Condition C violations in Thai and St'át'imcets

Zia van Blankenstein
University of Victoria
ziavanb@gmail.com

This paper aims to look at the relationship between different types of noun phrases in sentence structure. It focuses on Binding Theory, specifically, by outlining apparent Condition C violations found in both Thai and St'át'imcets. It presents examples of the apparent violations and consolidates restrictions found in previous literature on when these violations can occur. The violations in the two languages are compared and Dechaine and Wiltscko's (2002) pro-PhiP theory is used to account for both violations. Through applying this theory to the St'át'imcets violations, numerous issues are found and presented.
Keywords: Syntax; binding theory; condition c; pro-phiP

1   Introduction

In this paper I present grammatical Condition C violations in two languages, Thai and St'át'imcets. By reviewing key literature on this topic, I outline the restrictions on when Condition C violations can occur for each language and compare them. I then present Dechaine and Wiltschko's (2002) theory of a pro-PhiP, review how Larson (2005) has used this theory to explain the Condition C violations found in Thai and look at how this argument applies to the violations in St'át'imcets. Based on this, I explore core issues with the theory and sketch an outline for how to account for these inconsistencies. I conclude with the premise that Condition C is functional in both languages and while the pro-PhiP theory may superficially account for the Condition C violations found, it leaves many problems unexplained and therefore is not a complete analysis.

2   Background

In 1981 Chomsky proposed Binding Theory which accounts for the distribution of three types of NPs- anaphors (*herself, themselves, etc.*) pronouns (*he, they, etc.)* and R-expressions (*Sarah, London, etc.*)*. He showed how these Noun Phrases (NPs) are sensitive to different binding domains and proposed three principles for this. One of these principles, Condition C, states that R-expressions must be completely free; they cannot be bound like pronouns and anaphors. This means an R-expression cannot be coindexed and c-commanded by an antecedent. This is evident in (1) where the R-expression *Haley* cannot be bound by the pronoun *she*; an example of an ungrammatical Condition C violation.

(1) *She_i wanted to dance with Haley_i

The languages studied in this paper, Thai and St'át'imcets, were chosen as though they are both Condition C violating languages, they are very different languages structurally and historically. While Thai is the primary language in the country of Thailand, St'át'imcets is an Indigenous language in Southwestern British Columbia. St'át'imcets is an endangered language with an estimated 300 speakers (Census Canada, 2016).

3   Thai Condition C Violations and Restrictions

There have been a growing number of documented languages that exhibit permissible Condition C violations. This challenges the idea of a universal Binding Theory and suggests Condition C is not robust cross-linguistically. One of the first languages reported to exhibit these violations was Thai (Lasnik, 1989). In Thai, R-expressions can be bound in many domains. This is demonstrated in (2), where the R-expression *nɔ̀yi* 'Noi' can be bound by the identical R-expression *nɔ̀yi* 'Noi'.

(2) nɔ̀yi_i   khít   wâa   nɔ̀yi_i   càʔ   chanáʔ
     Noi   think   that   Noi   will   win
     'Noi_i thinks that she_i will win'   (Deen & Timyam, 2018)

Thai does, however, maintain certain restrictions on grammatical Condition C violations. First of all, Thai R-expressions cannot be bound by pronouns or anaphors as shown in (3) where khaw 'he' cannot bind John.

(3) *Khaw_i   chɔɔp   John_i
     he   likes   John
     '*He_i likes John_i'   (Lasnik, 1991)

Secondly, an R-expression in Thai cannot be bound by a different definite R-expression. This was proposed to be due to an exact-copy condition (Lee, 2003). Accordingly, the Condition C violations are only permissible when the bound variable is an exact copy of its antecedent, as seen in (2). However, Larson (2005) claims that this condition is insufficient; she presents cases where only part of the R-expression is copied. For example, in (4), only *aajan* 'teacher' is copied in the bound R-expression, not the entire R-expression *aajan Sid* 'teacher Sid'.

(4) Aajan   Sid_i   bɔɔk   waa   aajan_i   mâi   waang   phrungnii
     Teacher Sid   tell   COP[1] teacher   not   free   tomorrow
     'Teacher Sid_i said that he_i isn't free tomorrow'   (Larson, 2005)

---

[1] The following abbreviations are used in this paper: CAUS = causative (neutral) transitivizer, COMP = complementizer, DET = determiner, ERG = ergative (transitive) subject, IMPF = imperfective auxiliary, MID = middle intransitivizer, NOM = nominalizer, PART = particle, POSS = possessive, RED = redirective (applicative) transitivizer.

As Thai is a head initial language, Larson accounts for this by posing a Head Constraint which states that minimally, the antecedent head must be copied in the bound R-expression. She demonstrates that the complement alone cannot be copied by providing examples such as (5) where it is ungrammatical for the bound R-expression to copy only the name *Sid* when it is the complement in the antecedent R-expression.

(5) *Aajan   Sid$_i$ bɔɔk   waa   Sid$_i$ mâi   waang   phrungnii
     Teacher   Sid   tell   COMP   Sid   not   free   tomorrow
     'Teacher Sid$_i$ said that he$_i$ isn't free tomorrow'   (Larson, 2005)

4   St'át'imcets Condition C Violations and Restrictions

St'át'imcets also exhibits Condition C violations that are restricted to specific environments (Davis, 2009). An example of a Condition C violation in St'át'imcets is shown in (6), where *John* is bound by *snilh* 'he'.

(6) Tsúkw=t'u7   snilh$_i$ wa7   xát'-min'-as
     finish=PART   *s/he*   *IMPF*   *want-RED-3ERG*
     kw=a=s   nas   ts'úqwaz'-am   kw=s=John$_i$
     DET=(NOM)IMPF+ 3POSS   go   fish-MID   DET=NOM=John
     'Only he$_i$ wants that John$_i$ goes fishing'   (Davis, 2009)

The restrictions in St'át'imcets that limit violations of Condition C are as follows: First, Condition C violations only occur across clause boundaries, meaning the two coindexed elements must be separated by a clause boundary. We can see this in (7) where there is only one clause and therefore Condition C cannot be violated ('He loves John' cannot mean 'John loves himself').

(7) Wa7   xwey-s- -ás   kw=s=John
     IMPF   *love-CAUS-3ERG*   *DET=NOM=John*
     'S/he loves John'   (Davis, 2009)

The second restriction on Condition C violations states that the c-commanding element (the antecedent) must always be a pronoun. Davis (2009) also found additional constraints relating to bound variable anaphora when there are multiple possible referents available that are not relevant for the scope of this paper but worth noting.

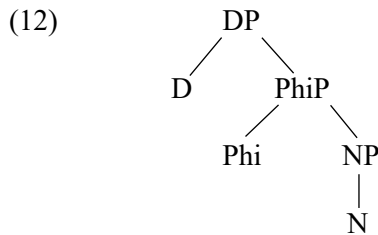## 5   Comparing Thai and St'át'imcets Violations

As Condition C violations in both Thai and St'át'imcets are highly restricted and systematic, we can conclude that both languages do have an operational Condition C in their language. Interestingly, Thai and St'át'imcets exhibit certain similarities in terms of their Condition C violations. Both languages show robust strong

crossover effects in long distance wh-movement, meaning that wh-traces cannot be A-bound either locally or non-locally. A St'át'imcets example of this is displayed in (8) where the wh-trace *swat* 'who' cannot be bound by *pro*. Thai demonstrates the same phenomenon in (9) where *khray* 'who' cannot be bound by *khaw* 'he'.

(8) Swat$_i$ ku=s-kw7íkwlacw-s   pro$_i$ [kw=s=cuz'   melyíh
    who   *DET=NOM-dream-3POSS*   *[DET=NOM=going.to   marry*
    e   kalál]?
    soon
    '*Who$_i$ did s/he$_i$ dream [*e* was going to get married soon]?'

                                                                (Davis,
    2009)

(9) *Khray$_i$ thii   khaw$_i$ khit   t$_i$ waa   Nit   rak   t$_i$
    who   *COMP*   *he*   *think*   *COMP*   *Nit*   *love*
    '*Who$_i$ does he$_i$ think Nit loves?'   (Lee, 2003)

It has also been proposed that both languages have a restriction on Condition C in terms of locality. As noted previously, St'át'imcets cannot demonstrate a grammatical Condition C violation within one clause. Similarly, Larson (2005) states that in Thai, the R-expression cannot be locally bound, as shown in (10).

(10) *Sak$_i$ dti   Sak$_i$
     Sak   *hit*   *Sak*
     'Sak$_i$ hit himself$_i$'   (Larson, 2005)

However, Larson's claim is contradictory to evidence provided by Lee (2003) where (11) is grammatical even though the R-expression is bound locally.

(11) John$_i$ konnaud   John$_i$
     John   *shaved*   *John*
     'John$_i$ shaved himself$_i$'   (Lee, 2003)

This suggests either one of their examples is inaccurate or something other than a locality constraint, perhaps a difference in verb class or a pragmatic effect, accounts for this contrast.

Thai and St'át'imcets also have differences in their Condition C violations. The primary difference being that in Thai, a pronoun cannot bind an R-expression whereas in St'át'imcets this is the only way Condition C violations can occur. The Most Dependent Hierarchy is a ranking which states that anaphors are the most dependent, R-expressions are the least and pronouns are in between the two (anaphor > pronoun > R-expression) (Safir, 2004). This means that St'át'imcets violates the Most Dependent Hierarchy whereas Thai does not. This idea of a binding hierarchy has been used to account for the restrictions in the Thai
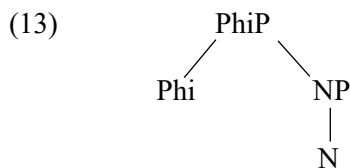
violations as an R-expression is never dependent on a pronoun, so it has been argued that the hierarchy is not violated (Larson, 2005). This is not true of the St'át'imcets restrictions. This hierarchy also does not explain why a Thai R-expression cannot be bound by a different R-expression antecedent; it can not fully account for the Thai data. Throughout my research I found the only explanation that seemed to successfully account for the Thai violations analyzes the structure of nominals (Larson, 2005). This also aligns with Chomsky's recommendation of analyzing Condition C violations by looking at the how the pronoun is constructed (personal communication, April 1, 2021).

6 Pro-PhiP Theory

Using Dechaine and Wiltschko's (2002) argument for three different types of pronouns, Larson (2005) proposed this explanation to account for Condition C in Thai. I refer to this argument as the Pro-PhiP Theory. Dechaine and Wiltschko (2002) propose that pronouns are determined morphosyntactically and that the three types of pronouns each have a distinct structure. The first pronoun, referred to as pro-DP, is claimed to be the most syntactically complex of the three, functioning as an R-expression and containing PhiP (ΦP) and NP as subconstituents, as shown in (12).

(12)

```
              DP
             /  \
            D    PhiP
                /    \
              Phi     NP
                       |
                       N
```

The second pronoun, pro-PhiP, is represented in (13). These pronouns only contain Phi- features (number, gender and person). They can function either as predicates or arguments but they do not act as full DPs.

(13)

```
            PhiP
           /    \
         Phi     NP
                  |
                  N
```
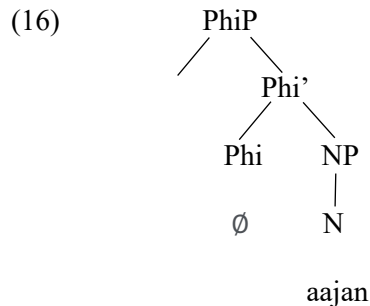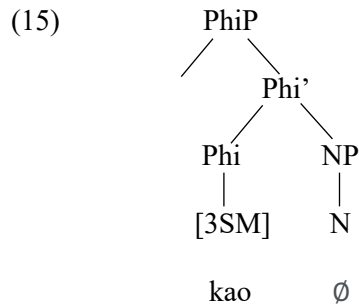
The third pronoun is a pro-NP, shown in (14). They have the same syntax as a lexical noun and are the simplest structurally, functioning as predicates only.

(14)   NP
          |
          N

Dechaine and Wiltschko (2002) claim that only pro-DP is visible to Condition C whereas pro-PhiP and pro-NP are not. Following this theory, the Most Dependent Hierarchy can be restated as: pro-NP > pro-PhiP > pro-DP.

6.1   Applying Pro-PhiP Theory to Thai Condition C Violations

Larson (2005) proposed that the bound R-expressions in Thai are not DPs but pro-PhiPs and the antecedent R-expressions are pro-DPs. For the bound pro-PhiP to gain its features from the antecedent, the antecedent is first spelled out as a pro-DP; it can then license the bound pro-PhiP. At this point, according to Larson (2005), Spell-Out occurs of the pro-PhiP. There are two options for Spell-Out: it can be just of the Phi features or it can be only of the noun. If Spell-Out is just of the Phi features, a bound pronoun surfaces. If it is of the noun, the copy of the antecedent's head surfaces. This is seen in the trees below where *kao* 'he' surfaces in (15) but *aajan* 'teacher' surfaces in (16).

(15)   PhiP
             /   \
                  Phi'
                 /   \
            Phi   NP
             | |
           [3SM]   N

            kao   Ø

(16)   PhiP
             /   \
                  Phi'
                 /   \
            Phi   NP
             | |
             Ø   N

                    aajan

Larson (2005) did not thoroughly explain why sometimes the noun will be spelled out and other times the Phi will be. She stated that a speaker can alternate between the two forms of pro-PhiP, suggesting this may be due to pragmatics. I found this to be an unsatisfying explanation for why one variant would be selected

over the other. This is the main gap in applying the Pro-PhiP analysis to Thai. However, this analysis does account for the Thai data which is why I wanted to explore it further in St'át'imcets.

6.2 Applying Pro-PhiP theory to St'át'imcets Condition C Violations

Davis's (2009) proposal for St'át'imcets' failure to demonstrate Condition C effects focuses on a parameterization of Safir's (2004) Independence Principle. This principle states that a dependent pronoun cannot c-command its antecedent. Davis's (2009) parameterization of this principle is highly language specific which aligns with his conclusion that binding domains should be assessed on a language-specific basis. This conclusion directly opposes the universality of Binding Theory. When asked about this idea of assessing Binding Theory on a language specific basis, Chomsky stated that this does not tell us anything as, despite any distinction in the language itself, we are still left with universal principles (personal communication, April 1, 2021). For this reason, I aimed to find another way to account for the St'át'imcets Condition C violations.

Applying Pro-PhiP theory to the previous St'át'imcets example (6) would mean that *snilh* 'he' would be a pro-DP whereas *John* would be a pro-PhiP.

(6) Tsúkw=t'u7 $snilh_i$ wa7   xát'-min'-as
     finish=PART   *s/he*   *IMPF*   *want-RED-3ERG*
     kw=a=s   nas   ts'úqwaz'-am   kw=s=John$_i$
     DET=(NOM)IMPF+ 3POSS   *go*   *fish-MID*   *DET=NOM=John*
     'Only he$_i$ wants that John$_i$ goes fishing'   (Davis, 2009)

This would explain why *John* is bound and invisible to Condition C. This would also resolve St'át'imcets' violations of the Most Dependent Hierarchy. For this to work, the bound pro-PhiP would pick up Phi features from its antecedent, the pro-DP, exactly as in the Thai examples where the pro-PhiP is spelled out like an R-expression. However, assuming that this is correct, there is no explanation for how an R-expression's phonetic form could surface if it is receiving all its features from a pronoun. In Thai, the R-expression could surface as the same form as its antecedent. This is not the case in St'át'imcets, however, as it does not have the same Head Constraint. To account for this, one possibility parallels Larson's (2005) idea of two options for pro-PhiP structures in Thai. St'át'imcets could have two possible pro-DP structures, one where the Phi head is spelled out and a pronoun surfaces, the other where the NP head is spelled out and the R-expression surfaces. This would mean that even if the pro-DP antecedent surfaces as a pronoun, there is still the structure of an R-expression for the bound pro-PhiP to "copy". This would account for why an R-expression can surface when it is bound by a pronoun. However, this is just a proposal and it would require further research to determine if it is a fruitful claim.

Another discrepancy in applying this theory to St'át'imcets is it does not explain why an R-expression cannot be bound by another R-expression. If all R-

expressions were pro-PhiP's in St'át'imcets then this would be logical as only a pro-DP can bind other variables. However, as shown (17), where the pronoun *snilh* 'she' can be bound to the R-expression *Mary*, R-expressions in St'át'imcets can be pro-DPs as they can bind pronouns in non-condition C violating sentences.

(17) Tsút=tu7   s=Mary$_i$ [kw=s=cuz'   snilh$_i$ nas
      say=then   *NOM=Mary*   *[DET=NOM=going.to*   *she*   *go*
      ts'úqwaz'-am   natcw
      fish-MID   *tomorrow*
      'Mary$_i$ said she$_i$ was going fishing tomorrow'   (Davis, 2009)

This could be resolved by simply stating that all bound variables in St'át'imcets are pro-PhiPs, but then we are left with the question of why Condition C violations only occur across clause boundaries. Though this theory has accounted for how Condition C violations occur, it has dismissed the uniqueness of these violations; if all bound variables are pro-PhiPs why would only those that violate Condition C have such specific restrictions?

7   Discussion

Looking at the Thai and St'át'imcets examples where Condition C has been thought to be absent, we can determine that Condition C is in fact present in the language. This question then becomes not whether Condition C exists in the language, but what is allowing for the apparent violations. To begin to answer this question I looked at the violations under the lens of pro-PhiP theory. I found that despite accounting for the data and allowing modification of the Most Dependent Hierarchy, it was still not a comprehensive enough. The theory could not account for the specific restrictions on the binding of R-expressions, especially those in St'át'imcets. Additionally, I had to propose two possible pro-DP structures to account for how a pro-PhiP R-expression could surface despite the antecedent being a pro-DP pronoun. Another large gap in the analysis, for both languages, was determining when and why each type of pronoun would occur and, when there are two possible structures, how a speaker alternates between the two variants. As a result, further research is needed to address these questions about pro-PhiP theory.

        Another option that could be explored for analyzing these violations would be to look at them from the view of phase theory. Phase theory (Chomsky, 2008) outlines how the syntactic derivation is constructed and sent off for interpretation in phases that once sent become inaccessible. Chomsky (2008) mentions looking at Condition C as a probe-goal (agreement) relation within phase structure. However, there is minimal research on Condition C specifically in relation to phase theory and there is a lack of research on how grammatical Condition C violations could occur under phase theory. With the phase impenetrability condition perhaps a phase including the antecedent is sent to Spell-Out first and then the bound variable is somehow invisible to Spell-Out, allowing it to violate Condition C. This is just a rough outline of how these violations may connect to phase theory. To

fully investigate this theory, it would also be important to see at what stage of the derivation Condition C applies, another topic disputed in the literature.

8   Conclusion

The evidence presented in this paper has confirmed findings that Condition C is present in languages that exhibit apparent Condition C violations. It has also shown that Condition C violations are highly restricted and that Thai and St'át'imcets, two structurally very different languages, show similarities in their limitations on violating Condition C. This was demonstrated in the strong crossover effect examples. We also saw how the Pro-PhiP theory could account for the data in both languages if pro-DP was adapted to have two variants for Spell-Out in St'át'imcets. However, through analyzing this data, shortcomings arose which provoked further question about the validity of this theory. In conclusion, assessing Condition C violations on a language-specific basis challenges the universality of Binding Theory and makes it necessary to analyze pronoun structure and other syntactic structures in the language that may account for these violations.

References

Chomsky, N. 1981. *Lectures on government and binding.* Dordrecht: Foris.

Chomsky, N. (2008). On phases. *Current Studies in Linguistics Series, 45*, 133.

Davis, H. (2009). Cross-linguistic variation in anaphoric dependencies: evidence from the Pacific Northwest. *Natural Language & Linguistic Theory, 27*(1), 1-43. Retrieved from
https://link.springer.com/article/10.1007%252Fs11049-008-9062-0

Déchaine, R. M., & Wiltschko, M. (2002). Decomposing pronouns. *Linguistic inquiry, 33*(3), 409-442. Retrieved from
https://direct.mit.edu/ling/article/33/3/409/160/Decomposing-Pronouns

Deen, K. U., & Timyam, N. (2018). Condition C in adult and child Thai. *Language, 94*(1), 157-190. Retrieved from
https://muse.jhu.edu/article/688304/summary?casa_token=y3jQnVp6trYA
AAAA:YhOu4KXBGe2_O5LwU1xru-DGVPoo5-
kIF6z_V0UuNeQizKIEN3ebjukJVnMURt7c1INpjp3mAQ

Larson, M. (2005). The Thais that bind: Principle C and bound expressions in Thai. *NorthEast Linguistic Society (NELS), 36*, 427–40

Lasnik, H. (1989). On the necessity of Binding Conditions 1986. In *Essays on anaphora* (pp. 149-167). Springer, Dordrecht.

Lee, F. (2003). Anaphoric R–expressions as bound variables. *Syntax, 6*(1), 84-114. Retrieved from
https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-
9612.00057?casa_token=TsqZW5pVF2cAAAAA:JT18HakiP-oESPt-
cO_YDQr0g-1yFX8X_FvpRyxXkmvFiypt22cE86R_ffy997PAizlGpf
PItjnRm5I

Safir, K. J. (2004). *The syntax of (in) dependence*. Cambridge: MIT Press.

Statistics Canada (2016). *Census Profile, 2016 Census-Language.* Retrieved from https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=PR&Code1=01&Geo2=PR&Code2=01&Data=Count&SearchText=Canada&SearchType=Begins&SearchPR=01&B1=Language&TABID=1