The morphological causative in Panãra: A Grammar Matrix implementation

Carly Crowther, Emily Luedke, Emily M. Bender
University of Washington

1 Introduction

Panãra (Jê, Brazil) exhibits complex verbal morphology with a strictly ordered set of polypersonal agreement affixes. The verb stem is inflected first with either the object (O) prefix[1] for transitive constructions, or the subject (S) prefix for intransitive constructions. The transitive agent (A) prefix attaches after. A causative morpheme can appear with an intransitive verb, increasing its valence to two core arguments. This morpheme is only compatible with intransitive verbs, as is the S argument inflectional position class. Contrastively, the A and O inflectional position classes are only compatible with transitive verb stems. The causative morpheme position class follows the would-be demoted O prefix, which poses a problem as the order of inflection would not allow the verb to select an object before 'knowing' that the valence is to be increased with the causative morpheme. However, due to the ergative-absolutive alignment of Panãra, the S and O agreement prefixes happen to be homophonous. Furthermore, the interaction with the Valence Change library provided by the Grammar Matrix preserves agreement information about an intransitive subject in the resulting transitive complement. We model the causative morpheme to take the S agreement affix as input, and allow the A agreement prefix to further inflect the verb to produce the desired fully inflected causative verb construction.

2 Background

2.1 Panãra

Panãra [ISO 639-3: kre] is a Jê language spoken by about 630 native speakers, between the states of Pará and Mato Grosso in Brazil (Lapierre 2023). Data for this project comes from Bardagil (2018)'s dissertation, which is based on primary field work and description of the Panãra language. This paper stems from a class project in which we used the Grammar Matrix customization system (Bender et al. 2010) and hand-edited TDL (Type Description Language) files to build an implemented grammar of Panãra.

2.2 The LinGO Grammar Matrix

The LinGO Grammar Matrix (Bender, Flickinger & Oepen 2002, Bender et al. 2010, Zamaraeva et al. 2022) is a grammar engineering framework for creating implemented grammars using Head-Driven Phrase Structure Grammar (HPSG) (Pollard & Sag 1994) and Minimal Recursion Semantics (MRS) (Copestake et al. 2005). The grammar fragments are written in TDL formalism and can be interpreted by other DELPH-IN software, including the Linguistic Knowledge Builder (LKB) grammar development environment (Copestake 2002).

3 Data

3.1 Argument roles, agreement, and case

Panãra has an ergative-absolutive syntactic alignment system. The A argument of a transitive verb is marked with the ergative morpheme *hẽ*.[2] The transitive verb's O argument and intransitive verb's sole S argument do not receive case marking. (1), (2), and (3) illustrate the first person singular pronoun serving as each of the three types of core arguments.[3]

(1) Jyrawâ inkjẽ.
Jy-ra-wâ **inkjẽ**
INTR-1SG.S-born 1SG
'I was born.' (Bardagil 2018:103)

(2) Karân kamẽrânpun inkjẽ.
Ka-rân ka-mẽ-r-ânpun **inkjẽ**
2SG-DU.ERG 2SG.A-DU-1SG.O-see 1SG

[1] Throughout this paper we use the term prefix to refer to the morphemes Bardagil (2018) refers to as clitics. This reflects the analysis we pursued of the morphemes as both syntactically and phonologically dependent, even if less phonologically integrated than other affixes.

[2] Dual and plural personal pronouns are case marked with an ergative suffix rather than *hẽ*, as seen in (2) (Bardagil 2018).

[3] The gloss line in all following IGT examples has been changed slightly from the reference material to reflect the S/O/A argument roles that we focus on in this paper.

'You two saw me.' (Bardagil 2018:121)

(3) Inkjẽ hẽ rêsunpa nãkãã.
Inkjẽ hẽ rê-s-unpa nãkãã
1SG ERG 1SG.A-3SG.O-fear snake
'I'm scared of snakes.' (Bardagil 2018:59)

As shown above, the first person pronoun only receives ergative case marking when it is the A argument of a transitive construction; the S and O arguments pattern together in the unmarked absolutive case.[4] This ergative/absolutive alignment extends to the verbal agreement prefix paradigm as well, described in Tables 1 and 2.

Person	SG	DU	PL
1	rê	rê...mẽ	nẽ
2	ka	ka...mẽ	ka rê
3	ti	ti...mẽ	nẽ

Table 1: Ergative agreement prefixes

Person	SG	DU	PL
1	ra (r)	mẽ...ra (r)	ra
2	a (k)	mẽ...a (k)	rê a (k)
3	ø (s/j)	mẽ...ø (s/j)	ra (r)

Table 2: Absolutive agreement prefixes
(Allophones for vowel-initial verbs in parentheses)

For Panãra intransitive verbs, the prefix that agrees with the sole S argument appears directly adjacent to the left edge of the verb root. For transitives, the O argument agreement prefix occurs in this same location; the A argument agreement prefix precedes it. Panãra also has a dual marker *mẽ-*, which marks agreement with a dual number value on the A, O, or both arguments. Intransitive verbs receive an additional verbal prefix attached to the left edge of the verb, which indicates the intransitivity of the verb. This pattern is summarized in Tables 3 and 4.

3.2 The causative morpheme

Panãra causatives are formed with the verbal prefix *ho-*, which attaches to intransitives only,[5] demoting the S to O and providing a new A argument. Causatives, being derived transitives, don't take the intransitive *jy-* prefix. This pattern is illustrated in (4)–(6).

A agr-	Dual-	O agr-	Verb root

Table 3: Transitive Verbs

Intrans-	Dual-	S agr-	Verb root

Table 4: Intransitive Verbs

(4) Ka jõpãã jysõti.
Ka jõpãã jy-ø-sõti
2SG child INTR-3SG.S-sleep
'Your child sleeps.' (Bardagil 2018: 108)

(5) Ka hẽ kahosõti ka
Ka hẽ ka-ho-ø-sõti ka
2SG ERG 3SG.A-CAUS-3SG.O-sleep 2SG
jõpãã.
jõpãã
child
'You made your child sleep.'
(Bardagil 2018: 108)

(6) *Inkjẽ hẽ rêhokuri inkjẽ
Inkjẽ hẽ rê-ho-ø-kuri inkjẽ
1SG ERG 1SG.A-CAUS-3SG.O-eat 1SG
jõpãã suasĩra jĩ.
jõpãã suasĩra jĩ
child peccary meat
'I made my child eat peccary meat.'
(Bardagil 2018: 174)

4 Analysis

4.1 The Panãra verb

To model the patterns in Tables 3 and 4 with the Grammar Matrix customization system's morphotactics library (Goodman 2013), we created a position class for each of these "slots" in a verb's inflection pattern. The position classes serve to determine the order that the morphemes appear in relation to the verb root. Figure 1 illustrates these chained position classes for both transitive and intransitive verbs. Within each of these position

[4]This pattern is identical for both pronouns and full NPs.

[5]Transitive verbs require a periphrastic construction for creating a causative semantic relation (Bardagil 2018).

classes are multiple lexical rule types (LRTs), one for each person/number combination distinguished in Panãra, with each instantiated by a lexical rule instance.[6]

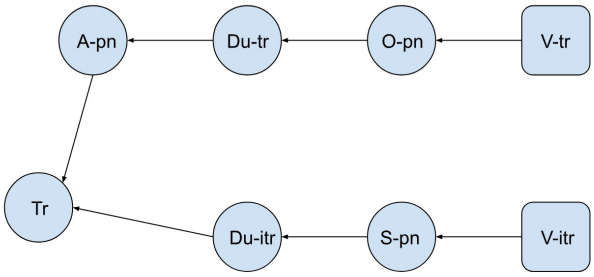

Figure 1: Chained verbal position classes

The transitivity position class at the end of the chain accepts both the output of the A-pn[7] and dual-intransitive position classes as its input. An LRT for transitive verbs requires a transitive argument structure with ergative case on the subject and absolutive on the object, and contributes no affix in this situation. For intransitive verbs, the LRT applies the intransitive prefix *jy-* and contributes an argument-structure constraint of a single, absolutive subject.

4.2 Moving towards a concise implementation of the causative

In analyzing the verbal inflection position classes implemented for the intransitive and transitive, we saw potential for a concise and accurate representation of the causative operation as a change from an intransitive verb to a transitive.

In the implementation described thus far, the S-pn and A-pn position classes will never apply to the same verb, since they have mutually exclusive requirements for their inputs. The S-pn position class takes only uninflected intransitive verb stems as input, while the A-pn position class takes inflected transitive verbs from the dual-transitive position class as input. Although they fill different semantic roles in this language, S and O arguments share a number of properties. They have the same orthographic forms for both full noun phrases and pronouns, as well the same absolutive verbal agreement prefix paradigm across person and number, as seen in Tables 1 and 2. The S-pn and O-pn position classes are also the first that the verb stem goes through — for intransitive and transitive verbs, respectively.

We analyzed the Panãra causative as a "switch" midway through the verbal inflection from the intransitive verb's chain of position classes to the transitive verb's chain. This switch is triggered by the *ho-* morpheme, which occurs between the S-pn and dual position classes. Figure 2 offers a visual representation of this analysis.

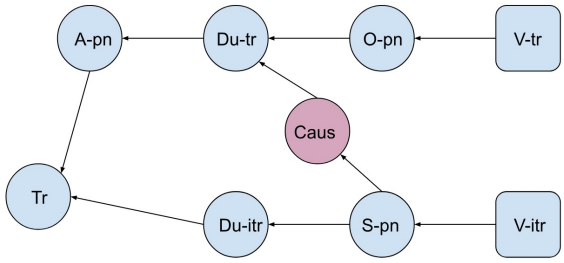

Figure 2: Panãra's causative as a "switch" midway through the verbal inflection

To model this phenomenon, we added a new optional verbal inflection position class called Causative. This position class takes as its input the output of the S-pn position class, and appears as the prefix *ho-* on the left edge of the verb complex. There is one LRT in this position class, which specifies that the subject must have ergative case. To employ the work of the valence-changing operation library (Curtis 2018), we added a subject-adding valence-changing operation to the lexical rule. The feature structure in Figure 3 illustrates this inflectional rule.

[6] Although in descriptive work, the S and O verbal prefixes are grouped together in the absolutive, it is necessary for us to model them with separate position classes here. The LRTs in the S-pn position class constrain agreement information on the *subject* while those in the O-pn position class do so for the *object*. Therefore, in an HPSG analysis, they are not the same morpheme.

[7] A-pn is the name of the position class for the verbal morpheme that agrees with the agent in person and number. The same naming pattern extends to the O-pn and S-pn position classes, which agree with the person and number information of the object and subject, respectively.

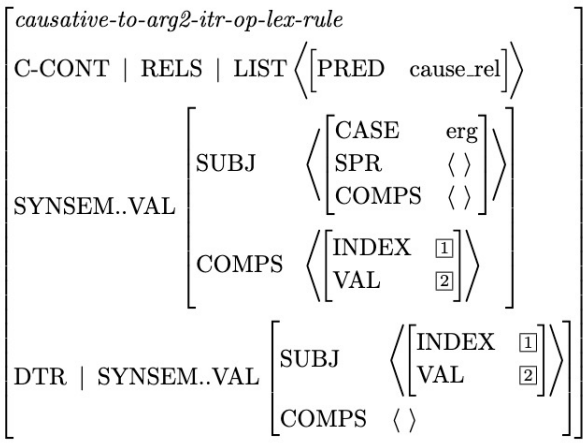

Figure 3: Subject-adding valence-changing inflectional rule

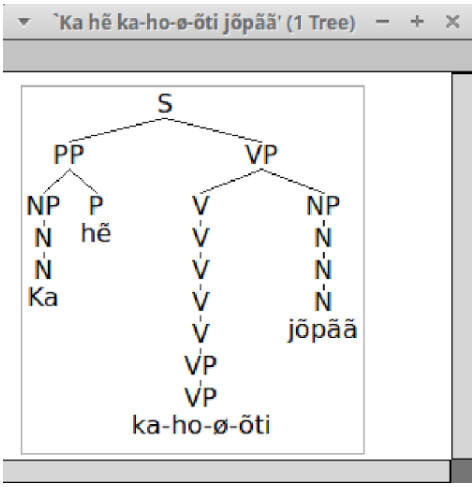

Figure 4: LKB output: Parse tree of sentence (7)

Information for both the valency and index is copied from the daughter's SUBJ list to the output's COMPS list. The resulting subject must have ergative case, as is consistent with transitive subjects. The non-empty SUBJ and COMPS lists in the output ensure the desired valency. Finally, this rule contributes a PRED value of **caus_rel** in the MRS (Copestake et al. 2005).

4.3 Validation

The analysis and corresponding implementation detailed above produces the desired behavior for causative constructions, as well as rules out ungrammatical structures.

After going through the Causative position class, and switching to the transitive path of position classes, the argument marked by the prefix closest to the root can no longer be interpreted as an S, but rather must be O. Consequently, the inflected verb is prevented from incorrectly taking the intransitive prefix *jy-*, which can only appear when the subject is absolutive, per the constraints of the LRTs in the Transitivity position class. The parse tree in Figure 4 illustrates a successful implementation of our analysis, using sentence (7)[8] as an example.

[8]The original data from (Bardagil 2018) was slightly altered to remove the possessive relation of 'your child' and instead use 'the child' because inalienable possession was not implemented in the grammar.

(7) Ka hẽ kahosõti jõpãã.
Ka hẽ ka-ho-ø-sõti jõpãã
2SG ERG 2SG.A-CAUS-3SG.O-sleep child
'You made the child sleep.'
(Based on Bardagil 2018: 108)

The default MRS of a causative structure output by the valence-changing operation library (Curtis 2018) produces a valid semantic representation for the sentence in Figure 5.[9] The Causative position class contributes a cause relation to the RELS list. The ARG0 of the O argument, **jõpãã_n_child_rel**, is identified with the ARG1 of the verb and the ARG2 of the cause relation. The ARG0 of the pronominal A argument is identified with the ARG1 of the cause relation.

5 Implementation

Using the LinGO Grammar Matrix (Bender, Flickinger & Oepen 2002) as a starting point, we modeled the grammar as described by Bardagil (2018) with the Grammar Matrix Customization System (Bender et al. 2010). We curated a test suite containing grammatical and ungrammatical sentences in the language over the course of 10 weeks. The final test suite held 230 items covering each of the grammatical features modeled — one of which is valence adjusting operations.

[9]This analysis treats **cause_rel** as a three-place relation. A two-place relation could also be implemented, if further empirical work with Panãra speakers shows it is more correct.

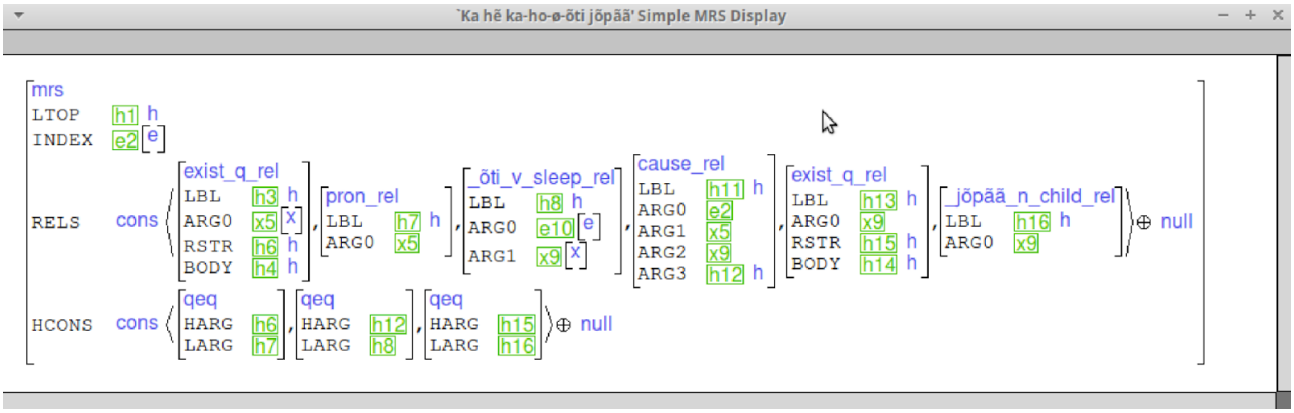

Figure 5: LKB output: MRS representation of sentence (9)

We used the [incr tsdb()] grammar profiling software (Oepen & Flickinger 1998) for measuring the grammars' coverage and overgeneration. Over 127 grammatical testsuite items, our grammar had 84.3% coverage (107/127). Over 103 ungrammatical testsuite items, it has 13.6% overgeneration (14/103). The average number of parses per parsed item was 1.44.

It is important to note that 142 of these testsuite items were examples that we constructed in order to isolate specific phenomena of interest and to include only phenomena that could be handled by the grammar during its incremental development. Each author-constructed example is based on the data and analysis from Bardagil (2018), but has not yet been vetted by speakers of the language.

There were 8 testsuite items (3 grammatical and 5 ungrammatical) constructed specifically to test the valence-changing operation analysis and implementation. On these items, the grammar had 100% coverage and 0% overgeneration; our implementation was successful, with no added ambiguity.

6 Conclusion

The support for modeling morphology, including valence changing morphology, based on the notion of position classes in the Grammar Matrix customization system, correctly predicts the interaction of morphemes for causative constructions in Panãra. The implementation of chained verbal inflection position classes realized our analysis of causatives as a jump between the inflection patterns of intransitive to transitive verbs. Since the O-pn position class is incompatible with intransitive verbs and attaches before the causative morpheme, we take advantage of the orthographic and syntactic parallels between S and O agreement affixes to create a construction with two core arguments and a Causative position class that moves the verbal inflection from intransitive rules to transitive rules. Thus, intransitive verb stems which encounter the causative are able to fully inflect with two argument agreement affixes and the correct transitivity and case-marking morphology. We believe this analysis and implementation succinctly represents the causative operation in Panãra.

References

Bardagil, Bernat. 2018. *Case and agreement in Panará*. University of Groningen dissertation.

Bender, Emily M., Scott Drellishak, Antske Fokkens, Laurie Poulson & Safiyyah Saleem. 2010. Grammar Customization. *Res. Lang. Comput.* 8(1). 23–72. https://doi.org/10.1007/s11168-010-9070-1.

Bender, Emily M., Dan Flickinger & Stephan Oepen. 2002. The Grammar Matrix: An Open-Source Starter-Kit for the Rapid Development of Cross-linguistically Consistent Broad-Coverage Precision Grammars. In *COLING-02: grammar engineering and evaluation*. https://aclanthology.org/W02-1502.

Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. CSLI Publications.

Copestake, Ann, Dan Flickinger, Carl Pollard & Ivan Sag. 2005. Minimal Recursion Semantics: An Introduction. *Research Language Computation* 3. 281–332. https://doi.org/10.1007/s11168-006-6327-9.

Curtis, Christian Michael. 2018. *A Parametric Implementation of Valence-changing Morphology in the LinGO Grammar Matrix*. University of Washington MA thesis.

Goodman, Michael Wayne. 2013. Generation of Machine-Readable Morphological Rules from Human-Readable Input. *University of Washinton Working Papers in Linguistics* 30. http://depts.washington.edu/uwwpl/vol30/goodman_2013.pdf.

Lapierre, Myriam. 2023. The Phonology of Panãra: A Segmental Analysis. *International Journal of American Linguistics* 89(2). 183–218. https://doi.org/10.1086/723642.

Oepen, Stephan & Dan Flickinger. 1998. Towards systematic grammar profiling: Test suite technology ten years after. *Computer Speech and Language: Special Issue on Evaluation* 12. 411–436.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press.

Zamaraeva, Olga, Chris Curtis, Guy Emerson, Antske Fokkens, Michael Goodman, Kristen Howell, T.J. Trimble & Emily M. Bender. 2022. 20 years of the Grammar Matrix: Cross-linguistic hypothesis testing of increasingly complex interactions. *Journal of Language Modelling* 10(1). 49–137. https://doi.org/10.15398/jlm.v10i1.292.