The PAPPI System: Lexical Semantics and Morpho-Syntax

Sandiway Fong
sandiway@research.nj.nec.com
NEC Research Institute, Inc.
4 Independence Way, Princeton NJ

PAPPI is a multilingual parsing engine in the Principles-and-Parameters framework (Chomsky, 1981), initially developed at MIT (Fong, 1991) and greatly expanded upon at NEC. Written in Prolog and supplanted with a set of grammatically-relevant primitives with a view to ready modification, it has been used for exploring parameterized syntax across a variety of languages. It is a freely downloadable system available to others for research and pedagogical purposes.[1] For example, recent and on-going collaborative projects have included Turkish (Birtürk, 1988), Chinese (Lin et al., 1998), Japanese[2] and Norwegian[3]. We highlight the two current projects.

1 NECI Lexicon Project

This is a project to build a large-scale English verb lexicon using semantic templates inspired by (Rappaport Hovav and Levin, 1998). The central idea is to start from a representation of verbs based on event templates, and to provide a formal mechanism for deriving subcategorization frames for use in parsing. Current work is focused on secondary predication and verb alternations (Fong et al., 2000a), also (2000b); in particular, the resultative construction. As (1) illustrates, activity verbs like *wipe* may take adjectival phrase (AP) resultatives expressing change of state:

(1) (a) John wiped the table
(b) John wiped the table clean

Alternations involving non-subcategorized arguments, i.e. normally oblique arguments that may appear in direct object position in conjunction with the original (demoted) object, as shown in (2), are another topic of investigation:

(2) (a) Mary cleared the driveway
(b) Mary cleared the driveway of leaves
(c) Mary cleared the leaves off the driveway

In the event template framework, both resultatives and shadow arguments are introduced via secondary templates. For example, *wipe* and *clear* in (1a) and (2a) have the primary, or basic, templates (3a) and (3b), respectively. Here, x encodes the agent, y a location, and z the optionally-realized shadow.[4] The difference in templates expresses the fact that de-adjectival *clear* encodes final state, whereas *wipe* is a manner verb.[5] *Clean* and *off* come with their own templates (3c), encoding change of state, and (3d), encoding material (z) and location (y), respectively:

(3) (a) x act$_{<wiping>}$ on y
(b) x caus y become <*clear*> (remove z)
(c) y be <clean>
(d) z off y

Templates are combined in (4a) (for (1b)) and (4b) (for (2c)), with the constraint that elements y and z of secondary templates must be checked off, or identified, against corresponding primary elements:

[1] The PAPPI project homepage is at `http://www.neci.org/homepages/sandiway/pappi/`.

[2] KUIS COE project: Researching and verifying an advanced theory of human language. `http://coe-sun.kuis.ac.jp/coe/coee.htm`

[3] Theoretical Norwegian Grammar Simulator (TGR-SIM), University of Bergen, Norway. Johnsen, L., A. Hestvik, S. Fong and D. Lebeaux. `http://vega.uib.no/tgr-sim/index.html`

[4] z is unrealized in (2a), realized in (2b).

[5] This explains why *clear* cannot accept a resultative: **Mary cleared the driveway clean.*

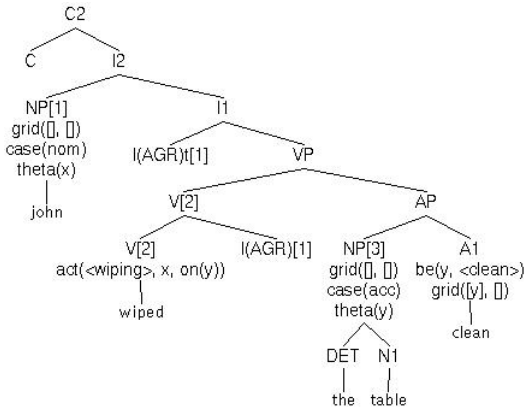

Figure 1: John wiped the table clean

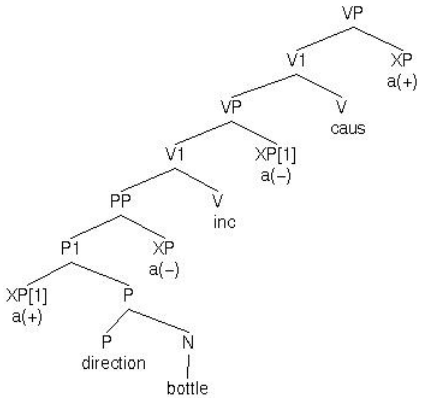

Figure 2: Bottle

(4) (a) x act$_{<wiping>}$ on y & y be <clean>
(b) x caus y become <*clear*> (remove z) & z off y

The check-off mechanism is invoked when the phrase representating the secondary template is combined with the verb phrase (VP) shell during parsing, e.g. in Figure 1 secondary *y* is checked off at the juncture between the AP and the lower VP.

2 The Syntax of Words

This is a project on derivational morphology. The LR(1) bottom-up parser underlying PAPPI can be used for near-linear-time morphological processing in the Strict Asymmetry framework based on X′-structure (Di Sciullo and Fong, 2000). Examples of both overt and zero-affix examples are given in (5):

(5) (a) V:*simplify*: A:*simple* + inc (*i*) + caus (*fy*)
(b) V:*formalize*: N:*form* + A:*al* + inc (*i*) + caus (*ze*)
(c) A:*formalizable*: N:*form* + A:*al* + inc (*i*) + caus (*ze*) + A:*able*
(d) N:*employer*/*-ee*: V:*employ* + N:*-er*/*-ee*
(e) V:*bottle*: N:*bottle* + P:*en* + inc + caus

In the X′-selection framework, morphemes head full X′-phrases containing both a specifier and a complement. Figure 2 illustrates the parse obtained for the verb *bottle*. Beginning with the simple noun *bottle*, a (non-overt) directional prefix *-en* is layered on top, providing an argument `a(+)` object position.[6] The causative/inchoative (`caus`/`inc`) complex selects for the object and provides a second `a(+)`-position for the agent. Thus transitive *bottle*, as in *John bottled the wine*, selects for two arguments with the interpretation that the wine was placed into the bottle.

[6] In Italian and French, this directional prefix is overt, e.g. *im*bottigliare and *em*bouteiller.

References

A. Birtürk. 1988. *A Computational Analysis of Turkish Using the Government-Binding Approach*. Ph.D. thesis, Department of Computer Engineering, METU, Ankara, Turkey.

N.A. Chomsky. 1981. *Lectures on Government and Binding*. Foris Publications.

A.-M. Di Sciullo and S. Fong. 2000. Morpho-syntax parsing. Technical Report 2000-078, NEC Research Institute, Princeton, New Jersey.

S. Fong, C. Fellbaum, and D. Lebeaux. 2000a. Distinguishing direct objects: shadows and ghosts. Technical Report 2000-053, NEC Research Institute, Princeton, New Jersey.

S. Fong, C. Fellbaum, and D. Lebeaux. 2000b. Semantic templates and transitivity alternations in the lexicon. In *Proceedings of TALN 2000, EPFL, Lausanne, Switzerland*.

S. Fong. 1991. *Computational Properties of Principle-Based Grammatical Theories*. Ph.D. thesis, Artificial Intelligence Laboratory, MIT.

K. H.-C. Lin, V.-W. Soo, and S. Fong. 1998. Dealing with nominalizations in Mandarin Chinese using a principles and parameters parser. *Computer Processing of Oriental Languages*, 11(3).

M. Rappaport Hovav and B. Levin. 1998. Building verb meanings. In Butt and Geuder, editors, *The Projection of Arguments: Lexical and Compositional Factors*. CSLI Lecture Notes No. 83.