A Computational Account of Modality-based Case Frame Transformation

Timothy Baldwin, Takenobu Tokunaga, Hozumi Tanaka
Tokyo Institute of Technology

{tim,take,tanaka}@cs.titech.ac.jp

1 Introduction

Verb modality presents a major processing obstacle in any NLP application, and can be overcome either by exhaustively listing all possible modal variants of any given verb, or by deriving any modality-based valency alterations *in situ*. Given the combinatorial complexity of any exhaustive listing attempt, we adopt this second methodology in proposing a modular set of valency frame transformations to dynamically recreate the appropriate valency frame.

Modularity in transformation produces the advantage that any permutation of modal transformation types can be modelled, while maintaining linguistic accountability within the rule sets. The drawback of rigidly adhering to a modular rule set is that any ambiguity of transformation for the modal types must be reproduced in the associated rule sets. This is facilitated by allowing any given rule set to output multiple valency frame derivations, and having multiple valency frame derivation candidates cascade through the rule sets in parallel. Final selection of the appropriate valency frame is then made by way of case slot correspondences with input, and selectional restrictions on each case slot. Selectional restrictions are indexed from the corresponding case slot in the original valency frame, and are considered to be unaffected by the derivation process, except where marked otherwise in the rule sets.

The proposed case frame transformational rule sets are applied sequentially, according to the linear (left-to-right) ordering of modal auxiliaries suffixing the verb stem. To take an example of the transformation process, the verb *tabe-sase-rare-ru* "to eat-CAUSE-PASS-PRES"[1] would involve the application of the causative, followed by the passive, rule sets. In the case of a causativised passive such as *sika-rare-sase-ru* "to scold-PASS-CAUSE-PASS", this ordering would be reversed.

The principle mechanism called upon in implementing the transformational rule sets is grammatical roles, and hence some indication of the grammatical role of each case slot must be contained within the original valency frame. Within the rule sets, the only alterations made in transformation are to grammatical roles and associated case marking.

[1]The following abbreviations are utilised in this paper: PASS = passive, CAUSE = causative, PRES = non-past, NOM = nominative, ACC = accusative, DAT = dative, COM = comitative, GEN = genitive.

2 Passive

The *(r)are* (passive) affix is associated with **subject honorific**, **potential/spontaneous**, and **passive** usages (Jacobsen 1992:140-56). Additionally, passive voice occurrences are found in **true**/direct, **adversative**/indirect, and **anchor-topical** readings. Honorific usages bring about no change to the original case frame, the potential and spontaneous senses produce case marking alternation, and only the various passive types transform the case frame. Thus, the handling of the *(r)are* affix can be partitioned into two sub-tasks: identifying which of these four patterns of usage has occurred, and transforming/modifying the case frame accordingly. Fortunately, there is some scope to disallow types according to auxiliary verb collocation, as adversative passives and spontaneous usages are incompatible with other valency transformations, and the subject honorific and potential readings must occur as the final auxiliary verb in the verb compound. These collocational restrictions are hence applied as a pre-processing filter.

True passives correspond to the traditional passive transformation of Direct or Indirect Object onto the Subject case slot, and the original Subject to an optional Passive Agent case slot; in this, they can be considered as *valence-reducing*.

(1) *A-ga* *B-o* *yonda.*
A-NOM B-ACC called
'A called B.'

(1′) *B-ga* *(A-ni)* *yobareta.*
B-NOM A-DAT was called
'B was called (by A).'

Adversative passives, on the other hand, are *valence-increasing* in the sense that the original Subject retains its obligatory status under transformation into the Passive Agent case slot (Hoshi 1994), and a new affected Subject is introduced into the valency frame.

(2) *A-ga* *syorui-o* *otosita.*
A-NOM document-ACC dropped.
'A dropped the documents.'

(2′) *B-ga* *A-ni* *syorui-o* *otosareta.*
B-NOM A-DAT was dropped
'B was adversely affected by A dropping the documents.'

Meanwhile, anchor-topical passives are *valence-*

maintaining. That is, while the Subject is transformed into the optional Passive Agent case slot, an 'anchoring' entity prefixing either the Direct Object or Perlative is mapped into the Subject case slot, and the 'anchored' element is maintained within the original case slot. This process is apparent in (3), in which the anchor is *B* and anchored element *asi*:

(3) *A-ga* *B no asi-o* *hunda.*
A-NOM B GEN foot-ACC trod on.
'A stepped on B's foot.'

(3') *B-ga* *(A-ni)* *asi-o* *humareta.*
B-NOM A-DAT foot-ACC trodden on
'B had his/her foot trodden on (by A).'

The anchored element is typically a body part (e.g. *atama* "head"), kinship term (e.g. *kodomo* "child"), or local relational noun (e.g. *ue* "top").

2.1 Sense disambiguation of the *(r)are* affix

First, we pre-analyse the inflectional content of the main verb, to exclude illegal inflectional contexts of the various *(r)are* senses (see above). Next, we apply the rule set in Figure 1 to derive valency frames for the remaining legal senses. During final disambiguation of the valency frame, we then apply heuristics to complement sortal restrictions in weighting the plausibility of the various *(r)are* sense candidates.

For the *(r)are* rule set, rule subsets for each legal *(r)are* reading are applied independently (rule subsets in Figure 1 are delineated by the labels corresponding to the various *(r)are* senses). Stative case marking for the spontaneous/potential readings refers to alternative dative marking on the subject and nominative marking on the direct object, noting that the spontaneous/potential readings are obtainable only for transitive verbs.

The given rule ordering for the true passive rule subset stems from causativised case slots being most salient in the passivisation process.[2] Failing to detect a causativised case slot, the Direct and Indirect Object case slots become the next candidates for passivisation. The reader will notice that the rules for the object case slots are set up conditional independently, such that, for ditransitive verbs, two distinct valency frame candidates will be generated by the true passive rule subset, corresponding to Direct and Indirect Object passivisation, respectively. One additional passivisation phenomenon which is inherently modelled by the given passive rule subsets, is the adversative passive semantics of passivised intransitive valency frames, a result which obtains from the requirement for an object case slot in both the anchor-topical and true passive rule subsets.

Lexical phenomena which are considered to suggest true passive voice in the context of a *(r)are* affix, are case marking with the *-ni yori/yotte* particles (used to mark the Passive Agent for only true passives), quotative ('*to*') case marking, and collocation with the causative morpheme. Lexical evidence which supports equally the various passive readings is the existence of a datively marked animate argument (corresponding to the Passive Agent).

On the reverse side, honorific suffixes such as *denka* "his/her highness" and *sama* "sir/madam" on the subject filler tend to point to subject honorification.

[2]We do not discredit Matsumoto's (1996) observation that the patient of the base verb can be the passive subject for passivised coercive causatives, but simply lay claim to the statistical improbability of such an eventuality.

3 Causative

Inflectional causatives in Japanese are much less troublesome than passives, as the *(s)ase* causative auxiliary unambiguously marks instances of causativisation. The causativisation process involves the subject being moved into the Causee or Authorisee case slot, and an animate causer being added in the Subject position.

Within Japanese causatives, there are two widely accepted sub-types, **inducing** and **permissive** causatives, which differ in their implicit degree of coercion and approbation, respectively (Shibatani 1990). This distinction in the degree of coercion involved between the subject and object of causativisation is marked by the separate Causee and Authorisee case-roles, corresponding to inducing and permissive causation, respectively. Inducing causatives are characterised by accusative marking on the Causee case slot, in the absence of an accusatively marked Direct Object, whereas permissive causatives generally mark the Authorisee datively. However, overlap exists between these two case marking paradigms. For our purposes, therefore, we make no assumptions as to case marking tendencies, and allow both accusative (where syntactically acceptable) and dative marking of the object of causativisation for the two causative types.

3.1 Case frame transformation

The rule subsets for the inducing and permissive causative types are essentially identical, except for the distinct case-role mapping of the object of causativisation onto the Causee and Authorisee case-roles, respectively. In combination with other transformational rule sets, the separation of these case-roles translates to passivised causatives being assumed to correlate to inducing-causative sense, with the Authorisee case-role not being passivisable. Additionally, when used in tandem with *mora(-u)*, the Authorisee maps onto the Subject for permissive causatives, whereas inducing causatives maintain separate Causee and Subject case-roles.

4 Resultatives

Resultatives are produced by way of the *-te ar(-u)* construction, and indicate some resultant/perfective state of an action. Resultatives can be classified into two types: **valence-changing** ('V-C') and **valence-maintaining** ('V-M') (Hasegawa 1996:86). With V-C resultatives, the Direct Object of a transitive verb is moved into the Subject position, overwriting the original Subject (cf. (4)); V-M resultatives, on the other hand, produce no change in the valency frame (cf. (5)).

(R)are rule set:

HONORIFIC: Return valency frame unchanged.

SPONT/POT.: IF (valency frame contains Direct Object) THEN indicate verb as stative, and mark case marking alternations appropriately;
ELSE Fail.

ANCHOR-TOP.: IF (body part, kinship term or local relational noun constitutes legal case filler for Direct Object or Perlative case slot) THEN map Subject onto optional Passive Agent case slot AND add animate Subject case slot;
ELSE Fail.

ADVERS.: Map Subject onto obligatory Passive Agent case slot AND add animate Subject case slot.

TRUE: Map Subject onto optional Passive Agent case slot;
IF (Causee case slot in valency frame) THEN map Causee onto subject;
ELSE
IF (Direct Object in valency frame) THEN map Direct Object onto Subject AND output valency frame;
IF (Indirect Object in valency frame) THEN map Indirect Object onto Subject AND output valency frame.

Causative rule set:

INDUCING: Map Subject onto Causee case slot AND add animate Subject.

PERMISSIVE: Map Subject onto Authorisee case slot AND add animate Subject.

Resultative rule set:

V-C: IF (Direct Object in valency frame) THEN delete Subject AND map Direct Object onto Subject case slot;
ELSE Fail.

V-M: Return valency frame unchanged.

Age(-ru) rule set:

AGERU: Add optional Beneficiary case slot.

Kure(-ru) rule set:

KURERU: Add optional Beneficiary case slot.

Mora(-u) rule set:

MORAU: Map Subject onto optional Agent case slot;
IF (Authorisee case slot in valency frame) map Authorisee onto Subject;
ELSE IF (animate-type Indirect Object case slot in valency frame) map Indirect Object onto Subject;
ELSE IF (animate-type Direct Object case slot in valency frame) map Direct Object onto Subject;
ELSE Add animate Subject.

Figure 1: Transformational rule sets

(4) *rozyō-ni* *kuruma-ga* *tomete-aru.*
on road-DAT car-NOM is parked
"A car is parked on the road."

(5) *A-ga* *kuruma-o* *tomete-aru.*
A-NOM car-ACC is parked
(lit.) "The car is in the state of having been parked by A."

The rule-based handling of resultatives consists simply of mapping the Direct Object onto the Subject position (overwriting the original Subject) for V-C resultatives, and retaining the original composition of the valency frame for V-M resultatives.

5 Empathy constructs

Empathy constructs are commonly used in Japanese to indicate that entity which benefits from the action described, or for whose benefit the action is performed. While empathy constructs can occur as main verbs, this paper is concerned solely with their auxiliary usages.

5.1 *Age(-ru)*

The *-te age(-ru)* affix is used to indicate that the action described by the stem of the verb compound, is performed for the benefit of another entity, described in the (optional) Beneficiary case slot. For example, in the case of *asonde-ageru* "to play (with)", the act of 'playing' is performed either for the benefit of the co-player, or a third party Beneficiary, with the Co-actor and Beneficiary case-roles seen to have been conflated in the first interpretation. The empathy, however, is always on the Subject, as seen in the ungrammaticality of:

(6) **A-ga* *watasi-to* *asonde-ageta.*
A-NOM I-COM played
"A played with me." (intended)

Age(-ru) is implemented simply by adding an optional Beneficiary case slot. The optionality of the Beneficiary case slot is crucial because of the potential for conflation between any non-Subject complement case slot and the Beneficiary.

Clearly, simple addition of a Beneficiary case slot in the rule set implementation brings about the possibility for multiple occurrences of the Beneficiary case slot within a single clause. However, this is precisely the way that the Beneficiary works, in the case of complex empathy constructs:

(7) *A-ni* *hon-o* *yonde-agete-kureru?*
A-DAT book-ACC can (you) read (to)
"Can (you) do (me) the favour of reading a book to B."

5.2 *Kure(-ru)*

The *-te kure(-ru)* affix closely resembles *age(-ru)* in that an action is performed for the benefit of the Beneficiary, and differs only in that the empathy is on the Beneficiary rather than the Subject:

(8) **watasi-ga* *A-to* *asonde-kureta.*
I-NOM A-COM played
"I played with A." (intended)

Similarly to *age(-ru)*, *kure(-ru)* is implemented simply by adding an optional Beneficiary case slot.

5.3 *Mora(-u)*

The *-te mora(-u)* affix produces transformation in the valency frame, rather than simple addition of a Beneficiary case slot as occurred for the other two empathy construct types. Specifically, the Subject is moved into an optional Agent case slot, and any of the Authorisee, Indirect Object and Direct Object case slots are moved into the Subject position, failing which a new animate Subject is added.

The strict preference for the Authorisee over other case-roles stems from the semi-idiomatic status of *-sasete-mora(-u)* "allow (me) to" in the permissive causative sense. Preference for animate Indirect Objects over animate Direct Objects is intended as a reflection of the inability to transform the Direct Object case slot for *syōkai(-suru)* "t"o the Subject position.

6 Conclusions

This paper was dedicated to the description of a modular rule set to model valency frame transformations arising form verb modality. Rule sets associated with each verb modal type are applied linearly, and any occurrences of analytical ambiguity are cascaded in parallel through the full transformation process. Final selection of the single most appropriate valency frame is made both heuristically and according to sortal preferences indexed from the various valency frame candidates.

The system of rule sets is currently implemented as component of a Japanese relative clause analyser (Baldwin 1998), although independent evaluation is left as an issue for future research.

References

BALDWIN, T., 1998. *The Analysis of Japanese Relative Clauses.* Master's thesis, Tokyo Institute of Technology.

HASEGAWA, Y. 1996. *A Study of Japanese Clause Linkage.* CSLI.

HOSHI, H., 1994. *Passive, Causive, and Light Verbs: A Study of Theta Role Assignment.* University of Connecticut dissertation.

JACOBSEN, W. M. 1992. *The Transitive Structure of Events in Japanese.* Kurosio Publishers.

MATSUMOTO, Yo. 1996. *Complex Predicates in Japanese: A Syntactic and Semantic Study of the Notion 'Word'.* CSLI & Kurosio Publishers.

SHIBATANI, M. 1990. *The Languages of Japan.* CUP.