A Non-local Attachment Preference in the Production and Comprehension of Thai Relative Clauses

Teeranoot Siriwittayakorn
Department of Linguistics, Faculty of Arts,
Chulalongkorn University,
Phayathai Road, Pathumwan,
Bangkok, 10330, Thailand
steeranoot@gmail.com

Edson T. Miyamoto
University of Tsukuba, Graduate School of
Humanities and Social Sciences, Tsukuba,
Ibaraki, 305-8571, Japan
miyamoto@alum.mit.edu

Theeraporn Ratitamkul
Department of Linguistics, Faculty of Arts,
Chulalongkorn University,
Phayathai Road, Pathumwan,
Bangkok, 10330, Thailand
Theeraporn.R@chula.ac.th

Heeyoun Cho
College English Program, Faculty of Liberal
Education, Seoul National University,
Seoul, South Korea
heeyoun.cho@gmail.com

Abstract

In parsing, a phrase is more likely to be associated with an adjacent word than to a non-adjacent one. Instances of adjacency violation pose a challenge to researchers but also an opportunity to better understand how people process sentences and to improve parsing algorithms by, for example, suggesting new features that can be used in machine learning. We report corpus counts and reading-time data for Thai to investigate an adjacency violation that has been reported in other languages for ambiguous relative clauses that can be attached to either of two nouns, namely, the local noun (which is adjacent to the relative clause) or the non-local noun (which is farther from the relative clause). The results indicate that, unlike English, Thai violates adjacency by favoring non-local attachment even though the two languages share many grammatical features that have been linked to a local-attachment preference (e.g., rigid SVO word order). We re-interpret previous proposals to suggest that a language favors the non-local noun if it passes at least one of two tests. (1) Modifiers can intervene between noun and relative clause. (2) Adverbs can intervene between transitive verb and direct object.

1 Introduction

We investigated the role of *locality* (or *proximity)* in processing decisions by comparing two languages (Thai and English) that have evolved largely independently but share grammatical features that have been claimed to be crucial in sentence comprehension.

A preference to associate words locally has been reported at least since the 1970s (Kimball, 1973; Gibson, 1998; *inter alia*). For example, in (1), the underlined relative clause (RC) can be attached to the non-local noun (N1, *daughter*) or to the local noun (N2, *colonel*).

(1) The journalist interviewed the daughter of the colonel <u>who had the accident</u>.

English readers prefer the RC to modify N2, whereas N1 is preferred in the corresponding construction in Spanish (Cuetos and Mitchell, 1988). Various typological differences have been used to predict which languages violate locality by favoring N1 in such *complex NP*s (i.e., N1 of N2 RC).

(2) A language *L* favors N1 attachment if:

a. L has no alternative construction for expressing the N1 interpretation (Frazier and Clifton, 1996);
b. L has flexible word order (Gibson et al., 1996);
c. L allows constituents (e.g., adverbs) to intervene between a verb and its direct object (Miyamoto, 1999);
d. L exhibits consistent use of relative pronouns (Hemforth et al., 2000);
e. L has pseudo-RCs (Grillo, 2012);
f. L allows constituents (e.g., adjectives) to intervene between the modified noun and the RC (schematically: *N adjective RC*, the *modifier-straddling hypothesis*, MSH, Cuetos and Mitchell, 1988).

All those competing proposals correctly predict that English does not violate locality as it favors N2. Thai is similar to English in a number of aspects. Word order is the same in the target construction (N1 of N2 RC) and a complementizer comparable to *that* (*thî:*) can be used as RC marker (there are two other RC markers, but *thî:* is the most frequent and has relatively few stylistic restrictions; Iwasaki and Ingkaphirom, 2009). The following properties are particularly relevant in the discussion on RC attachment.

(3)
a. Thai has at least two alternative unambiguous constructions to modify N1, namely, an RC-preposing construction (N1 RC of N2) and a compound-like structure (N1 N2 RC) resulting from the omission of the preposition.
b. Thai is a rigid SVO language, in particular, verb and direct object have to be adjacent.
c. The RC marker *thî:* has been claimed to be omissible in some environments (Iwasaki and Ingkaphirom, 2009; Kullavanijaya, 2010).
d. Pseudo relative clauses are not available in Thai.

The features in (3) together with the proposals in (2a-e) predict Thai to pattern with English in the comprehension of (1), thus resulting in a preference for N2 attachment.

In contrast, according to the MSH (see (2f)), if a language allows the sequence *N adjective RC*, the adjective can be generalized to other types of modifiers (e.g., *of N2*), hence weakening the adjacency bias and increasing the likelihood that the RC will skip the intervening modifier and attach to N1 (Cuetos and Mitchell, 1988; see the general discussion on some possible counter-examples). Unlike English, adjectives are postnominal in Thai and can intervene between the noun and the RC. This should lead Thai readers to favor N1 according to the MSH. Therefore, the goal of this paper is to test the MSH against the proposals in (2a-e), which predict Thai to be an N2-attachment language.

We report a corpus count and a self-paced reading experiment confirming the predictions of the MSH for *thî:*-marked RCs in Thai.

2 Corpus Count

A corpus count was conducted to determine production preferences in RC attachment in Thai taking the influence of context into consideration.

Since there are no plural markers or morphological agreement in Thai, ambiguity resolution is often based on plausibility. For this reason, surrounding context plays an important role in attachment. Although previous corpus counts on this topic have not included context as a factor, some studies have suggested that the matrix clause can favor N1 (e.g., by making the RC informative, Frazier 1990; increasing text coherence, Rohde et al., 2011; allowing for an alternative interpretation, see *pseudo RC*s in Grillo 2012; also Desmet et al., 2002b, on the matrix clause increasing the N1 preference in a norming questionnaire). Therefore, in order to measure the influence of the context surrounding the *complex NP*, tokens were classified according to whether information inside the complex NP was enough to determine attachment (*internally disambiguated*; e.g., *voice of men that was uttered)* or whether it was also necessary to consult the context surrounding the complex NP (*externally disambiguated*).

Moreover, it might be the case that together with context, other factors could affect attachment. One such a factor is the position of the *disambiguating context* (i.e., the information that indicates the attachment intended for the RC). Complex NPs are usually embedded in a larger context and the disambiguating context can come either before or after the complex NP. When the disambiguating context comes before the complex NP, N1 attachment might be favored in order to increase

text coherence. However, N1 bias might be weaker when context comes after the NP.

Another possible factor in attachment is the syntactic position of the target NP (subject or object). If, for example, the context provided by the preceding clause, I_{n-1}, has already given sufficient information about the subject of clause I_n, further subject modification (i.e., N1) of the clause I_n, might be unnecessary. Although the same reasoning can be applied to an object NP, because a subject tends to be a discourse-old entity (see Mattausch, 2011 on related discussion), it is predicted that the plausibility for the preceding context to be related to a subject is higher than that of an object. Therefore, the rate of attaching an RC to N1, in the subject position might be lower than that in the object position.

In sum, instances of complex NP were classified according to the following three factors.

- point of disambiguation (internally or externally-disambiguated)
- syntactic position of the complex NP (subject or object position)
- for externally-disambiguated items, point of disambiguation was further classified according to the position of the disambiguating context (early or late; i.e., before or after the complex NP).

2.1 Method

Segments with *thî*: preceded by *khɔ̌:ŋ* "of" within a three-word window were extracted from the six writing genres of the Thai National Corpus (Aroonmanakun et al., 2009), namely fiction (which contains 7,469,530 words), newspaper (5,029,019 words), academic text (8,894,650 words), non-academic text (5,342,092 words), law (1,190,516 words) and miscellanea (4,000,160 words).

Out of 23,726 sequences found, 4,800 instances (800 instances per genre) were randomly selected

and manually analyzed, and irrelevant cases discarded (e.g., if *thî*: was not used as an RC marker). From the 2,462 instances of *N1 of N2 RC* found, 356 instances (14.46%) were eliminated because the attachment site was not clear. Instances were also eliminated if the head nouns were not common nouns (481 instances, 19.54%, with proper names or pronouns, which are usually avoided in behavioral experiments) or were likely to attract the RC (308 instances, 12.51%; e.g., *khon* 'person' or *sìŋ* 'thing', see Wasow et al., 2011, for related discussion). The remaining 1,317 tokens were analyzed according to attachment.

Three native Thai speakers coded the sentences independently and disagreements (less than 5%) were settled after discussion.

2.2 Results

N1 attachments were more frequent than N2 attachments (χ^2 (1) = 42.3, p < .0001; see Table 1). The results held regardless of whether the complex NP was in subject or object position (subject position only: χ^2 (1) = 11.06, p < .001; object only: χ^2 (1) = 30.98, p < .0001).

To factor out the influence of the surrounding context, further analyses were conducted on the internally-disambiguated items. Attachments were more frequent to N1 than to N2 in all cases (overall: χ^2 (1) = 20.92, p < .0001; subject: χ^2 (1) = 6.8, p = .009; object: χ^2 (1) = 14.12, p < .001).

Analyses on externally-disambiguated items showed that when the disambiguating context came before the complex NP, the RC was more frequently attached to N1 than to N2 (χ^2 (1) = 51.58, p <.0001). The trend was the same when restricted to NPs in object position (χ^2 (1) = 47.26, p <.0001), and was marginally so for subject-position NPs (χ^2 (1) = 3.28, p = 0.07). Further analyses indicated such early contexts tended to favor N1. In the overall results (column *overall* in Table 1), the N1 bias went up

	Syntactic position		Overall
Point of disambiguation:	Subject	Object	
Internally-disambiguated	158 (58.09%)	518 (56.24%)	676 (56.67%)
Externally-disambiguated: early context	9 (81.82%)	74 (88.10%)	83 (87.37%)
Externally-disambiguated: late context	9 (81.82%)	9 (50.00%)	18 (62.07%)
Overall	176 (59.86%)	601 (58.75%)	777 (59.00%)

Table 1. Corpus frequency of N1 attachment according to point of disambiguation (internal or external), syntactic position (subject of object) and position of disambiguating context (early of late).

from 56.67% in the internally-disambiguated row to 87.37% in the row for externally-disambiguated items with early context (χ^2 (1) = 33.02, p <.0001). The trend was similar for the object-position NPs (from 56.24% to 88.10%, χ^2 (1) = 30.96, p <.0001), but it was not statistically reliable for subjects.

When context came after the complex NP, the frequencies of N1 and N2 attachments were not statistically different. There was only a marginal trend towards N1 attachment in subject position (χ^2 (1) = 3.28, p = 0.07).

Although there were few instances of N2 attachment among the externally-disambiguated tokens (overall: 23 tokens, subject: 4 tokens, object: 19 tokens), the results suggest that context can favor N2 attachment as well.

2.3 Discussion

There was a consistent preference for N1 attachment regardless of the different types of classifications used. Even after eliminating the influence of context, N1 attachment in both subject and object positions remains more frequent in Thai.

No previously-proposed grammatical factor except for the MSH (Cuetos & Mitchell, 1988) can explain the overall advantage for N1 attachment.

Some studies have suggested that animacy and concreteness can affect RC attachment (Desmet et al., 2002a; Desmet et al., 2006). However, more detailed analyses of the data suggest that they are not determining factors in Thai as there was a bias towards N1 attachment regardless of animacy and concreteness of the two nouns (see Appendix A).

3 Experiment

A reading-time experiment was conducted to investigate the on-line comprehension of RCs in Thai.

3.1 Method

Participants: Fifty-two native Thai speakers, undergraduate students at Chulalongkorn University, participated in the experiment for course credit. Since English is a compulsory subject in Thailand, the participants here and elsewhere in this paper are likely to have learnt it as a second language.

Stimuli: There were 112 test items divided into four types (28 items for each type) that varied according to the animacy of the nouns N1 and N2,

that the RC could modify (only concrete nouns were used for N1 and N2). Although care was taken to control for various factors, items were excluded from the analyses because of a number of confounding factors (e.g., plausibility of the interpretations, frequency of the words in the RCs). Therefore, we will report results for a subset of 20 items in which both nouns are animate. Each item had two versions (i.e., N1-attachment and N2-attachment versions). See (4) for an example pair.

(4)
a. N1 attachment
khunphô: fà:k khɔ̌:ŋ hâj | khunkhru: khɔ̌:ŋ
father leave thing give| teacher of
lû:kcha:j | thî: |sɔ̌:n wíʔcha: pha:sǎ:thaj
son | that |teach subject Thai language
"The father left something for the teacher of his son that teaches Thai."

b. N2 attachment
khunphô: fà:k khɔ̌:ŋ hâj | khunkhru: khɔ̌:ŋ
father leave thing give| teacher of
lû:kcha:j | thî: | sɔ̀:ptòk wíʔcha: pha:sǎ:thaj
son | that |fail subject Thai language
"The father left something for the teacher of his son that failed a Thai exam."

Because Thai lacks agreement morphology, attachment was disambiguated based on plausibility (e.g., in (4b), a student is more likely to fail an exam compared to a teacher). To avoid possible differences related to extraction position, all RCs were subject extracted (see Grodner and Gibson, 2005, and references therein for a discussion on English).

Norming: The test items were disambiguated based on plausibility. Therefore, a questionnaire was conducted to ensure that the plausibility manipulations were effective. This type of supplementary questionnaire is commonly used to verify the items used in the main experiment. For example, to make sure that the two interpretations in (5a) are equally natural, the two sentences in (5b, c) are compared in a questionnaire (example adapted from Desmet et al., 2002b).

(5)
a. The police interrogate the advisor of the politician who speaks with a soft voice.
b. The assistant has a soft voice.
c. The politician has a soft voice.

Note that RCs are usually not used in (5b, c) since we are only interested in the plausibility of the interpretations (e.g., how natural it is for an assistant or a politician to have a soft voice; but see Desmet et al., 2002b, who used RCs instead, thus potentially confounding plausibility with attachment preference).

Because the matrix clause can affect RC attachment, it was included as a separate sentence (see Desmet et al., 2002b, for questionnaires with and without the matrix clause). For each item pair in the main study, four versions were created in a 2 by 2 design (noun: N1 or N2; plausibility: plausible or implausible). The examples in (6) are the four versions created for the item pair in (4).

(6)
a. N1-plausible
khunphɔ̂: fà:k khɔ̌:ŋ hâj khunkhru: khɔ̌:ŋ
father   leave thing give teacher   of
lû:kcha:j | khunkhru sɔ̌:n wíʔcha: pha:sǎ:thaj
son | teacher teach subject Thai
 "The father left something for the teacher of
 his son. The teacher teaches Thai."

b. N1-implausible
khunphɔ̂: fà:k khɔ̌:ŋ hâj khunkhru: khɔ̌:ŋ
father   leave thing give teacher   of
lû:kcha:j | khunkhru sɔ̀:ptòk wíʔcha: pha:sǎ:thaj
son | teacher failed subject Thai
"The father left something for the teacher of
his son. The teacher failed a Thai exam."

c. N2-plausible
khunphɔ̂: fà:k khɔ̌:ŋ hâj khunkhru: khɔ̌:ŋ
father   leave thing give teacher   of
lû:kcha:j | lû:kcha:j sɔ̀:ptòk wíʔcha: pha:sǎ:thaj
son | son   fail   subject Thai
"The father left something for the teacher of
his son. The son failed a Thai exam."

d. N2-implausible
khunphɔ̂: fà:k khɔ̌:ŋ hâj khunkhru: khɔ̌:ŋ
father   leave thing give teacher   of
lû:kcha:j | lû:kcha:j sɔ̌:n wíʔcha: pha:sǎ:thaj
son | son   teach subject Thai
 "The father left something for the teacher of
 his son. The son teaches Thai."

As customary in Thai writing, spaces were used between sentences (indicated with vertical bars in (6)) but not between words.

By comparing (6a) and (6c) we can guarantee that the intended attachments were equally plausible.

By comparing (6b) and (6d), we can determine whether the unintended interpretations were equally implausible and thus equally unlikely to interfere by competing with the intended interpretations. A new group of 76 native Thai students at Chulalongkorn University who did not participate in the main experiment rated sentences on a five-point scale (1 implausible, 5 plausible).

The results for the plausible attachments (mean 4.26; median 5) and for the implausible attachments (mean 1.91; median 1) suggest that the overall plausibility manipulation worked as planned for the 20 items reported in the main study.

More importantly, according to an ordinal logistic regression analysis (Agresti, 2002), there was no difference when attachment site (N1 or N2) was included as a factor as the two plausible conditions (6a) (mean 4.37, median 5) and (6c) (mean 4.15, median 5) were equally plausible, and the two implausible conditions (6b) (mean 2.04, median 1) and (6d) (mean 1.78, median 1) were equally implausible (all p's > .25).

Procedure: Each participant in the main experiment saw a list of 112 test items following a Latin Square design so that only one version from each pair was included. Test items were shown in random order interspersed with 195 fillers. Fillers included sentences with *thî:* not followed by an RC, *N1 of N2* sequences (not followed by an RC), a single noun followed by an RC, and a variety of unambiguous sentences with one or two clauses. To make sure that participants were reading carefully, half of the test items and two-fifth of the fillers (78 items) were followed by a comprehension question.

Test sentences were segmented into four regions as indicated by the vertical bars in (4) and shown using a non-cumulative self-paced reading presentation on E-Prime 2.0. Most sentences were too long to fit on a single line, therefore all items were presented with a line break after the second region (i.e., after *N1 of N2* sequence) (previous results indicate that a pause between N2 and the RC marker is not associated with an N1 preference, e.g., Clahsen and Felser, 2006). The test session was divided into three sub-sessions with optional breaks in-between and lasted for about an hour.

Analyses: For the first three regions, analyses are reported with length-residualized reading times (based on a linear regression including all test items and fillers; Ferreira and Clifton, 1986). Data points beyond four standard deviations from condition-

region means were removed, affecting less than 1% of the test data (trends in the untrimmed results were similar to those with trimmed data).

Because the RCs (the critical region) differed in their words and plausibility biases, the reading times to the RC region were regressed against RC length, the judgments for the plausible and implausible conditions in the norming study, and the log-frequencies of words and bigrams obtained from the Thai National Corpus (Aroonmanakun et al., 2009). Residuals from this linear regression were trimmed in the same way as the whole data set (with less than 1% eliminated).

Reading times were analyzed with mixed-effects models using the lme4 package (Baayen et al., 2008, and references therein) on R (R Core Team, 2013). Wald chi-square was used to calculate p-values (function Anova in the package car; Fox and Weisberg, 2011). Pairwise comparisons with Tukey-adjusted p-values are reported (function lsmeans in the package lsmeans: Lenth, 2013).

3.2 Results

Comprehension accuracy of all test items and fillers was 96.70%. All participants scored over 88%, suggesting that they were paying attention during the experiment and therefore none of them was eliminated from further analyses. For the 20 animate-animate test items, response accuracy did not differ for the two types of attachment (N1: 96.54%; N2: 97.31%; mixed-model including random intercepts for subjects and items: $z < 1$).

Reading times: The mixed model included attachment as fixed factor and as random slope for participants and for items. To decrease correlation between the predictors in the model, a simple contrast-coding scheme was used for each categorical variable by comparing each level to the reference level and setting the intercept as the grand mean.

In region 1, N1 attachment was faster than N2 attachment (p=.015), but the difference was unexpected since attachment was not manipulated at this point, and it may have been caused by participants sometimes resting at the beginning of a new sentence. There were no differences in the next two regions (p's>.15). In the critical region (region 4), the RC was read faster when attached to N1 than to N2 (residualized reading times: $\chi^2 (1) = 4.166$, p = .0412).

3.3 Discussion

The results showed that when the two nouns were animate, N1 attachment was preferred. However, this advantage for the non-local noun should be interpreted with caution for two reasons. First, although RC reading times were residualized against corpus frequencies, the corpus interface restricted the searches in a number of ways (e.g., some words were more likely to be prefixes; e.g., *khwa:m,* an adjective nominalizer).

Second, sentences were presented with a line break between N2 and the RC marker, potentially enhancing the perception of a pause, and decreasing the adjacency advantage for N2. Such an effect would be compatible with the *implicit prosody hypothesis* (Fodor, 1998; but see Clahsen and Felser, 2006; also, English readers prefer N2 attachment even with a break after N2, Felser et al., 2003).

The reading-time advantage for N1 is partially in line with the corpus counts. Only concrete nouns were used for N1 and N2 in the test items of the reading experiment. In the corpus, although N1 attachments were more frequent than N2 attachment overall, the advantage for N1 was not reliable when both nouns were concrete and animate (see Appendix A for the breakdown by animacy and concreteness) but perhaps a coarse-grained count is used (e.g., collapsing across different animacy and concreteness patterns; but see Desmet et al., 2006, on the need for fine-grained counts). Clearly, further results are needed to address this point in more detail.

4 General Discussion

Both production and comprehension data indicate that, unlike English, there is a preference for N1 attachment in Thai. The corpus study also suggested that N1 bias was present in both subject and object position, and context tended to favor N1 when it preceded the RC.

The modifier-straddling hypothesis (MSH; Cuetos & Mitchell, 1988) is the only grammatical factor that correctly predicts the N1-attachment preference for Thai observed in the corpus and in the animate-animate condition of the reading experiment. However, the MSH cannot explain the non-local preference reported for languages in which modifiers do not intervene between noun and RC (e.g., Dutch: Brysbaert and Mitchell, 1996; German: Hemforth et al., 2000).

One solution is to extend the notion of modifier in the MSH to include both adjectives and adverbs. Therefore, we propose a *generalized MSH* that includes a second factor namely adverb intervention, as mentioned in (2c).

(7) Generalized Modifier-Straddling Hypothesis (GMSH). A language favors N1 attachment if at least one of the following two triggers is set.
- Trigger 1. Modifiers (e.g., adjectives) can intervene between head noun and RC (Cuetos and Mitchell, 1988).
- Trigger 2. Adverbs can intervene between transitive verb and direct object (Miyamoto, 1999).

The first trigger is directly related to RCs. The second trigger is related to previous observations that (i) verb-object clusters tend to have a closer relation than verb-subject ones across a variety of typologically-distinct languages (Tomlin, 1986) and (ii) whether a language allows adverbs to intervene between verb and object has been associated with a number of word-order properties (Pollock, 1989).

According to the GMSH, there are roughly four types of languages. English is among the most restrictive and has neither trigger. Thai has only trigger 1. Dutch and German have the second but not the first. The most lenient languages such as Romance languages have both triggers. The last three types of languages should all favor N1 attachment. It is not clear whether the triggers necessarily entail gradient preferences (e.g., the N1 preference is stronger with both triggers than with just one), but this would be a natural prediction that could be pursued in the future. It is possible that the triggers are just tests, convenient ways of checking for properties (e.g. RC attachment) that cluster together.

Another question that needs to be addressed in the future is whether the GMSH affects attachment preferences directly by dictating parsing decisions during comprehension, or whether it affects attachment preference indirectly by dictating production processes (hence, frequency of use), which in turn affect expectation during comprehension as in *exposure-based accounts* (Desmet et al., 2006; Kamide, 2012; MacDonald and Christiansen, 2002; Mitchell et al., 1995; *inter alia*).

The GMSH can be further tested in a number of ways. It makes predictions about individual differences in that speakers who tend to accept the two triggers in (7) are more likely to attach RCs to N1 than to N2. It also suggests that the triggers can be incorporated as features in machine learning in order to better predict RC attachment in the target language.

4.1 Cross-linguistic Variation

A crucial theme in the research of RC attachment has been the observation that preferences vary across languages. However, the corpus count suggests that surrounding context can play a role in RC attachment and the bias is often but not exclusively to N1. Therefore, it is difficult to ascertain how much of the differences observed across various languages are cross-linguistic variations in the way native speakers parse RCs rather than differences in the contexts that were used in the previous studies. This is particularly true for corpus counts because it is unclear how context affected attachment in previous results.

But the observation may also apply to previous behavioral results. Although Desmet et al. (2002b and references therein) reported that surrounding context did not affect online processing, previous results may have been affected by subtle differences in the materials used. For example, although similar sentences were used in the original study comparing English and Spanish (Cuetos and Mitchell, 1988), closer inspection suggests that many English RCs used the simple past (*was*), whereas the Spanish translations used two forms (the preterit *estuvo* or the imperfect *estaba*). This may have caused the N1 preference in Spanish to look stronger than it actually is. The imperfect does not include the start or end points of the event and tends to be more natural when accompanied by a time reference (see Zagona, 2012, for relevant discussion). The matrix event can provide a time reference especially when the RC is attached to N1, which as an argument of the matrix verb makes the connection between the two events clearer.

5 Conclusion

We reported corpus and reading time data indicating that N1-attachment is favored in Thai. We proposed a generalized version of the MSH in which intervening constituents can increase the preference

for associating an RC to a non-local head. The proposal can account for a range of cross-linguistic data. Cross-linguistic variation in RC attachment requires more careful studies given the possible influence of contexts used in previous results.

Acknowledgments

References

Agresti, Alan. (2002). *Categorical Data Analysis* (2 ed.). New York, NY: Wiley.

Aroonmanakun, Wirote, Tansiri, Kachen, & Nittayanuparp, Pairit. (2009, 6-7 August). *Thai National Corpus: A progress report.* Paper presented at the 7th Workshop on Asian Language Resources, ACL-IJCNLP 2009, Suntec, Singapore.

Baayen, Rolf Harald, Davidson, Doug J., & Bates, Douglas M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Brysbaert, Marc, & Mitchell, Don C. (1996). Modifier Attachment in Sentence Parsing: Evidence from Dutch. *The Quarterly Journal of Experimental Psychology Section A, 49*(3), 664-695.

Clahsen, Harald, & Felser, Claudia. (2006). Continuity and shallow structures in languge processing. *Applied Psycholinguistics, 27*, 107-126. doi: 10.1017.S0142716406060206

Cuetos, Fernando, & Mitchell, Don C. (1988). Cross-linguistic differences in parsing: Restriction on the use of the Late Closure strategy in Spanish. *Cognition, 30*, 73-105.

Desmet, Timothy, De Baecke, Constantign, Drieghe, Denis, Brysbaert, Marc, & Vonk, Wietske. (2006). Relative clause attachment in Dutch: On-line comprehension corresponds to corpus frequencies when lexical variables are taken into account. *Language and Cognitve Processes, 21*(4), 453-485. doi: 10.1080/01690960400023485

Desmet, Timothy, Brysbaert, Marc, & De Baecke, Constantijn. (2002a). The correspondence between sentence production and corpus frequencies in modifier attachment. *Quarterly Journal of Experimental Psychology, 55A*(3), 879-896. doi: 10.1080/20724980143000604

Desmet, Timothy, De Baecke, Constantijn, & Brysbaert, Marc. (2002b). The influence of referential discourse context on modifier attachment in Dutch. *Memory & Cognition, 30*(1), 150-157.

Felser, Claudia, Roberts, Leah, Marinis, Theodore, & Gross, Rebecca. (2003). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics, 24*, 453-489.

Ferreira, Fernanda, & Jr., Charles Clifton. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*(3), 348–368.

Fodor, Janet Dean. (1998). Learning to parse? *journal of Psycholinguistic Research, 27*, 285-317.

Fox, John. & Weisberg, Sanford. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage.

Frazier, Lyn. (1990). Parsing modifiers: Special purpose routines in the human sentence processing mechanism. In D. A. Balota, G. B. F. d'Arcais & K. Rayner (Eds.), *Comprehension process in reading* (pp. 303-331). Hillsdale, NJ: Erlbaum.

Frazier, Lyn, & Clifton, Charles. (1996). *Construal.* Hong Kong: The MIT Press.

Gibson, Edward. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1-76.

Gibson, Edward, Schütze, Carson T., & Salomon, Ariel. (1996). The relationship between the frequency and the processing complexity of linguistic structure. *Journal of Psycholinguistic Research, 25*(1), 59-92.

Grillo, Nino. (2012). Local and Universal. In V. Bianchi & C. Chesi (Eds.), *Enjoy Linguistics. Papers offered to Luigi Rizzi on the occasion of his 60th birthday*: CISCL Press.

Grodner, Daniel J., & Gibson, Edward. (2005). Consequences of the Serial Nature of Linguistic Input for Sentential Complexity. *Cognitive Science, 29*, 261–291.

Hemforth, Barbara, Konieczny, Lars, & Scheepers, Christoph. (2000). Syntactic Attachment and Anaphor Resolution: The Two Sides of Relative Clause Attachment. In M. W. Crocker, M. Pickering & J. Charles Clifton (Eds.), *Architectures and Machanisms for Language Processing* (pp. 259-281). Cambridge, United Kingdom: Cambridge University Press.

Iwasaki, Shoichi, & Ingkaphirom, Preeya. (2009). *A reference grammar of Thai*. Cambridge: Cambridge University Press.

Kamide, Yuki. (2012). Learning individual talkers' structural preferences. *Cognition, 124*, 66-71. doi: 10.1016/j.cognition.2012.03.001

Kimball, John. (1973). Seven principles of surface structure parsing in natural language. *Cognition, 2*, 15-47.

Kullavanijaya, Pranee. (2010). อนุประโยคขยายนาม :คุณานุประโยค และอนุประโยคเติมเต็มนาม [Clauses modifying nouns: Relative clauses and complement clauses] .In A. Prasithrathsint (Ed.), *Controversial constructions in Thai grammar: Relative clause constructions, complement clause constructions, serial verb constructions, and passive constructions* (pp. 7-64). Bangkok: Chulalongkorn University Press.

Lenth, Russell. V. (2013). lsmeans: Least-squares means (1.10-2). R package.

MacDonald, Maryellen C., & Christiansen, Morten H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review, 109*, 35–54. doi: 10.1037//0033-295X.109.1.35

Mattausch, Jason. (2011). A note on the emergence of subject salience. In A. Benz & J. Mattausch (Eds.), *Bidirectional Optimality Theory* (Vol. 279, pp. 73–96): John Benjamins Publishing Company.

Mitchell, Don C., Cuetos, Fernando, Corley, Martin M. B., & Brysbaert, Marc. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research, 24*(6), 469-488. doi: 10.1007/BF02143162

Miyamoto, Edson T. (1999). *Relative clause processing in Brazilian Portuguese and Japanese* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Pollock, Jean-Yves. (1989). Verb Movement, Universal Grammar, and the Structure of IP. *Linguistic Inquiry, 20*(3), 365-424.

R Core Team (2013). R: A language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing.

Rohde, Hannah, Levy, Roger, & Kehler, Andrew. (2011). Anticipating explanations in relative clause processing. *Cognition, 118*, 339-358. doi: 10.1016/j.cognition.2010.10.016

Tomlin, Russell S. (1986). *Basic Word Order: Functional Principles*. London: Croom Helm.

Wasow, Thomas, Jaeger, T. Florian, & Orr, David M. (2011). Lexical Variation in Relativizer Frequency. In H. J. Simon & H. Wiese (Eds.), *Expecting the unexpected: Exceptions in Grammar* (pp. 175-195). Germany: Mouton de Gruyter.

Zagona, Karen. (2012). Tense and aspect. In J. I. Hualde, A. Olarrea, & E. O'Rourke (Eds.), *The handbook of Hispanic linguistics* (pp. 355-372). Chichester, UK: Wiley Blackwell.

Appendix A.
See Tables 2 to 4 for the corpus frequencies according to animacy and concreteness.

		Types of N2				Total
		animate		inanimate		
		concrete (%)	abstract (%)	concrete (%)	abstract (%)	
animate	concrete	13 (54.17)	11 (47.83)*	7 (53.85)	0 (0.00)	31 (51.67)
	abstract	0 (0.00)	44 (95.65)*	1 (50.00)	1 (100.00)	46 (86.79)*
inanimate	concrete	54 (50.94)	25 (92.59)*	57 (52.78)	9 (56.25)	145 (56.42)*
	abstract	157 (59.70)*	107 (70.39)*	120 (46.33)	70 (46.98)	454 (55.16)*
Total		224 (56.42)*	187 (75.40)*	185 (48.43)	80 (48.19)	676 (56.67)*

Table 2. N1 attachments in internally-disambiguated sentences. (N1 bias in% each cell: *: p <% .05; +: p < .10 according to exact binomial texts).

		Types of N2				Total
		animate		inanimate		
		concrete (%)	abstract (%)	concrete (%)	abstract (%)	
animate	concrete	2 (40.00)	5 (50.00)	1 (33.33)	0 (0.00)	8 (44.44)
	abstract	0 (0.00)	6 (100.00)*	0 (0.00)	0 (0.00)	6 (75.00)
inanimate	concrete	17 (70.83)+	4 (80.00)	16 (55.17)	3 (60.00)	40 (63.49)*
	abstract	30 (68.18)*	20 (71.43)*	28 (44.44)	26 (54.17)	104 (56.83)+
Total		49 (66.22)*	35 (71.43)*	45 (46.88)	29 (54.72)	158 (58.09)*

Table 3. N1 attachments in subject position in internally-disambiguated sentences. (N1 bias in% each cell: *: p <% .05; +: p < .10 according to exact binomial texts).

		Types of N2				Total
		animate		inanimate		
		concrete (%)	abstract (%)	concrete (%)	abstract (%)	
animate	concrete	11 (57.89)	6 (46.15)	6 (60.00)	0 (0.00)	23 (54.76)
	abstract	0 (0.00)	38 (95.00)*	1 (100.00)	1 (100.00)	40 (88.89)*
inanimate	concrete	37 (45.12)	21 (95.45)*	41 (51.90)	6 (54.55)	105 (54.12)
	abstract	127 (58.00)*	87 (70.16)*	92 (46.94)	44 (43.56)	350 (54.69)*
Total		175 (54.18)	152 (76.38)*	140 (48.95)	51 (45.13)	518 (56.24)*

Table 4. N1 attachments in object position in internally-disambiguated sentences. (N1 bias in% each cell: *: p <% .05; +: p < .10 according to exact binomial texts).