Search for a Minimal Agent Predicate Link Preference in Recursive Agent Distribution Strategy for Embedded Clauses

Abhijit Debnath and Gautam Sengupta
University of Hyderabad

1. Introduction

Studies in initial preference of control (restricted however to the obligatory control types, subject and object control) have ascertained how certain structural, lexical as well as general processing restrictions in the memory may be capable of initiating default predictions of antecedent assignment to PRO before as well as after the control information, in the control verb, is available to the processor. This behaviour has been used to understand and make explanatory assumptions regarding internal processing strategies that facilitate the information comprehension and production mechanism in the language processing and production system in the human mind. Findings in processing of control sentences in Spanish show a significant preference towards Object Control compared to Subject Control (Betancort et al. 2006) supporting the preference for a structurally close antecedent for PRO. Japanese shows that in case of a retrieval task, performed after control sentences (where the task of recalling the antecedent of PRO immediately follows the Control Verb in a left branching verb final structure), matrix clause objects rather than subjects are capable of being recalled and retrieved more easily and accurately as the antecedents for PRO (compared to the matrix subjects) (Sakamoto 2001). This however showed up in both canonical as well as scrambled word order, suggesting the role of a grammatical function in controlling the preference, which remained consistent across word orders, canonical and scrambled. The current work reports two self-paced reading experiments in Bangla, carried out in order to ascertain whether a *Minimal Agent-Predicate Link* could be a probable processing preference that results in the increase of processing complexity when the number of association links between any agent and the predicates (the verbs either in matrix clause or embedded clause) increases. Bangla provides a very sensitized design for the tests by allowing a sentence final location for the matrix verb (that provides the actual control information) at the end of a sentence (like Japanese), after the initial control choices can be taken to have already been processed.

Findings in Spanish: Findings in Spanish sentences having a Control-Verb ∼ PRO sequence suggest that there is an immediate usage of control information to recover the antecedent of PRO. The control information at the matrix verb is infered to be immediately used to select the controller NP. It is therefore concluded that this initial selection of controller NP for the EC seems to rest on lexical factors (at the matrix verb). A preference for the nearest NP filler is also found in the Spanish sentences showing the effect of structural factors. Therefore both lexical and structural factors are found to be active in affecting the processing of the integration of an antecedent with PRO.

Findings in Japanese: Spanish has a Control-Verb ∼ PRO sequence. But the influence on preference is bound to be different in a different kind of sequence. In languages like Japanese such a sequence

(PRO ∼ Control-Verb) is available. A study in Sakamoto (2001) involved a comparative analysis of four experiments. The first two had post-stimulus recognition tasks whereas the third and fourth had post-stimulus retrieval tasks. Both the recognition task as well as retrieval task experiments had a variation in the disambiguating point for *subject control matrix verb* and *object control matrix verb* at the end of the sentence. Findings showed that grammatical functions of the antecedent (i.e., object or subject) is significant for building a preference of the control type that may be initially built by the parser before encountering the final disambiguating control verb. This remained consistent even in scrambled versions as well. But consistency in this preference could not be traced across types of response task. The experiment basically tested two types of post-stimulus behavioural tasks, recognition and retrieval where the preference for control type varied significantly according to the type of post-stimulus task as well. These two sets of experiment were studied (involving two types of response tasks), each having sentences varying for usual SOV -Vs- scrambled OSV word order. In case of the recognition task, the subjects were given a stimulus sentence in audio mode and after it finished playing they were given an audio stimulus which mentioned one of the arguments. They had been instructed to identify, and press from given choice of Yes/No response button and indicate whether the later mentioned argument was the doer of the action in the infinite clause (i.e., the correct antecedent for PRO). In case of the retrieval task, the subjects were given similar sentences in audio mode. They had been instructed to speak out the antecedent after the sentence finished playing. Time to respond as well as accuracy of response in both the task types was measured. There was a significant difference in preference only according to the type of task where retrieval task indicated Object Control preference and recognition task indicated Subject Control preference.

2. Replication of tests in Bangla:

In the current Experiments two types of issues were considered

- Non-finite Subordinate clause Vs Finite Subordinate Clause

- Object Control Vs Subject Control (within Non-Finite Subordinate Clauses)

2.1. Experiment 1

The first experiment was designed to test whether the preference can exist independent of [+- Finite] factor. Therefore two types of sentences (2.1.1 and 2.1.2) were prepared where the first had a non-finite subordinate clause and the second had a finite subordinate clause. For the current test, only a comparison between Object Controlled PRO in a Non-Finite Clause Vs matrix-subject bound pro in a finite clause was available for a minimal pair design. The counter part design, i.e., Subject Controlled PRO in a Non-Finite Clause Vs object bound pro in a finite clause , was not available as a suitable grammatical sample of the latter was not available in Bangla. Processing difference was measured at the matrix verbs, (for RT of button-press activity in a masked self-paced reading at the target word in msec), in the two types of sentences. One required a dropped pro at the subject position of the embedded clause and the other required a controlled PRO in the embedded clause.

1. John$_i$ Mary-ke$_j$ boleche [PRO$_{*i/j}$ *kaal bikel-e harmonium bhaara kor-e* gan kor-**te**] $^{OBJ\ CON}$
John Mary-Dat/Acc tell-ppl [tomorrow evening-loc har... rent do-conj.prt song do-inf]
John told Mary (Mary) to sing *having rented a harmonium next evening*.

2. John$_i$ Mary-ke$_j$ boleche [*pro$_{*i/j}$ kaal bikel-e harmonium bhaara kor-e* gan kor-**be**] $^{SUBJ\ BOUND}$
John Mary-Dat/Acc tell-ppl [tomorrow evening-loc har... rent do-conj.prt song do-Fut(3rd)]
John told Mary (that John) will sing *having rented a harmonium next evening*.

Sentence type (2.1.1) has an infinite embedded clause having an object controlled PRO whereas sentence type (2.1.2) has a finite embedded clause and needs a matrix subject bound *pro* at its subject position. The sentences are framed into a minimal pair by minimally varying for the last syllable of the verb (**-te** and **-be**) at the sentence end. Therefore the sentence remains temporarily ambiguous till before the last NV complex predicate. This complex predicate is the disambiguation point which has

been properly shifted to the end of the sentence and the actual disambiguation has been placed at the ending syllable of the whole sentence. Thus if the processor either forms preferences and predictions regarding the binding of the EC's (PRO or pro) or has inbuilt processing strategies for comprehension of the binding as such, then the whole sentence till before the end, shall allow the processor to already formulate an identity for the performer of the embedded action accordingly. If a verb at the end, with the control disambiguation, does not indicate a conformity with such a formulation, RT at that verb is expected to be more than that at a verb that indicates a conformity.

40 subjects aged 20 to 35 participated in the experiment. Comparison of RT at the last word showed a preference for an object controlled PRO (mean RT: aS-te (926.4 msec) < aS-be (1118.0 msec), difference of mean = 191.6 msec, t-value > 2) suggesting that Matrix-Object is preferred as an antecedent for PRO. This suggests that the closest antecedent might be preferred for PRO in Bangla. But in a free phrase order language like Bangla, scrambling has also be considered.

2.2. Experiment 2

The second experiment (Debnath 2011) pseudo-replicated Sakamoto and Oda et. al.'s experiment using both canonical and scrambled word orders (for the sequence of subject and object in the matrix clause), to test for control preference. Processing difference was measured at the last verb, in RT, in the two types of sentences. Default as well as Scrambled word order was also used for each type of control.

1. ami oke [$_{INF\ CLAUSE}$ boi-Ta di-te] baddho **kor**-l-am $^{OBJ\ CON}$

   (a) ami$_i$ o-ke$_j$ [PRO$_{*i/j/*k}$ boi-Ta di-te] baddho kor-l-am
   1^{st}SgNom 3rd-Dat/Acc [PRO book-Cl give-inf] compel do-pst-1st
   I compelled him to give the book.
   OR

   (b) ami$_i$ *pro$_j$* [PRO$_{*i/j}$ o-ke boi-Ta di-te] baddho kor-l-am
   1^{st}sg-Nom pro [PRO 3rd-Dat/Acc book-Cl to.give] compelled-do(pst)-1st
   I compelled *someone* to give him/her the book.

2. ami$_i$ [$_{INF\ CLAUSE}$ PRO$_{i/*j}$ o-ke boi-Ta di-te] baddho **ho**-l-am $^{SUBJ\ CON}$
   1^{st}SgNom [PRO 3rd-Dat/Acc book-Cl give-inf] compel be-pst-1st
   I was compelled to give him the book.

(2.2.1) has an object controlled PRO in the embedded infinite clause and (2.2.2) has a subject controlled PRO. The sentences formed a minimal pair by minimally varying for the first syllable of the last verb (**-kor** and **-ho**). The scrambled counterparts that were prepared are as follows.

(1') <u>o-ke</u> <u>ami</u> boi-Ta di-te baddho kor-l-am

(2') <u>o-ke</u> <u>ami</u> boi-Ta di-te baddho ho-l-am

In Japanese sentences, additional information had been introduced between the first and the second NP through a PP, in order to increase the information load between the first two NPs, the two potential antecedents for PRO. This effect has been suitably achieved in Bangla by introducing a conjunctive participle between the subject and object of the matrix clause. In the following sample the conjunctive participle 'shokal-e librari-te giye' (morning-loc library-loc go.pft) meaning '*after having.gone to the library*' is used (as shown in the finally prepared sentence sample)

- ami **shokal-e librari-te giye** oke boi-Ta di-te baddho kor-l-am

- ami **shokal-e librari-te giye** oke boi-Ta di-te baddho ho-l-am

- oke **shokal-e librari-te giye** ami boi-Ta di-te baddho kor-l-am

- oke **shokal-e librari-te giye** ami boi-Ta di-te baddho ho-l-am

Comparison of RT at the last word of the two types of sentences showed a preference for object controlled PRO (mean RT: *kor-l-am* (1627 msec) < *ho-l-am* (2113 msec), difference of mean = 486 msec, t-value > 2) across word orders.

This suggested that an antecedent having the grammatical function of object in the matrix clause, seems to be strongly preferred as the controller for PRO.

3. Spanish, Japanese and Bangla in Comparison

The methodology in the currently studied experiments can be grouped under the following major stimulus→response features.

1. **Stimulus → Response**

   - **Spanish:** Sentences were displayed and the eye-movements were measured on-line while the subjects read the sentences and comprehended them.

   - **Japanese:** The Retrieval Task experiment provided an audio stimulus to the subjects and immediately after the completion of the audio, (according to introductory instructions) they had to speak out the name of the performer of the embedded action.

   - **Bangla:** Self-paced reading task measured the button-press RT patterns at the verbs in the end of the sentence. So immediately after the display of the target stimulus, the response was made.

2. **Stimulus$_1$ → Stimulus$_2$ → Response**

   - **Japanese:** The Recognition Task experiment provided an audio stimulus (same as that in Retrieval Task). But, an additional audio stimulus, containing the name of a probable antecedent, was provided immediately after the first one. The subjects had to respond to the second audio stimulus, and the response was collected in the form of a Yes/No answer to whether or not the uttered name could be a correct antecedent for PRO. The distance of response collection from the original target control verb is more in this case where there is an extra stimulus to process between the verb and the response.

Therefore a consistent pattern of behavioural correlate can be inferred according to the two types of stimulus presentation and response collection strategy. (1) seems more likely to conform to an online measurement, immediately at and after the stimulus where as (2) seems to conform to a post stimulus, offline measurement of response. The difference in processing preference in the Japanese findings seems to be a result of such a difference in measurement techniques collecting responses at different locations, after the critical/target word. (3.1) seems to be relatively online in all three cases where as (3.2) seems to be off-line. Online measurement requires the target word or stimulus to immediately precede the response and in such a case, object control seems to be preferred.

3.1. A consistent preference in on-line processing:

Preference in the on-line processing may need to be analysed in terms of certain causal properties in the language processing faculty, as an embedded default strategy. The preference seems to be controlled by certain elements of information embedded in the control-verb, antecedent-EC relation ship type as well as the properties of the antecedent. This preference seems to persist along the lines of some internal processing phenomenon that has either led to the emergence of such a categorical choice across languages pertaining to the type of antecedent that the EC, PRO, should be associated with, or has come to exist as a result of other physical factors in the memory that makes certain algorithms function to result in sequential information processing. In order to ascertain the reason behind such a preference, we, for now, try to answer possible processing issues, perhaps conforming solely to complexity issues. Complexity of a matrix subject antecedent as a controller may be compared to that in case of a matrix object. In such an explanatory approach, the concept of preference needs to be decomposed in terms of

steps of processing, where preference may play no role but a pre-set ordering of processing steps may be expected to be existing as a processing strategy that enables systematic comprehension as an efficient mechanism, probably established over time. Such a strategy may be claimed to have come to existence due to certain activity minimization strategies dealing with expected processing complexity, according to requirements of the conceptualization of entities, antecedents, performers or agents and the events that they engage in.

Preference...?   It would be treading into a domain of uncertainty if preference strategy were to remain explained in terms of descriptive methods of visibly observed behavioural phenomenon which are assumed to be correlates of a certain internal processing. It is rather a requirement to explain the preference in terms of the causes behind these observations in terms of certain measurable increment of processing load, where the behavioural correlate does not pertain to any preference but is simply a correlate to **lesser** -Vs- **increased** amount of processing steps or levels for a particular component of information whose comprehension has to be completed. Such an incremental approach to understanding this dynamic nature in the variability of complexity can be implemented by attempting to detect required explanatory features in the information input (as in the sample sentence types), precisely correlated to such a measure of complexity.

3.2. A Feasible Hypothesis

To accommodate such an explanatory direction the following hypothetical claim may be considered to be pivotal for the observed preferences in the literature concerning processing of control.

The strategy of conforming to a minimal number of agent-predicate links may be the basic criterion for measuring the complexity that may have resulted in such a preference. The object control information in the matrix verb makes the processor posit a single association chain from the matrix object, linking it with the verb of the embedded clause only after already having posited an association between the matrix verb and the matrix subject. A subject control however makes the processor posit two association links from the matrix subject, that with the matrix verb as well as the embedded verb. Probably a kind of *performer identity in-general* agent concept gets postulated at the infinite verb already before the processor encounters the control information in the matrix verb and may be its identity remains uncontrolled by any overt antecedent.[1] Then at the next stage the control information in the matrix verb reveals the actual control requirements. If the matrix control verb does not require the identity of the agent of the embedded clause to be controlled by any overt antecedent within the matrix clause, then the comprehension is complete (which predicts that PRO-Arb should be most preferred) after reaching the matrix control-verb which has already been linked to a subject in the matrix clause. If the control verb suggests an association of the embedded verb with any matrix clause argument, then (according to all findings) matrix object is the next preferred antecedent for that identity. This second preference may be explained as the matrix object building its first association with the embedded verb with the first increment in association load. Subject control is therefore more complex and probably lesser preferred of the two, because here the association with the embedded verb/action somehow leaves the processor with a requirement for processing the load of two agent-predicate associations where the subject ends up getting linked with both the matrix as well as embedded verb. Another aspect in which it varies from the object control as well, is that in the object control sentence, both the matrix clause NPs are associated with a predicate (either or not embedded) where as in subject control, one of the argument (the matrix object) is left unlinked to or unassigned a predicate. At the moment it is not clear however whether this nature of being unlinked may also be able to affect the processing in terms of any other probable requirement failure (that may probably be understood in terms of other issues of processing)

This hypothesis, however, also predicts that PRO Arb shall show behavioural correlates that can be argued to indicate lesser processing discrepancy compared to both the major types of obligatory control, i.e., object as well as subject control.

Integration of this hypothesis with the findings seems to indicate that this consistent on-line processing preference is actually due to a preference towards a minimal number of agent-predicate links,

---

[1]   This may be guided by the experience in either linking difficulty or certain deeply embedded processing pathways controlled by some information processing and comprehension strategies in the brain.

required to be associated by the processor and not any distance factor or grammatical function of an antecedent. (2.1.2) requires the processor to associate the subject 'John' as an agent of two predicates, 'bole-ch-e' as well as the embedded 'gan kor-be'. Likewise in (2.2.2), the subject 'ami' has to be associated as an agent of 'baddho ho-l-am' as well as the embedded 'di-te'. Thus, diverging associations emerge from the matrix agent to both matrix as well as embedded predicates. (2.1.1) however requires the subject 'John' to be associated with only one predicate, 'bole-ch-e' and the object 'Mary-ke' with only one predicate, 'gan kor-te'. Likewise in (2.2.1), the subject 'ami' is to be associated with 'baddho kor-l-am' and either 'oke' (as in 2.2.1(a)) or, if possible, 'pro' to be associated with the embedded 'di-te'. Increase in the number of agent-predicate links projected from the agent entity seems to in-turn increase the processing cost (which is less in case of object control in both the experiments due to lesser number of agent predicate links).

Japanese sentences suggest a similar situation and as such such an explanation may be claimed to accommodate the findings in that language. Languages having SVO word order have the control verb preceding PRO. In such a case, the control verb already provides the information regarding whether there will be one agent to one predicate linking from the probable agents of the matrix clause (object control) or one gent to two predicate links (subject control). The initial assumptions regarding the linking shall not be formed however at the infinite verb here. They would rather not exist. But the control information at the matrix verb shall prepare the processor to expect the link properties. The matrix subject would get linked with the matrix verb as soon as the verb is encountered. If it is an object control verb, no more upcoming predicates are expected to link up with the same matrix subject but a subject control verb will give rise to such an expectation. It seems plausible to claim that a search for a linkable predicate for this subject is initiated at a subject control verb. This initiation may be understood to be the cause for the increased gaze and other eye-movement measures at the location near or immediately after the subject control verb as well as the total reading time measure in the Spanish subject control sentences.

4. The Search for a Minimal Agent-Predicate Link
4.1. Recursive Agent Distribution Strategy

While processing different types of events which are embedded as sub-events within another event, the processor has to engage in a recursive loop, into the task of assigning associations and links between agents and events, multiple times. The hidden strategy may be better explored when cases of controlled ECs are considered, where the embedded event must be associated with either of the overt agents in the matrix clause. The allotment of agents to embedded events may require consideration of a plausible but non active agent from the matrix clause (as in Object Control) or a reconsideration of the plausible agent which has already been allotted to an event (as in Subject Control). In such a case of recursive association formations, there may be claimed to be an existing strategy that involves in the distribution of agents to the available events.

The current experiment enables us to consider such a satiation factor from the point of view of the agents, rather than the events. The phenomenon still becomes a bit clearer at this point when an algorithm concerning some satiation factor upon such agents may be considered such that during comprehension of events and agents engaged in those events, the processor may reach satiation with least number of events allotted to agents. Perhaps extra processing functions and algorithms have to be activated in order to allow and hold allotments of more events to the same agent when the information requires the comprehension to do so.

4.2. Minimal Agent Predicate Links

Conformation to Minimal Agent Predicate Links may be considered as one of the integral processing strategies that enables an efficiency check in the complexity and adequate comprehensive simplicity in any incoming information. It can be considered to be capable of claiming that the properties of structural configuration may not be the final decisive factor capable of dictating any preference of processing. The minimality in the number of links seem to have a better explanatory adequacy in situations where ECs are the subjects of embedded clauses.

References

Betancort, Moises., Carreiras, Manuel., Acuña-Fariña, Carlos. (2006). "Processing controlled PROs in Spanish." Cognition, 100 217–282.

Betancort, Moises. and Carreiras, Manuel. (2004). "The empty category PRO: Processing what can't be seen," The on-line study of sentence comprehension: Eyetracking, ERP, and beyond, eds. Carreiras, Manuel & Clifton Jr, Charles (eds.), Brighton: Psychology Press.

Debnath, Abhijit. (2011). "Processing of Ambiguous Control of PRO in Bangla", Online Proceedings of GLOW in Asia Workshop for Young Scholars 2011, eds. Koichi Otaki, Hajime Takeyasu, and Shin-ichi Tanigawa. Available online at http://faculty.human.mie-u.ac.jp/~glow_mie/Workshop_Proceedings/05Debnath.pdf

Sakamoto, Tsutomo. (2001) "Processing Filler-gap Constructions in Japanese: The Case of Empty Subject Sentences," Sentence Processing in East Asian Languages, ed. Mineharu Nakayama, 189-221. CSLI Publications.

Proceedings of the 31st West Coast Conference on Formal Linguistics

edited by Robert E. Santana-LaBarge

Cascadilla Proceedings Project   Somerville, MA   2014

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Debnath, Abhijit and Gautam Sengupta. 2014. Search for a Minimal Agent Predicate Link Preference in Recursive Agent Distribution Strategy for Embedded Clauses. In *Proceedings of the 31st West Coast Conference on Formal Linguistics*, ed. Robert E. Santana-LaBarge, 134-140. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #3015.