How Computer Selects Antecedent?

Seung-Chul Moon

Hankuk Aviation University
200-1 Hwajon-dong, Deokyang-ku,
Koyang-city, Korea, 411-791

scmoon@mail.hangkong.ac.kr

1. Introduction

This study provides a computer-simulated optimal selection of antecedent of reflexive pronouns based on the Visual C++ computer programming language. The algorithm that I use for this implementation is based on the optimal theoretic approach to the proper interpretation of reflexive pronoun roughly adopted in Moon (1999). I show that the complex behavior of reflexive pronouns can nicely be explained in the Optimality Theory (henceforth, OT) and the optimal selection of reflexive pronoun can be implemented on computer simulation.

Following the ideas of Hendriks & Hoop (1999) and Moon (1995, 1999), I claim that the seemingly intricate anaphoric interpretation better be explained by the interaction of the six ranked violable constraints-Thematic Hierarchy Constraint (THC), Larger Domain Preference Constraint (LPC), Subject-Oriented Constraint (SOC), C-Command Constraint (CCC), Locality Condition (LOC), and Discourse Binding Constraint (DBC).

The organization of this study is as follows: in section two I briefly introduce the core of optimality theory developed in Prince and Smolensky (1993); in section three I explore the major characteristics of unbounded reflexives; in section four I explain how OT is applied to the proper interpretation of reflexive pronouns; and finally in section five I show the computer implementation algorithm that I used for the optimal antecedent selection of reflexive pronoun.

2. The Basic Mechanism of OT

The current OT mechanism is made up of mainly four parts: the first part is The Input, the second part is Generator, the third part is Evaluator and the fourth part is Constraints. If we compare OT with Minimalist Program (henceforth, MP), the Input and Generator part correspond to the Lexicon and the numerations of lexical items of MP, and the Evaluator and Constraints part correspond to the PF, LF and Interface Levels of MP.

Input (Lexicon, or Vocabulary) creates linguistically well-formed objects. Generator creates a candidate set of potential outputs. Evaluator selects the best (optimal) output for that input. Constraints provides language particular ranking of constraints for Evaluator.

(1) Basic OT Tableau

INPUT	Constraint 1	Constraint 2
candidate 1	*!	
candidate 2		*!

(*=violated, !=dropped off, shaded area=irrelevant,   =optimal output)

In Tableau (1), Constraint 1 dominates Constraint 2. This means that for any two candidates with otherwise identical constraint violations, one which violates Constraint 1 will be discarded in favor of one which satisfies Constraint 1 and violates Constraint 2 instead. Constraint violations are shown with "*", and the constraint violation which eliminates a candidate from further consideration is marked with "!". Thus in Tableau (1) the violation of Constraint 1 by candidate 1 eliminates it so that it cannot be chosen as an optimal output. Although it violates Constraint 2, candidate 2 best satisfies the constraint hierarchy in (1). This is represented by   to the right of relevant candidate.

3. Unbounded Reflexive Pronoun

Reflexives are those which show both short-distance and long-distance binding and thus are not regulated by the current standard Binding Theory. They include monomorphemic anaphors such as Korean *caki*, Japanese *zibun* and Chinese *ziji*. They may behave in a manner similar to the English reflexive anaphor *himself*, pronominal *him* or even PRO. Reflexives allowing different choices of antecedents are illustrated in (2)-(4) for Korean, Japanese, and Chinese. The superscript numbers 1, 2, 3 represent a preference order relation; thus 1 represents the best choice among potential antecedents for reflexive, 2 represents the next best choice, and so on.

(2) Korean
    John1-un [Tomj2-i cakii/j-lul pipanhayss-ul ttay] amwu-malto an hayssta.
    John-top Tom-nom self-acc criticize-comp when anything did not say
   'John did not say anything when Tom criticized himself.'

(3) Japanese (From Iida 1992)
    Tarooi-wa [Hanakoj-ga zibuni/j-o hihansita noni] damatteita.
    Taroo-top [Hanako-nom self-acc criticized though kept-silent
   'Taroo said nothing though Hanako criticized him/herself.'

(4) Chinese (From Y. Huang 1994)
    Johni yiwei Billj huaiyi Tomk ai shang le zijii/j/k.
    John think Bill suspect Tom love resultative verb ASP self
   'John thinks that Bill suspects that Tom has fallen in love with him/himself.'

These reflexives have been noted in English as well. Chomsky (1986:174) noted that the English reflexives *each other* and *himself* behave similarly in the so-called picture noun construction:

(5) Theyi told usj that [[pictures of each otheri/*j] would be on sale].
(6) Johni thinks that [[pictures of himselfi] would be on sale].

In the above examples reflexives are bound by the NPs outside the embedded clauses, suggesting that they behave in a similar way to Korean type reflexives.

In dealing with these type of reflexives which allow multiple antecedents, I assume that the interpretation proceeds from a total freedom. This means that the anaphoric interpretation of sentences like (2-6) are associated with the number of possible antecedents. If there are three possible antecedents in the sentence, all possible interpretations are evaluated with respect to certain constraints in a parallel fashion to obtain the best antecedent. The intuition for this is that if there is syntactic information, use it to determine the best interpretation; if there is a thematic information, use it to determine the best interpretation; if there is a salient discourse information, use it to determine the best interpretation, etc.. Crucially, we will see in what follows that many of the constraints based on the above-mentioned types of information can be violated during the interpretation process.

Moon (1995, 1999) formulates a series of constraints that play a role in the interpretation of anaphoric expressions through their mutual interaction with each other. First constraint defines that the anaphoric interpretation is subject to the Thematic Hierarchy Constraint stated in (7):

(7) Thematic Hierarchy Constraint (THC)
    Antecedent for reflexive pronoun must be thematically higher.

For our convenience, I assume the following thematic hierarchy (8) roughly advocated in Grimshaw (1990):

(8) Thematic Hierarchy
Agent > Experiencer > Goal, Theme, Patient, Source > Locative

THC (7) and Thematic Hierarchy (8) dictate that the element which is higher in the thematic hierarchy is qualified to be a reflexive's antecedent. When there is only one possible antecedent, it must be thematically higher than the reflexive. If a possible antecedent NP is thematically lower than, or equal to reflexive, it violates THC. If there are more than two possible antecedents in a sentence which are

thematically higher than reflexive, they are all qualified to be its antecedent. When reflexive is included in an element, I assume that it bears the thematic role of the NP that includes it. Consider an example (9) which shows theme's inability to be agent's antecedent:

(9) *Cakii-ka(agent) Johni-ul(theme) pipanhayssta.
     self-nom   John-acc criticized
     'He criticized John'

THC predicts straightforwardly that the sentence is unacceptable since *caki* taking an agent role cannot be bound by *John* taking a theme role.

    Second constraint proposed is the Larger Domain Preference Constraint (LPC) stated in (20) to which the anaphoric interpretation is subject:

(10) Larger Domain Preference Constraint. (LPC)
     Antecedent for reflexive pronoun must be in the larger domain.

LPC states that reflexive generally prefers to seek its antecedent in the larger of the domains. Thus in the following example, the antecedent in the larger domain will be chosen as the best choice of antecedent for reflexive pronouns:

(11) [S2 Johni-un [S1 Billj-i cakii/j-lul miwehanta-ko] mitnunta.]
         John-top   Bill-nom self-acc hate-comp believe
         'John believes that Bill hates him.'

This LPC plays a crucial role in determining the best choice of antecedent when two antecedents with identical thematic roles compete with each other to be a proper antecedent.

    Third constraint proposes that anaphoric interpretation is subject to the Subject Orientedness Constraint (SOC):

(12) Subject-Orientedness Condition (SOC)
     Antecedent for reflexive pronoun must be a subject.

Subject-orientedness occurs explicitly in the passive construction. Consider the contrast in example (13) and its passive counterpart (14):

(13) Johni-i Billj-ul cakii/*j-uy pang-eyse poassta.
     John-nom Bill-acc self-of room in saw
     'John saw Bill in self's room.'

(14) Billi-i Johnj-eyuyhay cakii/*j-uy pang-eyse po-ieci-essta.
     Bill-nom John-by self-of room-in was seen
     'Bill was seen by John in self's room.'

In the active sentence (23), the object *Bill* cannot be an antecedent, whereas in passive sentence (24) it can be. Binding in passives strongly suggests that subject-orientedness is crucially relevant to anaphoric interpretation.

    Fourth constraint proposes that reflexive is subject to the C-Command Constraint (CCC) as in (15):

(15) C-Command Constraint (CCC):
     Antecedent for reflexive pronoun must be in c-commanding position.

CCC accounts for why anaphoric interpretation is impossible in the following sentence, where reflexive pronoun is not c-commanded by the antecedent *John*:

(16) *Cakii-ka Johni-ul miwehanta.
     self-nom John-acc hate
     'Himself hates John'

Moreover CCC explains why both the subject *John* and object *Bill* may serve as antecedents in sentence (17), where reflexive pronoun *caki* is c-commanded by both *John* and *Bill*:

(17) Johni-i Billj-ul cakii/j-uy pang-eyse ttayliessta
     John-nom Bill-acc self-of room-in hit
     'John hit Bill in self's room.'

Fifth constraint proposes that the reflexive interpretation is subject to the Locality Condition (LOC) as in (18):

(18) Locality Constraint (LOC):
     Antecedent for reflexive pronoun must be in local domain.

This new constraint LOC is set up for languages like English which show a strong locality tendency in anaphoric interpretation. In contrast to LPC, LOC finds its antecedent in local domain as we see in the following example:

(19) Johni thinks that Tomj criticizes himself*i/j.

In discourse-oriented languages like Korean, LOC is more often violated than the configurative languages like English as we see in the example given above. Violability degree is different according to the languages under analysis.
    Even though LOC seems to be a weak-motivated constraint, I assume that it exists universally in both languages. Different constraint ranking will predict the best output in OT mechanism. LOC plays a more important role in English, thus it takes the fourth rank in English and the lowest rank in Korean.
    Final constraint proposes that reflexive interpretation is subject to the Discourse Binding Constraint (DBC) as in (20):

(20) Discourse Binding Constraint (DBC)
     Antecedent for reflexive pronoun must be a prominent discourse NP.

DBC explains how *John* can be *caki*'s antecedent straightforwardly in discourse linked example (21) or arbitrary interpretation example (22):

(21) Discourse-linked interpretation
     a. Nwu-ka **Johni**-eykey yenge-lul kaluchiesseyo?
        who-non John-dat English-acc taught
        'Who taught John English?'

     b. **Cakii**-ka honca paywesseyo.
        self-nom alone learned
        'He did alone.'

(22) Arbitary interpretation
     a. [Cakiarb-ka cakiarb-lul kukpokha-nun] il-un taytanhi elyepta. (epigram)
        self-non self-acc overcome-comp thing-top very hard
        'It is very hard to overcome oneself.'

     b. Cakiarb-uy coy-lul hoykayhay-ya kwuwen-ul etnunta. (Bible)
        self-of sin-acc repent-must salvation-acc secure
        'One must repent one's sin to secure one's salvation.'

Reflexive's discourse-linked interpretation occurs quite consistently and productively in the common speech, which provides a reasonable motivation to form a discourse- motivated constraint to predict it. DBC will naturally capture the general discourse binding property of reflexive. In (21b) where there is no sentential antecedent, DBC tells the grammar to search for a most prominent NP in the previous sentences, which is *John* in this case. Thus the sentence obtains a proper interpretation on the basis of discourse information.

4. Application of OT to Anaphoric Interpretation

In this section, I will demonstrate that the constraints proposed in the previous section interact to produce the best choice of antecedents. Recall that the interaction of multiple constraints and the constraint ranking are one of the key notions in selecting the optimal output. For the best result of the anaphoric interpretation I tested all possible combination of the constraints. For the sake of our space convenience, I simply propose the best constraint ranking (23a) for Korean and (23b) for English:

(23) a. THC > LPC > SOC > CCC > DBC > LOC for Korean
      b. THC > SOC > CCC > LOC > LPC > DBC for English

Given that the ranking (23a,b) correctly predict the anaphoric interpretation, let us look at the Korean simplex sentences:

(24) John-i caki-lul pipanhayssta.

INPUT	THC	LPC	SOC	CCC	DBC	LOC
caki=John		*			*	
caki=Dis NP	*!	*	*	*		*
caki=	*!	*	*	*	*	*

(25) John-i Bill-ul caki-uy pang-eyse poassta.

INPUT	THC	LPC	SOC	CCC	DBC	LOC
caki=John		*			*	
caki=Bill		*	*!		*	
caki=Dis NP	*!	*	*!	*		*

(26) a. Nwu-ka **John**i-eykey yenge-lul kaluchiesseyo?
      'Who taught John English?'
      b. **Caki**i-ka honca paywesseyo.
        'He did alone.'

INPUT	THC	LPC	SOC	CCC	DBC	LOC
caki=John	*	*	*	*	*!	*
caki=Dis NP	*	*	*	*		*
caki=	*	*	*	*	*!	*

Let's turn to the following English sentences and their tableaus corresponding to the Korean ones:

(27) John hates himself.

INPUT	THC	SOC	CCC	LOC	LPC	DBC
himself=John					*	*
himself=DisNP	*!	*	*	*	*	
himself=	*!	*	*	*	*	*

(28) Picture of himself annoyed John.

INPUT	THC	SOC	CCC	LOC	LPC	DBC
himself=John					*	*
himself=DisNP	*!	*	*	*	*	
himself=	*!	*	*	*	*	*

(29) John thinks that pictures of himself would be on sale.

INPUT	THC	SOC	CCC	LOC	LPC	DBC

himself=John					*		*
himself=DisNP	*!	*	*	*	*	
himself=	*!	*	*	*	*	*

(30) John told Bill that pictures of himself would be on sale.

INPUT himself=John	THC	SOC	CCC	LOC	LPC	DBC
				*		*
himself=Bill	*!	*		*		*
himself=DisNP	*!	*	*	*	*	

OT can also correctly predicts the optimal output in the psych sentences. If we assume that THC > LPC > CCC (brief form for our convenience) in constraint hierarchy is correct in Korean, one can obtain the right anaphoric interpretation in psych construction as illustrated in tableu (31):

(31) [Billi-i cakii/j-lul piphanhayssta-nun] sosik-i Johnj-ul silmangsikiessta.
'The news that Bill criticized self disappointed John.'

INPUT caki=John	THC	LPC	CCC
			*
caki=Bill		*!	
caki=Dis NP	*!	*	*

In tableau (31), we see only *John* satisfies THC and LPC, thus being selected as an optimal antecedent, while *Dis NP* and *Bill* drop off in the THC and LPC column respectively.

Now let us consider English psych constructions:

(32) The news that Billi criticized himselfi/*j disappointed Johnj.

INPUT himself=John	THC	CCC	LOC
		*!	
himself=Bill			*
himself=DisNP	*!	*	*

In (32) *John* satisfies THC but violates CCC because it stands structurally lower than 'himself.' On the other hand, *Bill* satisfies both THC and CCC, thus being selected to be an optimal antecedent. LOC is irrelevant constraint in this psych constructions because selection procedure stops at CCC column. Discourse antecedent is out of question since there are two more competitive (more potential in a traditional sense) antecedents in the sentence.

Consider another examples which shows that the constraint hierarchy is correct in the following sentence:

(33) [S2 Johni-i [S1 Billj-i cakii/j-lul cohahanta-ko] malhayssta.]
John-nom Bill-nom self-acc like-comp said
'John said that Bill liked him/himself.'

INPUT caki=John	THC	LPC	SOC	CCC	DBC	LOC
					*	*
caki=Bill		*!			*	
caki=Dis NP	*!	*	*	*		*

The explanation is similar to the psych-verb construction case, since in the first column, *John* and *Bill* satisfy THC due to their agent thematic roles. In the LPC column, only *John* can satisfy LPC because *John* is in the larger domain S2. Here *John* is selected as an optimal output and *Bill* as the second best choice. Notice that they both satisfy the SOC and CCC, which is irrelevant in selecting both optimal outputs.

If we apply OT to the English counterpart, the same accurate anaphoric interpretation is obtained as in (34):

(34) John said that Bill likes himself.

INPUT	THC	SOC	CCC	LOC	LPC	DBC
himself=John				*!		*
himself=Bill					*	*
himself=DisNP	*!	*	*	*	*	

It is quite surprising that the constraint hierarchy can select the best candidate *Bill* of subordinate clause subject as an optimal output, whereas Korean example picks up *John* of the higher clause subject.

In sum, this section showed that the seemingly intricate anaphoric interpretation can be captured by the interaction of the ranked violable constraints-Thematic Hierarchy Constraint (THC), Larger Domain Preference Constraint (LPC), Subject-Oriented Constraint (SOC), C-Command Constraint (CCC), Locality Condition (LOC), Discourse Binding Constraint (DBC).

5. Computer Implementation Algorithm

If grammar as a whole is essentially an optimality system, as argued in Prince and Smolensky (1993), such a system must be computer-implemented to show that this program really works in natural languages. The whole process of computer algorithm will be presented by a notebook computer in the actual presentation for better understanding.

References

Chomsky, Noam. 1992. A Minimalist Program for Linguistic Theory. *MIT*

Dalrymple, Mary. 1993. *The Syntax of Anaphoric Binding*. CSLI Lecture Notes No 36. Stanford   University: Stanford.

Grimshaw, Jane. 1990. *Argument structure*. Cambridge, Massachusetts: The MIT Press.

Grimshaw, Jane. 1996. Projection, Heads and Optimality, *Linguistic Inquiry* 28:373-422

Grimshaw, Jane. 1998. Constraints on Constraints in OT Syntax. Unpublished manuscript,   Rutgers /University, New Brunswick, NJ.

Hendriks   Petra,   &   Helen   de   Hoop.   1999.   *Optimality   Theoretic   Semantics*.   [ROA, http://ruccs.rutgers.edu/roa.html]

Huang, Yan. 1994. *The Syntax and Pragmatics of Anaphora; A study with special reference to   Chinese*. Cambridge: Cambridge University Press.

Iida, Masayo. 1992. *Context and Binding in Japanese*. Doctoral Dissertation. Stanford University.   Stanford.

Moon,   Seung-Chul.   1995.   *An   Optimality   Approach   To   Long   Distance   Anaphors*,   Doctoral Dissertation, University of Washington.

Moon, Seung-Chul. 1999. Optimal Anaphoric Interpretation in Psych-verb Construction. *Studies in Generative Grammar*: vol.9:no.1, Hankuk MwunHwaSa. Seoul

Prince, Alan S., & Paul Smolensky. 1993. *Optimality Theory; Constraint Interaction in*   *Generative Grammar*. Ms. Rutgers University, New Brunswick, NJ.

Rene Kager, 1998. *Optimality Theory, A textbook*. To appear at Cambridge University Press

Yang, Dong-Whee. 1983. The Extended Binding Theory of Anaphors. *Language Research*.   19:169-192. Seoul National University.