfortable chair with their head placed against a chin-rest and forehead-rest. The distance from the monitor to the chinrest was 63cm. Each character formed about a one-degree visual arc. An experimental session began with calibration of the eye-tracker on a nine-point grid. To initiate sentence display, subjects were instructed to fixate at a left-center square, in the same spot where also lay the first character of a sentence. All sentences were presented in one line. Whenever fixation failed, a re-calibration procedure was performed.

Data Analysis

We report analysis for two regions of text. The target region consisted of the target reflexive, while the spillover region contained the following three words, for example, "的毕业" in (2). We calculated four reading time measures at each region. *First fixed duration* is the duration of the first fixation within a region during first-pass reading, while *first-pass reading time* (or *gaze duration*, if the region for which the measure is computed contains a single word) is the summed duration of fixations within a region during its first inspection until it is exited to the left or right. *Regression path duration* is calculated by summing the duration of each fixation, starting with the first fixation when a region is entered from the left, up until but not including the first fixation in a region to the right. In addition, we also calculated the *total duration*, which summed all fixations made within a region.

Results

We deleted the data for which the first fixation duration is 0 or the total duration is 0. This accounted for 23.41% of the overall data. Besides, we also deleted outliers for each reading time measure at each region that was out of the range of Mean ± 2.5 SD (Standard Deviation). This accounted for 7.48% of the original data. Overall, the data we report includes 550 entries in the target region and 554 entries in the spillover region.

Analysis was conducted using linear mixed-effects models (see Baayen, 2008; Baayen, Davidson, & Bates, 2008) with the lme4 package in R. For each measure, a mixed-effects model was fit with fixed main effects of NP1 (match vs. mismatch), NP2 (match vs. mismatch) and their interactions. Subject and item random intercepts and random slopes for each fixed effect were fitted using a maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013). Significant interactions between NP1 and NP2 in two regions provide insight into the processing of *ta-ziji*.

Summaries of the reading time data are presented in Table 1. Figure 1 and Figure 2 are the visualization of the descriptive data at each region.

In regression pass duration at the spillover region, there was a significant main effect of NP1, with longer reading times when the NP1 mismatched in gender with the reflexive compared to when there was a gender match. Besides, the interaction effect between NP1 and NP2 was significant, too. A similar interaction pattern was also observed at the target region. However, none of the results were significant when we further conducted the simple main effect. No further main effects or interactions were significant in any measure.

Discussion

Reading times after the reflexive were significantly affected by the typical gender of NP1 but not the gender of NP2. In the regression path times, reading times were longer when NP1 mismatched in typical gender with the reflexive. And we did not observe any significant

effects of NP2, either in terms of facilitatory or inhibitory interference, in any reading time measure. Although in both the target region and the spillover region we saw the significant interaction between NP1 and NP2, the results of the simple main effect show no significance. This suggests that upon encountering the reflexive, participants preferentially retrieved NP1 instead of NP2. These results are compatible with Chow et al. (2014) and suggest that syntactic constraints constitute highly weighted cues to antecedent retrieval during reflexive resolution.

The perceptual span in Chinese reading extends from 1 character on the left of the fixated character to up to 4 characters on its right (Pan et al., 2017; Yan et al., 2010), depending on the frequency (Yan et al., 2010) and predictability (Zang et al., 2016) of parafoveal words. Given that ta and *ziji* are highly frequent (11853.78 words/million for masculine ta, 6600.96 words/million for feminine ta, and 1728.29 words/million for *ziji*; Cai & Brysbaert, 2010) and that these two words form natural collocations (i.e. *ta-ziji*), an early congruence effect(s) on processing *ta-ziji* might be detected when the eyes are fixated to the local verb preceding the reflexive (Chang et al., 2020). However, in our experiment on the aging population, the significant effect was only observed in the spillover region but not in the target region. We believe this was due to cognitive decline in the elderly. They responded to the target reflexive slowly and thus was only shown in the spillover region.

In sum, our experiment supported the structure-favoring accounts, demonstrating that syntactic constraints are more important cues for the processing of the compound reflexive *taziji* in Chinese in aging populations compared to gender agreement. In addition to this, our experiments also showed that the elderly are unique in reflexive processing, since such an effect appeared only in the spillover region, which is after the reflexive.

Conclusion

Our eye-tracking experiment on Chinese compound reflexive revealed a preference for

syntactic cues in the reflexive processing of the aging population, though such effects occurred only at the spillover region. Our findings are mostly consistent with the structure-favoring accounts, where the structural cues outweigh the gender cues in the resolution of *ta-ziji*. An open question for future work has to do with other non-structural retrieval cues, such as animacy/humanness information of nouns. In our experimental sentences, only proper nouns representing human names were used as candidate antecedents for *ta-zij*. But animacy/humanness is likely to be a stronger retrieval cue than gender (Dillon et al., 2016). Thus, future work might explore the potential effects of gender or animacy information on the processing of *ta-ziji*.

References

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Baron, A., & Cerella, J. (1993). Laboratory tests of the disuse account of cognitive decline. In J. Cerella, J. M. Rybash, W. Hoyer, & M. L. Commons (Eds.), *Adult information processing: Limits on loss* (pp. 175–203). Academic Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, *5*(6), e10729. https://doi.org/10.1371/journal.pone.0010729

Chang, W., Duan, Y., Qian, J., Wu, F., Jiang, X., & Zhou, X. (2020). Gender interference in processing Chinese compound reflexive: Evidence from reading eye-tracking. *Language, Cognition and Neuroscience*, *35*(10), 1355–1370. https://doi.org/10.1080/23273798.2020.1781213

Chomsky, N. (1993). *Lectures on government and binding: The Pisa Lectures* (7th ed.). Mouton de Gruyter.

Chow, W. Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00630

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes*, *28*(1–2), 188–219. https://doi.org/10.1080/01690965.2010.548391

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*(2), 85–103. https://doi.org/10.1016/j.jml.2013.04.003

Frazier, M., Ackerman, L., Baumann, P., Potter, D., & Yoshida, M. (2015). Wh-filler-gap dependency formation guides reflexive antecedent search. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01504

Huang, C. T. J., Audrey Li, Y., & Li, Y. (2014). *The syntax of Chinese*. Cambridge University Press.

JÃ¤ger, L. A., Engelmann, F., & Vasishth, S. (2015). Retrieval interference in reflexive processing: Experimental evidence from Mandarin, and computational modeling. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00617

Kemper, S., & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology and Aging*, *16*(2), 312–322. https://doi.org/10.1037/0882-7974.16.2.312

Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language*, *58*(2), 239–261. https://doi.org/10.1016/j.jml.2007.09.003

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25

Logačev, P., & Vasishth, S. (2015). A Multiple-Channel model of Task-Dependent ambiguity resolution in sentence comprehension. *Cognitive Science*, *40*(2), 266–298. https://doi.org/10.1111/cogs.12228

Pan, J., Yan, M., & Laubrock, J. (2017). Perceptual span in oral reading: The case of Chinese. *Scientific Studies of Reading*, *21*(3), 254–263. https://doi.org/10.1080/10888438.2017.1283694

Parker, D. (2019). Cue combinatorics in memory retrieval for anaphora. *Cognitive Science*, *43*(3). https://doi.org/10.1111/cogs.12715

Parker, D., & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language*, *94*, 272–290. https://doi.org/10.1016/j.jml.2017.01.002

Stine, E. A. L., & Hindman, J. (1994). Age differences in reading time allocation for propositionally dense sentences. *Aging, Neuropsychology, and Cognition*, *1*(1), 2–16. https://doi.org/10.1080/09289919408251446

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*(3), 542–562. https://doi.org/10.1016/s0749-596x(02)00536-3

van Dyke, J. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of*

Memory and Language, *49*(3), 285–316. https://doi.org/10.1016/s0749-596x(03)00081-0

van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*(3), 247–263. https://doi.org/10.1016/j.jml.2011.05.002

Wingfield, A., Stine, E. A., Lahar, C. J., & Aberdeen, J. S. (1988). Does the capacity of working memory change with age? *Experimental Aging Research*, *14*(2), 103–107. https://doi.org/10.1080/03610738808259731

Yan, M., Kliegl, R., Shu, H., Pan, J., & Zhou, X. (2010). Parafoveal load of word N+1 modulates preprocessing effectiveness of word N+2 in Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1669–1676. https://doi.org/10.1037/a0019329

Zang, C., Wang, Y., Bai, X., Yan, G., Drieghe, D., & Liversedge, S. P. (2016). The use of probabilistic lexicality cues for word segmentation in Chinese reading. *Quarterly Journal of Experimental Psychology*, *69*(3), 548–560. https://doi.org/10.1080/17470218.2015.1061030

Table 1

Four Reading Times at the Target Region and the Spillover Region

	First Fixation Duration	First Pass Reading Time	Regression Path Duration	Total Duration
Target Region				
NP1 Match, NP2 Match	255	460.3	697.6	460.3
NP1 Match, NP2 Mismatch	246.9	420	623	420
NP1 Mismatch, NP2 match	247	435.4	571.7	435.4
NP1 Mismatch, NP2 Mismatch	254.7	450.8	714.9	450.8
Spillover Region				
NP1 Match, NP2 Match	279.2	493.6	3735	734.3
NP1 Match, NP2 Mismatch	270.3	449.5	3189	654.3
NP1 Mismatch, NP2 match	274	472.6	3857	674.3
NP1 Mismatch, NP2 Mismatch	274.9	510.2	4487	727.6

Figure 1

Box Plot of the Four Reading Time Measures at the Target Region

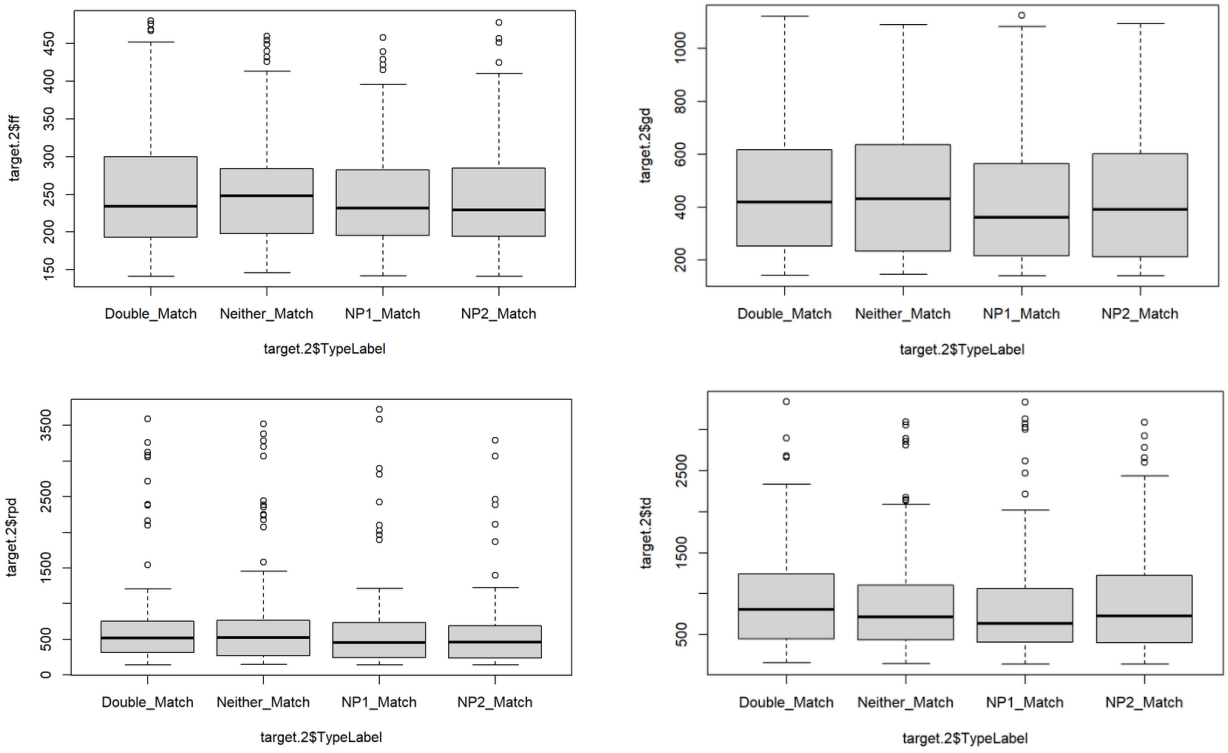

Figure 2

Box Plot of the Four Reading Time Measures at the Spillover Region