A computational view into the structure of attachment ambiguities in Chinese and Korean*

So Young Lee & Aniello De Santo

Miami University, University of Utah

1. Introduction

In the generative syntactic literature, it is common to be presented with multiple, contrasting analyses of the same phenomena, often differing in terms of subtle, fine-grained details. Of course, it is crucial to evaluate such analyses in terms of their constantly with broader theoretical commitments, and coverage (and predictions) over broader sets of phenomena. However, if one thinks that adequate syntactic representations should also help in explaining linguistic behavior then it should be possible to approach and evaluate linguistic constructions that are more theoretically opaque in terms of their ability to explain experimental profiles–for example, in terms of off-line sentence processing results. In order to leverage experimental data as evidence in the comparison of contrasting syntactic analyses however, what is needed is a transparent link between syntactic representations, processing mechanisms, and behavioral results (Rambow and Joshi 1994, Kobele et al. 2013).

Here, we argue that a computational parsing model grounded in the rich grammar formalism of Minimalist grammars (MG, Stabler 1996) can serve such purpose, by providing an interpretable link between syntactic assumptions and off-line processing results (Kobele et al. 2013, Graf et al. 2017, De Santo 2020).

As a case study, we model the preferences reported for relative clause attachment in Korean and Mandarin Chinese. Various syntactic analyses of relative clauses have been distinguished in terms of the relation between the relative clause head and the relative clause internal gap. Among many such accounts of relative clauses, the *wh*-movement analysis (*head external analysis*) and the promotion analysis (*head raising analysis*) have been frequently adopted for a wide range of languages. Thus, using relative clause attachment preferences as a starting point, here we leverage a computational parsing model to investigate whether sentence processing results can reveal which analysis is more compelling between these two competing accounts. In doing so, we provide empirical support to the claim that parsing models can give us insights into the effects that different syntactic choices have on processing predictions.

*We thank the anonymous reviewers and the audience of NELS 52 for valuable feedback and comments on this work.

2. Relative clauses in Mandarin and Korean

The cross-linguistic variation in attachment ambiguity preferences for relative clauses is one of the most extensively studied topics in sentence processing. Consider Chinese and Korean sentences corresponding to English sentence '*Someone shot the servant of the actress who was standing on the balcony*' in (1) and (2):

(1) *Chinese*

mouren kaiqiang [$_{RC}$ dasi-le zhanzai yangtaishang-de] nüyanyuan$_{LA}$-de puren$_{HA}$.
Someone shot standing-on-balcony-de$_{\text{REL}}$ actress-de$_{\text{GEN}}$ servant

(2) *Korean*

nwukwun-ka [$_{RC}$ palkhoni-ey se-iss-nun] yepaywu$_{LA}$-uy kacengpwu$_{HA}$-lul sswassta.
Someone$_{\text{NOM}}$ balcony$_{\text{LOC}}$ stand$_{\text{Prog-Rel}}$. actress$_{\text{GEN}}$ maid$_{\text{ACC}}$ shot

These sentences are both ambiguous between two interpretations, depending on whether the relative clause modifies NP *the servant* (known as high attachment, HA) or NP *the actress* (known as low attachment, LA). While these sentences are globally ambiguous–that is, both interpretations are theoretically possible–it is well known that users of the language will have a strong preference for one interpretation over the other. Crucially, these preferences are not consistent across languages, and possibly vary even within the same language–a fact that raises important questions about the relation between the relation between parsing mechanisms and grammatical representations. In our case, for instance, Mandarin Chinese and Korean differ in their reported preferences: while Chinese is attested as a low attachment language (Kwon et al. 2019), high attachment preferences have been reported for Korean (Lee 2021:a.o.).

2.1 Preverbal RCs and Head Directionality

This difference in attachment preferences between the two languages is of particular interest because of a structural similarity the two share. Relative clauses can be categorized according to the linear relation of the relative clause and the noun they modify.

As can be seen from the examples in 1 and 2, both Mandarin and Korean are usually classified as *prenominal* languages: a RC is linearly followed by its head noun (in contrast, for example, with English, where the RC follows the noun). Additionally, while Chinese is a head initial language (SVO word order language) like English, Korean is a head final language (SOV word order language).

Thus, evaluating relative clauses analyses across there two languages provide an interesting contrast in terms of structural differences and processing preferences. We could wonder whether the differences in head directionality enough to account for the processing differences across the two languages, and whether structural details of different type matter at all. As a preliminary attempt to address these questions then, we look at two contrasting analyses of RC structure for the two languages.

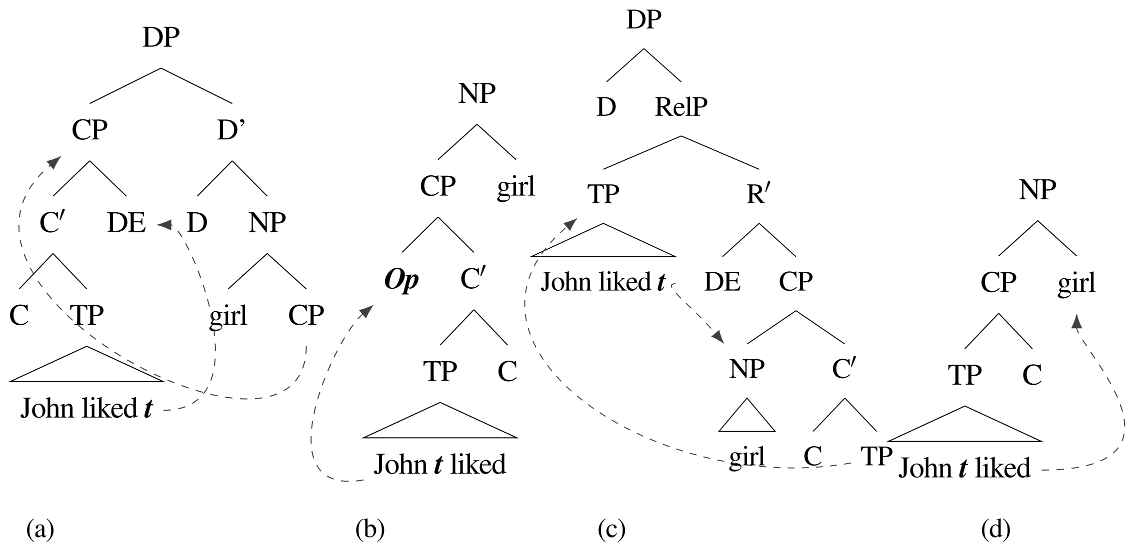


Figure 1. *Wh*-movement analysis and Promotion analyses adapted to Chinese *(a,c)* and Korean *(b,d)*

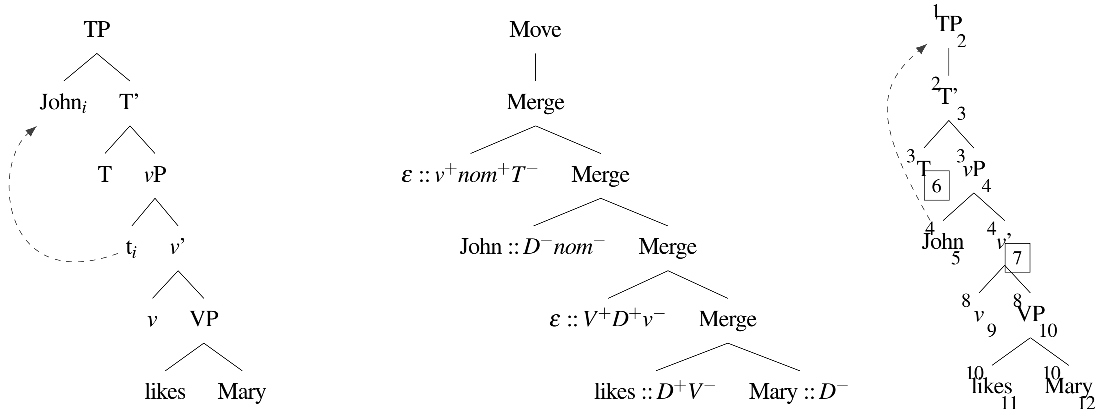
2.2 Two relative clause analyses

The *Wh*-movement analysis (Chomsky 1977) and the promotion analysis (Kayne 1994) are leading syntactic analyses of relative clauses (Bianchi 2002) and differ in a few subtle ways. The core difference between the two is that the element undergoing movements is either the *wh*-element or the head NP. A sketch of this approach is shown in Figure 1 for both Chinese and Korean. According to the *wh*-movement analysis, the head noun is not part of the RC. The *wh*-element starts out in the embedded RC, and undergoes movement from the base position to Spec, CP. Following a promotion analysis instead, the head noun is generated in the embedded RC and moves into the Spec, CP. The *wh*-element (i.e. *who* or a silent *wh*-operator) fills the complementizer position.

3. A top-down parser for MGs as a psycholinguistic model

Computational models grounded in rich grammatical formalism can provide transparent, interpretable links between syntactic assumptions and processing behaviors (De Santo 2020). Here, we adopt a top-down parser for MGs, combined with a set of metrics to connect the way the parser navigates the geometry of the tree to memory usage (Kobele et al. 2013, Gerth 2015). MGs (Stabler 1996) are a formalization of an earlier version of Minimalist syntax. MGs consist of a finite set of lexical items (LIs), each with a phonetic form and a finite, non-empty string of features. Importantly, the central data structure in MGs are *derivation trees* which encode the sequence of Merge and Move operations needed to produce a specific phrase structure representation (see 2).

In a derivation trees, a Merge operation is represented as binary branching nodes, but a Move operation is represented as a unary branching node. Since the tree is only encoding the movement operation, but leaves the actual displacement to a linearization procedure, simply reading the leaf



(a) (b) (c)

Figure 2. Phrase structure tree *(a)*, MG derivation tree *(b)*, and annotated derivation tree *(c)* for *John likes Mary*. Boxed nodes in *(c)* are those with tenure value greater than 2, following (Graf and Marcinek 2014).

nodes of derivation trees from left to right usually does not produce the actual word order of the sentence.

Since the surface order of lexical items in the derivation tree is not matched with the surface order of lexical items from the phrase structure tree, simply scanning the terminal nodes of the derivation tree from left to right yields the incorrect word order. Thus, the MG top-down parser keeps tracking the derivation operations that affect the linear word order while searching and scanning the nodes. The annotated version of the derivation tree in (2c) shows how the MG parser works. For details about the tree traversal procedure the reader is referred to Kobele et al. (2013).

In the annotated tree, the superscripted (index) and subscripted (outdex) numbers on each node indicate when the node is introduced into memory and when the node comes out from memory respectively. For instance, since the scanned word order of the terminal nodes should be the same with the order of terminal nodes read from left to right in the phrase structure tree, the first leaf node which must be scanned in (2c) is *John*. In order to search *John*, the MG parser starts to travel from the root node TP and moves downward and left to right until it reaches to the node *John*. While traveling, all the intermediate nodes from the upper most node TP to *John*'s sister node *v'* are introduced into memory. They stay in memory until they can be discharged following the word order of the phrase structure tree. In this way, the parser will scan every terminal node in the derivation tree. The index and outdex on the annotated tree can inform us about the memory usage of each node. This allows the MG model to quantify processing difficulty based on how the structure of a derivation tree affects memory usage during a parse.

Stabler's top-down parser for MGs, combined with a set of metrics quantifying memory usage, has already been proved to successfully explain processing difficulty across a variety of cross-linguistic phenomena (Kobele et al. 2013, Gerth 2015, Lee 2018:a.o.). We will use the same metrics in our analysis. The MG parser defines complexity metrics associating syntactic structure and processing difficulty based on general cognitive notions of memory usage: i) how long a node stays in memory (Tenure), ii) the length of a movement dependency spanning nodes over the tree (Size). Complexity metrics then can be evaluated over a full syntactic derivation, giving us a measure of off-line processing cost as modulated by differences in grammatical structure. The relevant metrics to our study are MAXT–the maximum amount of time any node stays in memory during processing–and SUMSIZE–the total length of the movement dependencies over a derivation.

Building on this previous work, in the next section, we will evaluate Chinese and Korean sentences including relative clause attachment ambiguities with the MG model.

4. Evaluating Atatchment Preferences

We tested sentences like in (3a) and (3b), modulated across attachment choice and syntactic analysis. Note that our study focuses on object relative clauses only.

(3) a. *Chinese*

[$_{RC}$ wǒ jiànguò de] yīshēng de érzi
I saw de$_{REL}$ doctor of son

'the son of the doctor that I saw'

b. *Korean*

[$_{RC}$nay-ka bo-n] uysa-uy atul
I$_{NOM}$ saw$_{REL}$ doctor-of son

'the son of the doctor that I saw'

We evaluated four different structures (= 2 attachment choices x 2 syntactic analyses) in Chinese and Korean, respectively. The annotated derivation trees for Chinese and Korean RCs are given in Figure 3 through Figure 6.

(4) *Summary of results for Mandarin Chinese*

Wh-movement analysis	SumSize score	Promotion analysis	SumSize score
HA	16	HA	16
LA	15	HA	16
Prediction	LA preference	Prediction	no preference

First, the results from Chinese data in (4) show that the MG parser successfully predicts a LA preference only when the trees were built following a *wh*-movement analysis. As we can see from Figure 3, under a *wh*-movement analysis three nodes (DE, I and CP) undergo movement. Intuitively, for the HA, CP has to move across more lexical items, than for the LA. As a result, LA has the lower score of the SUMSIZE metric, which results in lower processing difficulty than HA structure, Hence, the model's prediction matches the psycholinguistic results.

When the trees are built following a promotion analysis in Chinese, the number of nodes undergoing movements are also three (*DP*, *I* and *TP*). Contrary to the results of the *wh*-movement analysis, since the SUMSIZE scores of HA and LA structures are tied, the model fails to predict any processing preference. In the promotion analysis, the movement lengths of each node (*DP*, *I* and *TP*) were the same in both HA and LA structure. In sum, the model predicts the different

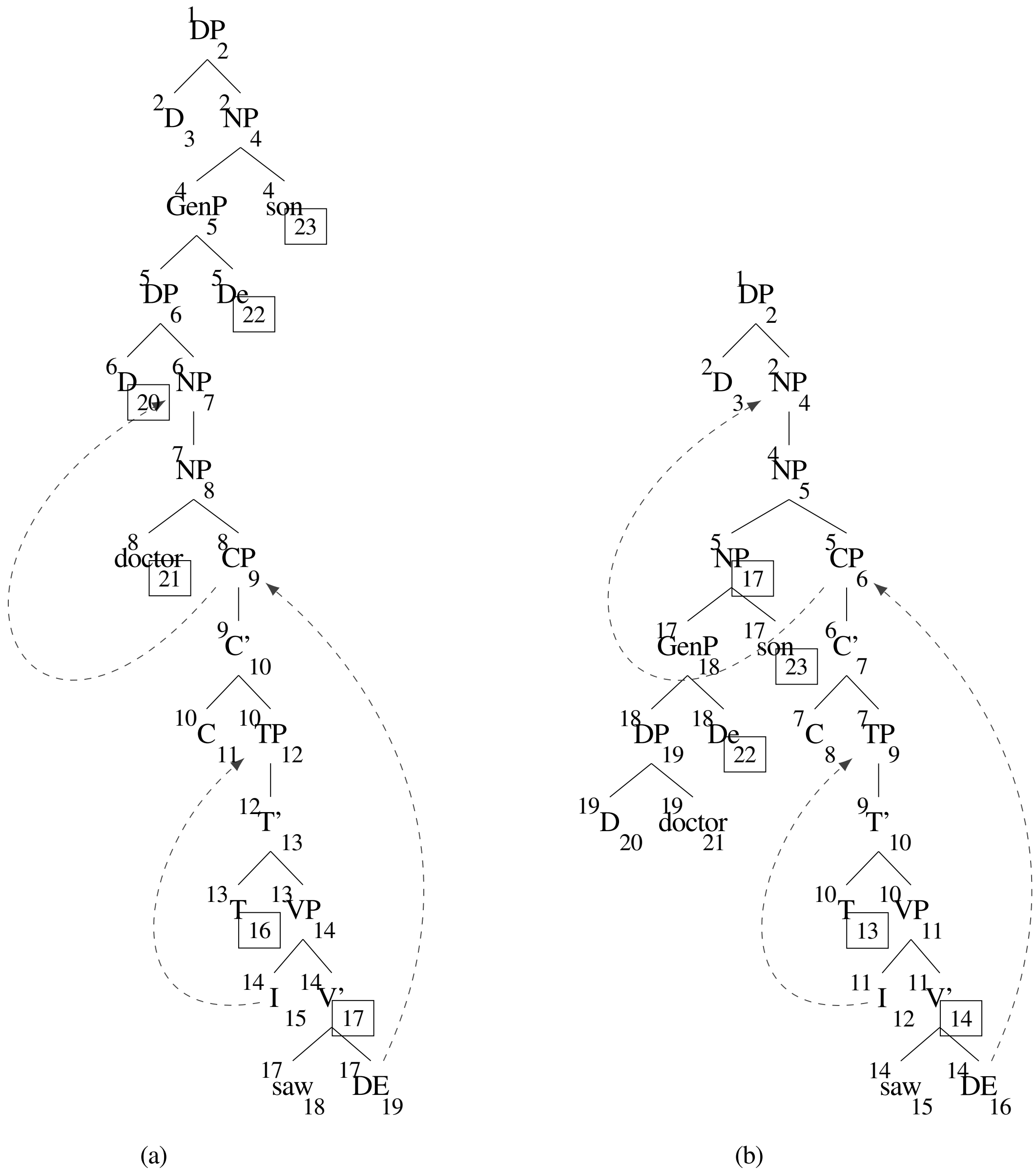
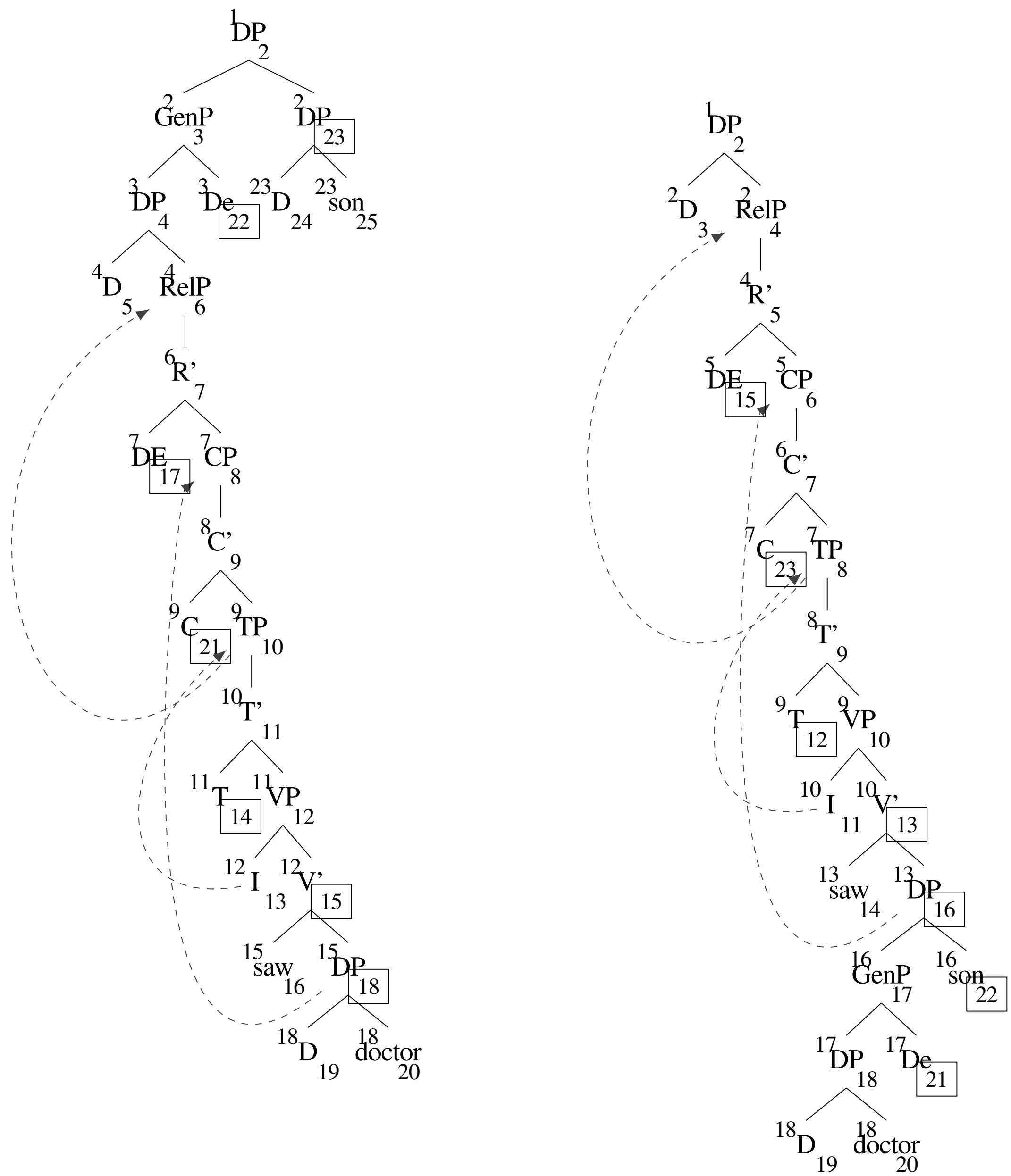


Figure 3. *Wh*-movement analysis: LA structure *(a)* and HA structure *(b)* for *the son of the doctor that I saw* in Chinese.

results depending on the syntactic analysis in Chinese. These are due to the differential effects of the length of the movement and memory commitment caused by different intervening DP structures depending on the analysis.

Now, let us move on to the Korean results. In Korean, tenure related metrics produced the correct predictions.



(a) (b)

Figure 4. Promotion analysis: LA structure *(a)* and HA structure *(b)* for *the son of the doctor that I saw* in Chinese.

(5) *Summary of results for Korean*

Wh-movement analysis	MaxT score	Promotion analysis	MaxT score
HA	12	HA	13
LA	16	LA	15
Prediction	HA preference	Prediction	HA preference

First, the trees built following the *wh*-movement analysis are in Figure 5.

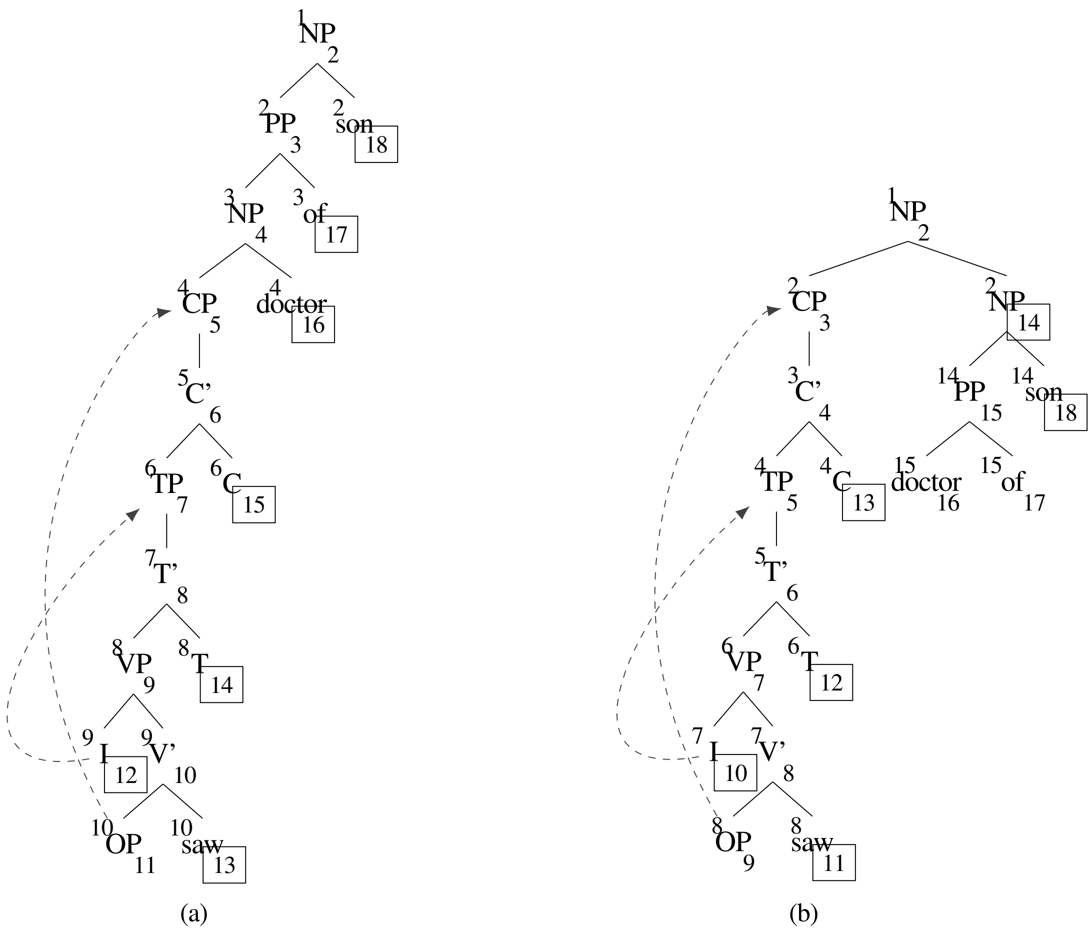
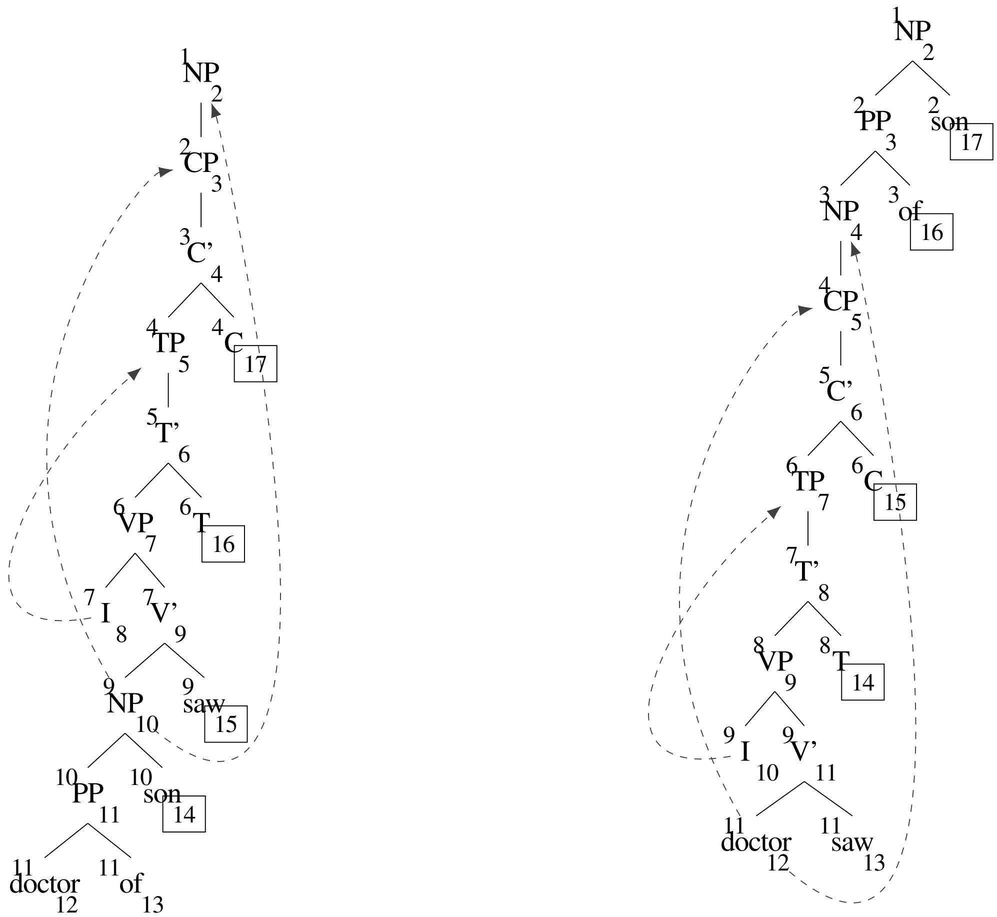


Figure 5. *Wh*-movement analysis: LA structure *(a)* and HA structure *(b)* for *the son of the doctor that I saw* in Korean.

The nodes *NP* and *son* show the highest tenure in both LA and HA structures. As we can see by their indexes, the two nodes are put into memory at step 2. In the high attachment structure, the *NP* stays in memory until the full relative clause is processed. In the low attachment structure, *son* has to wait in memory until the high *NP* (*the doctor*) is additionally processed after processing the relative clause. This yields the higher score (=16) of the MAXT for the LA structure than for the HA structure (= 12). The model thus successfully predicts HA preference. This prediction is matched with the preferences of Korean speakers as reported in previous psycholinguistic studies.

Finally, two trees following a promotion analysis are in Figure 6. The nodes that show the highest tenure in each structure are *C* and *son*. These two nodes are put into memory at step 4 and step 2 respectively and stay until all the lexical items are processed. As the MAXT score is 13 for the HA structure and 15 for the LA structure, the model predicts HA preference again.

Thus, in Korean the model predicts the HA preference successfully regardless of the syntactic analysis. This is because in low attachment interpretation structure in both analyses, the high NP



(a) (b)

Figure 6. Promotion analysis: LA structure *(a)* and HA structure *(b)* for *the son of the doctor* in Korean.

son is predicted at an early stage (like step 2) and needs to wait until the relative clause and the low NP are discharged from memory. This delay causes more memory usage than HA.

5. Conclusion

This paper presents a preliminary evaluation of alternative syntactic analyses for relative clauses in prenominal RC languages with different head directionality, through the lens of a computational parsing model connecting structural details to behavioral results via memory metrics.

The model predicted the processing effect successfully when the trees were built following *wh*-movement analysis in both languages. Importantly, even though the overall results seem to be in support of the *wh*-movement analysis, what we want to draw attention to is how the parsing approach highlights crosslinguistic differences–possibly linked to head directionality variation–in the relation between syntactic choices and behavioral predictions.

While memory metrics successfully predicted the different processing preferences for each language, investigating in detail the relationship between the specific memory metrics and processing mechanisms in individual languages remains for future research. Crucially though, our results suggest that the computational model adopted in this study can link the syntactic details of each analysis to processing performance and inform us of our syntactic choices, and thus opens the road to a closer relevance of psycholinguistic results towards the development of syntactic theories.

References

Bianchi, Valentina. 2002. Headed relative clauses in generative syntax - part i. *Glot International* 6:197–204.

Chomsky, Noam. 1977. On wh-movement. In *Formal syntax*, ed. by A. Akmajian P. Cullicover, T. Wasow, 71–132. Academic Press.

De Santo, Aniello. 2020. Structure and memory: A computational model of storage, gradience, and priming. Doctoral dissertation, State University of New York at Stony Brook.

Gerth, Sabrina. 2015. *Memory limitations in sentence comprehension: A structural-based complexity metric of processing difficulty*, volume 6. Universitätsverlag Potsdam.

Graf, Thomas, and Bradley Marcinek. 2014. Evaluating evaluation metrics for Minimalist parsing. In *Proceedings of the 2014 ACL Workshop on Cognitive Modeling and Computational Linguistics*, 28–36.

Graf, Thomas, James Monette, and Chong Zhang. 2017. Relative clauses as a benchmark for Minimalist parsing. *Journal of Language Modelling* 5:57–106. URL `http://dx.doi.org/10.15398/jlm.v5i1.157`.

Kayne, Richard S. 1994. *The antisymmetry of syntax*. MIT Press.

Kobele, Gregory M, Sabrina Gerth, and John Hale. 2013. Memory resource allocation in top-down Minimalist parsing. In *Formal grammar. FG 2013 2012. Lecture Notes in Computer Science*, ed. by G. Morrill and MJ. Nederhof, volume 8036, 32–51. Springer.

Kwon, Nayoung, Deborah Ong, Hongyue Chen, and Aili Zhang. 2019. The role of animacy and structural information in relative clause attachment: Evidence from Chinese. *Frontiers in Psychology* 10:1576. URL `https://www.frontiersin.org/article/10.3389/fpsyg.2019.01576`.

Lee, So Young. 2018. A Minimalist parsing account of attachment ambiguity in English and Korean. *Journal of Cognitive Science* 19:291–329.

Lee, So Young. 2021. The effect of honorific affix on processing of an attachment ambiguity. *Japanese/Korean Linguistics* 28:1–10.

Rambow, Owen, and Aravind K Joshi. 1994. A processing model for free word order languages. In *Perspectives on sentence processing*, ed. by K. Rayne C. Clifton Jr., L. Frazier, 267–301. Lawrence Erlbaum Associates.

Stabler, Edward P. 1996. Derivational Minimalism. In *International Conference on Logical Aspects of Computational Linguistics*, 68–95. Springer.

So Young Lee, Aniello De Santo
lees44@miamioh.edu, aniello.desanto@utah.edu