Representing the Continuum between Arguments and Adjuncts within Predicate-Frames

Pierre Marchal, Thierry Poibeau, Yves Lepage

Representing the Continuum between Arguments and Adjuncts within Predicate-Frames

Pierre Marchal[1, 2], Thierry Poibeau[3], Yves Lepage[2]

[1] ER-TIM, INaLCO, Paris, France

[2] IPS, Waseda University, Kitakyûshû, Japan

[3] LaTTiCe, CNRS – ENS – Université Paris III, Montrouge, France

pierre.marchal@inalco.fr, thierry.poibeau@ens.fr, yves.lepage@waseda.jp

Introduction

In the context of an automated task of acquisition of predicate-frames in Japanese, we introduce a method to observe and evaluate distinction criteria between arguments and adjuncts on a large scale. Rather than a binary classification, we are interested in a representation of the continuum between arguments and adjuncts.

complements = arguments ∪ adjuncts

Method and resources

We aim at modeling the two extremes of the continuum between arguments and adjuncts.

Prototypical argument: a type of complement which appears in every predicate-frame (*i.e.* usage) of a given verb.

Prototypical adjunct: a type of complement which never appears in any predicate-frame of a given verb.

Here, the goal is to build two list of verbs, one list to retrieve prototypical arguments, one another to retrieve prototypical adjuncts. We rely on an existing resource: 動詞項構造シソーラス. We consider two case markers: *wo* (prototypical argument), and *de* (prototypical adjunct).

	Verbs	Predicate-frames
wo-list of verbs	2,560 49 %	4,761 46 %
de-list of verbs	4,954 95 %	9,416 91 %
動詞項構造シソーラス	5,190 100 %	10,364 100 %

Table 1. Comparison of the two lists of verbs with 動詞項構造シソーラス

For each list of verbs we retrieve instances of predicate-frames. We process about 1.9M sentences from the BCCWJ with a dependency parser to build examples as in figure 1.

[...] とてもうれしそうな顔で私を見下ろしています。
[...] he is looking down at me with such a happy face.
↓
見下ろす　デ格:顔　ヲ格:私　アスペクト:テイル　101728176

Figure 1. Instance of a predicate-frame

We build two lists of examples. In the *wo*-list, every complement marked with the particle *wo* is considered as an argument. In the *de*-list, every complement marked with the particle *de* is considered as an adjunct.

	Count	Arguments / adjuncts
wo-list of examples	1,041,818	504,391 (*wo*)
de-list of examples	1,890,151	144,481 (*de*)

Table 2. The two lists of examples

Experiment

We apply two existing argument/adjunct distinction criteria to our lists of examples to compute a measure of autonomy expressing the placement of complements along the continuum. The autonomy measure ranges from 0 to 1. 0 corresponds to an argument.1 corresponds to an adjunct.

Ordering: arguments tend to be closer to the verb than adjunct.

$$autonomy_{ordering}(c) = \frac{distance(c)}{number\ of\ complements}$$

Head dependence: arguments tend to cooccur with a narrower range of verbs than adjuncts.

$$autonomy_{head\_dependence}(c) = \frac{productivity(c)}{count(c)}$$

Complements are sorted according to their degree of autonomy toward the verb.

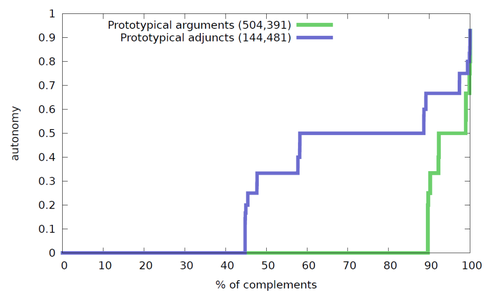

Figure 2. Ordering

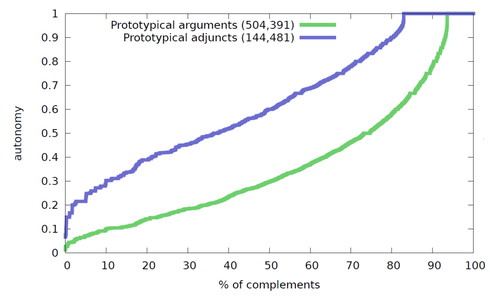

Figure 3. Head dependence

With both criteria it appears that prototypical arguments tend to have a lower autonomy than prototypical adjuncts.

Evaluation

We evaluate distinction criteria on 31,531 examples, all containing at least one complement marked with the *wo* particle and one complement marked with the *de* particle.

We add an extra criterion which is the combination of ordering and head dependence (*i.e.* the product of their degree of autonomy).

Our evaluation method rely on the idea of a continuum between arguments and adjuncts. For a criterion to be relevant to this task, it needs to order properly the complements along the continuum. That is, even if a criterion assigns a low autonomy to an adjunct it can still be accurate if it assigns – in the same context – an even lower autonomy to an argument (as shown in table 3).

	ヲ格 (*wo*)	デ格 (*de*)
Ordering (C1)	0	0.5
Head dependence (C2)	0.29	0.32
C1 + C2	0	0.16

Table 3. Evaluation on the example of figure 1

It appears that accuracy can be slightly improved by combining different criteria.

	Accuracy
Ordering (C1)	81.40
Head dependence (C2)	75.07
C1 + C2	**81.72**

Table 4. Results of the evaluation

NINJAL International Symposium on Valency Classes and Alternations in Japanese, Tôkyô, Japan, 2012