Pronoun Incorporation in Matsigenka

David Inman
University of Washington

1 Abstract

In this paper I review the literature for pronominal incorporation (PI), and what PI means from an HPSG perspective. I conclude that PI is simply a special case of argument optionality, and contrast the HPSG approach with the LFG analysis in Bresnan and Mchombo (1987), and show that the HPSG approach is lexically more parsimonious. Finally, I show how PI phenomena interact with information structure and motivate non-cancellation of valence lists within Matsigenka [mcb], an Arawakan language of Southeastern Peru. These rules can be used to generate (at least in some cases) syntactically-constrained agreement across coreferential information structural constituents.

2 Background

Research into pronoun incorporation (PI) dates back at least to Givón (1976), who proposes a typological pathway, using evidence from Bantu, for languages to develop verb agreement from isolating syntax: passing from pronoun topicalization to pronoun incorporation to agreement markers.

Formal analyses of pronoun incorporation (PI) are deeply indebted to Jelinek (1984), who (within a Government and Binding framework) argues that apparent agreement markers in the Warlpiri auxiliary complex are actually pronouns fully satisfying a finite verb's arguments. For Jelinek, the distinguishing factor between PI and agreement is exactly this argument satisfaction. She uses morphological evidence (differentiable case marking: erg/abs in the auxiliaries, and nom/acc in the nouns) and Warlpiri's discontinuous syntactic phrases to argue for the necessity of full PI in the auxiliary complex. She concludes that nominals and related parts of speech fulfill adjunct, rather than argument roles of the verb.

Bresnan and Mchombo (1987) elaborate on PI from an LFG perspective and, crucially, give methods to distinguish between PI and agreement within a single language, Chichewa. As with Jelinek, the distinction they make is that PI fuflills a verb's arguments. They offer clear tests to distinguish the two phenomena from each other, and show that Chichewa's object suffixes can act as both incorporated pronouns and agreement markers, and give a straightforward LFG analysis of the syntactic difference. One consequence of this analysis is they end up doubling the number of affixes in the lexicon to cover the two cases.

3 Methods

The data available for this research was provided through the work of Lev Michael and other field linguists (Michael et al, *pending citation*). I reviewed 4 of the 170 texts manually to look for relevant syntactic constructions, and developed scripts to sift through the remaining to find particular phenomena, which were then manually reviewed.

4 Evidence for Pronoun Incorporation

Bresnan and Mchombo (1987) give two ways to distinguish an Incorporated Pronoun (PI) from an agreement marker:

1. An incorporated pronoun enters into an anaphoric relationship with a coreferential NP. Anaphoric relationships are non-local, while agreement is obligatorily local. Non-locality between the coreferential NP and verb demonstrate that the marker must be an incorporated pronoun and not an agreement marker.

2. An incorporated pronoun occurs in island configurations that would be prohibited for unbounded dependencies.

Both these tests depend on locality constraints, and indeed the conditions of one seem to imply the other, especially since their analysis of the Chichewa data ends up treating non-local anaphora as

topics.[1] Subject and object affixes in Matsigenka both pass Bresnan and Mchombo's tests for PI. NPs coreferential with the marker can occur anaphorically bound outside the VP they agree with (test 1):

(1) y-oga=ri i-tovai-re o-onpera-ig-ak-a=ro
3MS-DEM.MED=CNTR 3MP-other.PRO-ALIEN.POSS 3FS-order.about-PL-PERF-REALIS.A=3FO
o-shinto-egi o-ok-ja-ig-u-e=ri=ra oaku
3FP-daughter-PL 3FS-throw-CL:fluid-PL-RET-IRREALIS=3MO=SUB river
The others[masc., fish] she ordered her daughters to throw them[masc] into the river.' [mcb]

In (1), *yogari itovaire* "those others" is agreeing in gender with the PI object of a verb two levels of embedding below it (indeed, it agrees with the only masculine argument in the entire sentence).

The Matsigenka markers also show up in island constructions (test 2), as below.

(2) maika naro no-kog-ak-e no-n-kamant-ako-ak-e=ri=ra
now 1.PRO 1S-want-PERF-realis 1S-IRREALIS-tell-APPL:INDR-PERF-IRREALIS=3MO=SUB
i-kematsa-ig-i=rira o-ntiri y-ovetsik-ge-ig-i=rira matsigenka
3MS-believe.in-PL-REALIS=REL 3FS-COORD 3MS-make-DSTR-PL-REALIS=REL person
'Now, I want to tell them what they believe in and what they make, the Matsigenka. (The Matsigenka are doing the believing and making.)

The morpheme *=rira* (glossed as =REL) is a second-position clitic that relativizes an entire VP, making it an NP (Michael, 2008, p.407−408). In (2), the subjects of these relativized verbs agree in gender with *matsigenka*, which appears outside the coordinated relative clauses.

So by the tests outlined by Bresnan and Mchombo (1987), both subject and object markers in Matsigenka fall into the category of incorporated pronouns.

5 Pronominal Incorporation and Argument Optionality

Both Jelinek (1984) and Bresnan and Mchombo (1987) agree that incorporated pronouns are unique from agreement markers in that they satisfy some valency requirement of a verb. This is in fact the *only* formal distinction between the two, and causes Bresnan and Mchombo (1987) to double the lexical entries for every ambiguous morphological affix. HPSG already has a formal mechanism for handling this type of ambiguity, and PI in HPSG can be shown to be simply a special case of argument optionality, with no need for lexical doubling.

Saleem (2010) gives a library for the broader phenomenon of argument optionality within the LinGO Grammar Matrix (Bender et al., 2002, 2010). She defines argument-dropping in HPSG as "the shortening of a valence list without the overt realization of the argument" and gives unary phrase-structure rules (*head-opt-subj-phrase* and *head-opt-comp-phrase*) to cancel items off valence lists (Saleem and Bender, 2010). The specific instances in which an argument can be cancelled off a valence list is left to language-particular restrictions inherited from these general phrase-structure rules.

Taking the above definition of argument optionality, does PI fall into the same category? The answer is yes: PI is a particular case when items can come off a valence case just in case some marker is present on the verb. There are however a few interesting specific restrictions in the case of PI, as opposed to general argument optionality:

1. In PI, an argument is only missing from valence lists in the case where its semantics have already been constrained by an element from the morphology (namely, the PNG constraints given by the PI). That is, PI phenomena do not include completely unconstrained dropped arguments (as in the dropped object of the English sentence *I ate*).

2. Verbal arguments satisfied via PI can enter into anaphoric relationships with external nouns (Jelinek, 1984) that they agree with, often with special information status (Bresnan and Mchombo,

[1] If topic phrases are modeled as a special type of adjunct to a sentence and inside their own phrase - and this is the prevailing model in the formalisms I know of - it becomes extremely difficult to formulate a scenario that would pass test (1) that does not itself represent an island condition.

1987). This is true for the Matsigenka data as well. Completely dropped arguments cannot have this relationship.

(1) suggests that the analysis should limit optionality based on morphological rules. (2) suggests that it should *not* remove these items from valence lists, so that their PNG values can still be accessed outside of their phrase by topicalized NPs (and other NPs carrying special information structure).

Saleem (2010) creates unary rules *head-opt-subj-phrase* and *head-opt-comp-phrase* based on a typological survey of argument optionality. I leverage the same rule definitions from which I inherit for the Matsigenka PI system. The relevant pieces of the rules are reproduced here.[2]

(3)
$$\begin{bmatrix} \textit{head-opt-subj-phrase} \\ \textsc{synsem} \mid \textsc{subj} \quad \langle\,\rangle \\ \textsc{hd-dtr} \mid \textsc{subj} \quad \langle \begin{bmatrix} \textsc{opt} & + \end{bmatrix} \rangle \end{bmatrix}$$

(4)
$$\begin{bmatrix} \textit{head-opt-comp-phrase} \\ \textsc{synsem} \mid \textsc{comps} \quad \boxed{A} \\ \textsc{hd-dtr} \mid \textsc{comps} \quad \langle \begin{bmatrix} \textsc{opt} & + \end{bmatrix} \rangle \ \oplus \boxed{A} \end{bmatrix}$$

The value of OPT (+, −) is determined by the lexical rules. I posit that the verb valence specifications are defeasibly OPT −, and only become OPT + via the lexical rules below.[3] Lexical rule supertypes are given in (5) and (6), with actual PNG values given by inheriting rules as in example (7).

(5)
$$\begin{bmatrix} \textit{verb-subj-pi} \\ \textsc{synsem} \mid \textsc{subj} \quad \langle \boxed{1} \begin{bmatrix} \textsc{opt} & + \end{bmatrix} \rangle \\ \textsc{hd-dtr} \mid \textsc{subj} \quad \langle \boxed{1} \rangle \end{bmatrix}$$

(6)
$$\begin{bmatrix} \textit{verb-obj-pi} \\ \textsc{synsem} \mid \textsc{comps} \quad \langle \boxed{1} \begin{bmatrix} \textsc{opt} & + \end{bmatrix} \rangle \ \oplus \boxed{A} \\ \textsc{hd-dtr} \mid \textsc{comps} \quad \langle \boxed{1} \rangle \ \oplus \boxed{A} \end{bmatrix}$$

(7)
$$\begin{bmatrix} \textit{3rd-masc-subj-pi} \\ \textsc{phon} \quad \textit{i-} \\ \textsc{synsem} \mid \textsc{subj} \quad \langle \begin{bmatrix} \textsc{png} & \begin{bmatrix} \textsc{pernum} & \text{3rd} \\ \textsc{gender} & \text{masc} \end{bmatrix} \end{bmatrix} \rangle \end{bmatrix}$$

The above rules generate a system whereby an argument can be dropped just in case a morphological process has added a PI marker to the verb. Otherwise, the verb's argument *must* be overtly realized in the syntax, either through a direct SUBJ or COMP-filling phrase structure rule or gapping. This is exactly what we see in the Matsigenka data.

In addition to PIs optionally occurring with an NP for the argument, subject and object PIs can occur with both an agreeing NP within the VP, and a topicalized corefferent NP. The relevant examples in (8) and (10) are given with explicit coreference shown in (9) and (11) below.

(8) ovashi iriro=ri i-sure-an-ak-a iriro=rika
so 3M.PRO=CNTR 3MS-think-ABL-PERF-REALIS.A 3M.PRO=INDEF.TEMP
m-an-ak-e=ro i-*jina matsontsori
bring-ABL-PERF-REALIS=3FO 3MP-wife jaguar
i-jinant-ak-enpa=ro=ra
3MS-take.as.wife-PERF-IRREAL.REFL=3FO=SUB

[2]Saleem (2010) actually specifies that *head-opt-subj-phrase* can only apply if the head daughter has an empty COMPS list. I do not impose that restriction, since I need to leave items on valence lists, as described below.

[3]The DELPH-IN tools do not allow for defeasibility, but the same effect can be achieved through a *verb-no-subj-pi* or *verb-no-obj-pi* rule that add OPT −. I do not think there is a significant difference between these two implementations.

'Then he thought perhaps the jaguar brought his wife to take her as his own wife.' [mcb]

(9) he bring=3FO his-wife jaguar
SUBJ bring=OBJ OBJ SUBJ

"The jaguar brought his wife"

(10) [y-oga=ri i-ketyo=rira y-ovetsik-ak-e tasorintsi matsigenka]
[3MS-DEM.MED=CNTR 3MS-first=REL 3MS-make-PERF-REALIS god person]
i-pait-ig-ak-e=ri Kinteroni o-ntiri Pareni
3MS-name-PL-PERF-REALIS=3MO NAME 3FS-COORD NAME.

'The first people God made he named Kinteroni and Pareni.'[4,5] [mcb]

(11) that first.one 3MS-make=REL god person.
OBJ SUBJ-make=relativizer SUBJ OBJ

"Those first persons God made."

These sentences both appear to show an argument being filled by two noncontiguous NPs that agree both with each other and with the verbal PI in PNG value. However on closer inspection one of the NPs seems to be a topic.[6] I could find no instances of 3 coreferent NPs, which is to be expected if the above sentences are analyzed as containing topics (or some other form of information status) rather than NP adjuncts as Jelinek (1984) claims for Warlpiri.

One analysis for this phenomenon would be to admit an unconstrained topic NP as an adjunct to the VP. However, this will not be able to account for the required agreement of PNG values between topic, NP, and incorporated pronoun. There is however a way to do this within the syntax, and that is through non-cancellation of valence lists.

Bender (2008) adopts non-cancellation of valence lists to account for nonconfigurationality in Wambaya. I adopt the same analysis, but for the slightly different goal of permitting topics to agree with verbal arguments. I modify the typical *head-subj-phrase* and *head-1st-comp-phrase* in the same way as Bender (2008), marking them as INST + (preventing them from being instantiated multiple times). These arguments can then be removed from their appropriate valence list either by the *head-opt* rules in (3) and (4), or by a topicalizing rule and placed on the SLASH list to then later be filled via *head-filler-phrase*. I give examples rules for SUBJ, from which the COMP versions can be easily extrapolated:

(12)
$$\begin{bmatrix} \textit{head-subj-phrase} & \\ \text{SYNSEM} \mid \text{SUBJ} & \langle \boxed{1} \begin{bmatrix} \text{OPT} & + \\ \text{INST} & + \end{bmatrix} \rangle \\ \text{HD-DTR} \mid \text{SUBJ} & \langle \boxed{1} \begin{bmatrix} \text{INST} & - \end{bmatrix} \rangle \\ \text{NON-HEAD-DTR} \mid \text{SYNSEM} & \boxed{1} \end{bmatrix}$$

[4]The syntax of this merits some additional explanation. *Yogari iketyorira yovetsikake tasorintsi matsigenka* "Those first people God made" is relative clause that is nominalized. *=rira* as mentioned above is a second-position clitic that takes the entire verb clause and makes it into an NP with the verb's subject as the head (Michael, 2008, p.407−408). Refer to the brackets in the IGT for the appropriate chunking of the relative clause.

[5]The word *ketyo* is in fact a pronoun with gender information. In some cases where it has feminine agreement, it can be used as an adverb (since verbs always take feminine agreement) - but it is marked as masculine here, making an adverbial parse impossible.

Despite being clearly part of the nominal complex *ketyo* actually encodes aspectual information about an event − that the event occurred first with respect to some sequence. Closely-related Nanti, and presumably Matsigenka as well, also has two other nominal aspect markers in addition to this one: one indicating recency and another indicating temporal overlap (Michael, 2008, p.371−375). This leads to some difficulty giving a direct translation for the NP alone, thus the awkward "that one.first."

[6]Field linguists working on this and related languages have claimed that certain of the free pronouns *always* carry special information status and can only occur in privileged places outside the VP's direct arguments *and* coreferent with what I am analyzing here as incorporated pronouns (Michael, 2008, p.346−352). Meanwhile non-pronominal elements can occur in a wider range of positions (see the relative ordering of the post-verbal NPs in exs 8 and 10). The privileged topic pronouns are illustrated nicely in these examples by *ririori* and *yogari*.

(13)
$$\left[\begin{array}{l}\textit{topic-subj-phrase} \\ \text{SYNSEM} \left[\begin{array}{ll}\text{LOCAL} \mid \text{SUBJ} & \langle\,\rangle \\ \text{NON-LOCAL.SLASH} & \left\langle \left[\text{LOCAL} \mid \text{HOOK} \left[\begin{array}{ll}\text{INDEX} & \boxed{1} \\ \text{ICONS-KEY} & \text{topic}\end{array}\right]\right]\right\rangle\end{array}\right] \\ \text{HD-DTR} \mid \text{SUBJ} \left\langle \text{LOCAL} \mid \text{INDEX} \boxed{1} \right\rangle\end{array}\right]$$

The item on the SLASH list from (13) can then be discharged through a *head-filler-phrase* instantiating the topic. There is now the possibility for two NPs to share an INDEX in the semantics (which guarantees identical PNG features), but can still have different PRED values, since PRED is a feature accessed through LTOP and not INDEX. This is precisely what is desired for coreferentiality.

6 Conclusion

The analysis presented here is superior to the LFG analysis in Bresnan and Mchombo (1987) in its ability to account for the data in a lexically minimal fashion. I also have connected pronominal incorporation directly with the broader phenomena of argument optionality, and repurposed an HPSG analysis of the latter (Saleem, 2010) to accomodate and explain the former. The argument optionality connection also offers a tempting link to linguistic change along the lines proposed in Givón (1976): the movement from PI to agreement marker can be expressed only by changing OPT + to OPT − within the relevant lexical rules. I have presented a reason to allow for rules with non-cancellation of valence lists within the HPSG formalism, which permits syntactically-encoded anaphoric agreement across multiple constituents. The increasing literature that shows the usefulness of keeping arguments on valence lists suggests this may be a way more broadly applicable way forward in HPSG analyses for non-local phenomena.

References

Bender, Emily M. 2008. Radical non-configurationality without shuffle operators: An analysis of wambaya. In *Proceedings of the International Conference on Head-Driven Phrase Structure Grammar*, ed. Stefan Müller, 6–24. Stanford: CSLI Publications. URL `http://cslipublications.stanford.edu/HPSG/9/`.

Bender, Emily M, Scott Drellishak, Antske Fokkens, Laurie Poulson, and Safiyyah Saleem. 2010. Grammar customization. *Research on Language and Computation* 8:23–72.

Bender, Emily M, Dan Flickinger, and Stephan Oepen. 2002. The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In *Proceedings of the 2002 workshop on Grammar engineering and evaluation-Volume 15*, 1–7. Association for Computational Linguistics.

Bresnan, Joan, and Sam A Mchombo. 1987. Topic, pronoun, and agreement in chichewa. *Language* 741–782.

Givón, Talmy. 1976. Topic, pronoun and grammatical agreement. *Subject and topic* 149:188.

Jelinek, Eloise. 1984. Empty categories, case, and configurationality. *Natural Language & Linguistic Theory* 2:39–76.

Michael, Lev. 2008. Nanti evidential practice: Language, knowledge, and social action in an Amazonian society. Doctoral Dissertation, University of Texas.

Saleem, Safiyyah. 2010. Argument optionality: A new library for the grammar matrix customization system. Master's thesis, Citeseer.

Saleem, Safiyyah, and Emily M Bender. 2010. Argument optionality in the lingo grammar matrix. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, 1068–1076. Association for Computational Linguistics.