Subcategorisation frame acquisition, selectional preferences and alternations

Timothy Baldwin and Dominic Widdows
Subcategorisation frames

A subcategorisation (subcat) frame is a statement of what types of arguments a verb ... takes as objects, infinitives, *that*-clauses, participal clauses and subcategorised PPs (Manning 1993):

*John wants Mary to be happy*
*John hopes that Mary is happy*
*John wants that Mary is happy*
*John hopes Mary to be happy*
Applications of subcat information

• Subcat information can lead to attachment disambiguation:

  John put [the cactus] [on the table]

• Core component of type hierarchy in linguistically-precise grammars

• Empirical evidence for lexicalised subcat information improving the performance of statistical parsers, WSD systems, information extraction engines, etc.
From Grammar to Lexicon: Unsupervised Learning of Lexical Syntax

Michael R. Brent
Computational Linguistics 19(2)

14 November, 2003
Basic method

1. Identify verb tokens through a variety of heuristics

2. For each verb type, use high-precision lexico-syntactic patterns to identify evidence for 6 different subcat frames

3. Use a statistical filter to remove noise in the extracted subcat data
Identification of verb tokens

• Very rough and heuristicky — (just) before the days of reliable POS tagging

• Focus on base and present participial verb forms

• Problems in distinguishing between base-form verbs and singular nouns (e.g. record — only workaround a filter on the immediately preceding word)
Lexico-syntactic patterns

- Based on closed-class words (pronouns, determiners, complementisers, auxiliaries, punctuation)
- NPs captured in the form of pronouns or sequences of capitalised words
- VPs based on auxiliaries and the verbs learned in step 1
Statistical filtering (1)

- Assumption that the probability of false evidence for a given subcat frame $S$ (e.g. transitive) occurring is equal for all verbs incompatible with $S$ (e.g. snore, put, say, ...)

- NOTE: probability of false evidence ($\pi_{-S}$) constant for a given $S$ but varies across different subcat frames

- Null hypothesis: the verb does not belong to subcat class $S$, i.e. it is $\neg S$
Statistical filtering (2)

- **Binomial test:** the probability of an event with probability $p$ occurring exactly $m$ out of $n$ times is given by

$$P(m, n, p) = \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m}$$

- The probability of the event occurring $m$ or more times out of $n$ is given by

$$P(m+, n, p) = \sum_{i=m}^{n} P(i, n, p)$$
<table>
<thead>
<tr>
<th>$\frac{m}{n}$</th>
<th>$P(m, n, p = 0.1)$</th>
<th>$P(m+, n, p = 0.1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0/10</td>
<td>0.349</td>
<td>1.000</td>
</tr>
<tr>
<td>1/10</td>
<td>0.387</td>
<td>0.651</td>
</tr>
<tr>
<td>2/10</td>
<td>0.194</td>
<td>0.264</td>
</tr>
<tr>
<td>3/10</td>
<td>0.057</td>
<td>0.070</td>
</tr>
<tr>
<td>4/10</td>
<td>0.011</td>
<td>0.013</td>
</tr>
<tr>
<td>5/10</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>6/10</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>7/10</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>8/10</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>9/10</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>10/10</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Statistical filtering (3)

- Given $n$ and $p \ (= \pi_{-S})$, we can apply a threshold $\theta$ to determine $m$ such that verbs which occur with subcat frame $S$ at least $m$ times can be classified as $+S$ with $(1 - \theta)$ confidence.

- In practice we don’t know $\pi_{-S}$ for each subcat frame $S$.

**SOLUTION:** set $\theta$ and $n$, and estimate $p$ based on the histogram distribution around each $m$; select the $p$ which best fits the binomial distribution.
Estimating $\pi - S$

1. For a sample of verbs and $N$ occurrences of each, count the number of times each verb occurs with a particular subcat frame

2. For each $j = 0, 1, ..N$ estimate $p$ as:

$$\sum_{i=0}^{j} H[i] \frac{i}{N} \text{ where } H[i]' = \frac{H[i]}{\sum_{i=0}^{j} H[i]}$$
and estimate the fit by way of:

\[
Fit(j) = \sum_{i=0}^{N} (P(i, N, p) - \text{Obs}(i))^2
\]

where

\[
\text{Obs}(i) = \begin{cases} 
H'[i] & \text{if } i \leq j \\
0 & \text{otherwise}
\end{cases}
\]

3. Set $\pi_S$ to the value $p$ which produces the best fit (lowest value)
Shortcomings of the Brent approach

- Assumption of $\pi_{-S}$ being equal for all verbs given a class $S$ shown to be flawed due to verb detection method
- Applicability of method to low-frequency words
- Scalability of method to other subcat frames
An update on more recent research

• Greater coverage of subcat frames (up to 160)

• Simple frequency shown to be at least as effective as binomial test at filtering out noise

• Verb sense shown to interface closely with subcategorisation properties

• AND YET the Brent method still has remarkable currency to this day
Open questions

- How to deal with low-frequency occurrences of subcat frames
- How well do the proposed methods port to other word classes (adjectives, nouns, ...) and languages
- Challenges for subcat acquisition in pro-drop languages (e.g. Japanese)
Generalizing Case Frames Using a Thesaurus and the MDL Principle

Hang Li and Naoki Abe
Computational Linguistics
The problem ...

Generalize from corpus examples of ‘x fly/flies’ to work out what class of things are likely to fly in new examples

Useful for

- Semantic modelling
- Disambiguation (lexical and structural)
- Recognizing metaphors?
The MDL principle

Description length = model DL + data DL

‘Simplicity vs Accuracy’

• Too much ‘accuracy’ to the data already seen results in overfitting

• How can we compress / generalize this distribution?
Tree cut models

- Use a taxonomy / tree structure (‘thesaurus’)

- Label the nodes that occur in the corpus examples with their corpus frequencies

- Try to generalize to nodes above, see if this distorts the model

- Lots of examples in the paper
PP attachment disambiguation

• Interesting problem which combines lexical and structural properties

• *saw the star with the telescope* vs. *saw the elephant with the trunk*

• *telescopes* attach to *seeing* but *trunks* attach to *elephants*
Lexical ambiguity

Related question: How do we cope with the ambiguity of *trunk*, which can also be an ‘artifact’ like telescope?

In paper, probability mass is distributed among possible senses

- Selectional preferences can be used as a guide to disambiguation

- Can problems become solutions?
Generalization and ‘disjunction’

• Every ‘logic’ is a ‘lattice’

• Disjunction $\iff$: generalization / union / ‘join’

• In Boolean logic, every set union is also a set

• In other logics (eg. quantum logic), disjunction also brings in other elements (and sometimes overgeneralizes)

• MDL tree cuts are ‘somewhere in between’
Alternations
Definition of alternation

- A regular mapping between argument positions in subcategorisation frames (generally assuming preservation of case-roles)

- Alternations involve *at least* one of:
  1. word order/(prepositional, case, etc.) marking variation between corresponding case slots
  2. case slot deletion
  3. case slot insertion
Example English alternations

(1) Kim loaded the truck with hay
    Kim loaded hay on the truck  \textit{Spray/load}

(2) Kim sold the car to Sandy
    Kim sold Sandy the car \textit{Dative}

(3) The dog walks
    Kim walks the dog \textit{Causative}

(4) Kim sliced the meat
    The meat sliced easily \textit{Middle}
Example Japanese alternations

(5) Kim-ga doa-o akeru / doa-ga aku
Kim-NOM door-ACC opens door-NOM opens
‘Kim opens the door’ ‘The door opens’

(6) Kim-ga doa-o hiraku / doa-ga hiraku
Kim-NOM door-ACC opens door-NOM opens
‘Kim opens the door’ ‘The door opens’

(7) Kim-ga doa-o akeru / doa-ga ake-rareru
Kim-NOM door-ACC opens door-NOM opens-PASS
‘Kim opens the door’ ‘The door is opened’
Types of alternations (1)

- **Analytical/diathesis**: alternation unmarked on the verb (e.g. *hiraku* “open\textsubscript{trans}” / *hiraku* “open\textsubscript{intrans}”)

- **Lexical**: alternation marked on the verb stem by predictable lexical variation (e.g. *akeru* “open\textsubscript{trans}” / *aku* “open\textsubscript{intrans}”)

- **Synthetic**: alternation marked by verbal inflection or a verb morpheme (e.g. *taberu* “eat” / *tabe-saseru* “make eat”)

14 November, 2003
Types of alternations (2)

- **Cognitive**: distinct verb forms but regularised pattern of alternation/simple change in focus, empathy, etc. (e.g. kau “buy” / uru “sell”)

- Focus on analytical, lexical and synthetic in this research
Alternations and verb semantics

• Verbs with similar alternation behaviour shown to cluster together semantically

• Semantically-similar verbs shown to alternate similarly
Alternation-based Lexicon Reconstruction

Timothy Baldwin and Francis Bond
TMI 2003
Basic Method

- Use selectional preferences to automatically extract alternations from a Japanese-English valency dictionary

- **Underlying hypothesis:** selectional preferences on alternating slots are the same

- Focus on Japanese verbs

- Analyse both the success of the method and what alternations we unearth
The bigger picture

- Move from a flat Japanese–English transfer dictionary to a hierarchical, language-modular dictionary structure
- In each monolingual lexicon, maximise structure sharing through analysis of alternations
- Assume no pre-defined alternation set (cf. EVCA), no supervision in alternation extraction
Subcat acquisition, selection preferences and alternations

p_1: A gathers B in C

p_2: D recruits E

p_3: F gathers G in H

p_4: I gathers in J
Source dictionary

- Goi-Taikei Japanese–English valency dictionary

- Valency frame described in form of case frame headed by verb

- Each case slot annotated with:
  - set of prototypical case markers
  - POS (NP or S)
  - set of selectional restrictions (→ Goi-Taikei thesaurus)
  - set of lexical fillers

14 November, 2003
Constraints on alternations

1. The selectional restrictions and lexical fillers on matching case slots are preserved under alternation

2. Alternations are monotonic in valency terms

3. A given alternation type has fixed direction: assume valency decreasing, and normalise direction alphabetically for valency-maintaining alternations (over-constraint ▼)
Extraction procedure

1. Generate all legal alternation candidates for each case frame pairing \((S, T)\) where \(S\) and \(T\) share some common kanji prefix

2. Score each, and return the highest scoring from among them

3. Accept only non-negatively-scoring alternations

4. In case of tie, select that alternation that preserves case marking the most
Dealing with inconsistency

- Possibility of annotational inconsistency between case frames:
  - non-preservation of selectional restrictions
  - unmotivated variation in case marking
  - unmotivated variation in lexical fillers

- Want to be able to deal with these robustly, and use such information in enhancing base lexicon
Scoring alternations

• Score linked case slots $S$ and $T$ according to their relative *conceptual cohesion*:

$$
\text{cohesion}(n_q) = -\log P(n_q) = -\log \frac{\sum_{\text{lex}_{p,i} \in n_q} \text{freq}(\text{lex}_{p,i})}{\sum_{\text{lex}_{p,i} \in n_o} \text{freq}(\text{lex}_{p,i})}
$$

$$
\text{classmatch}(n_j, n_k) = 3 \cdot \text{cohesion}(\text{sub}(n_j, n_k)) - \text{cohesion}(n_j) - \text{cohesion}(n_k)
$$

• Sum up the individual scores

14 November, 2003
\[
\text{classmatch}(a, a) = 3 \times 5.0 - 5.0 - 5.0 = 5.0 \\
\text{classmatch}(a, c) = 3 \times 0.9 - 0.9 - 5.0 = -3.2 \\
\text{classmatch}(a, b) = 3 \times 0 - 1.0 - 5.0 = -6.0
\]
Evaluation

- From a total of 13,880 valency frames, 2,777 alternation tokens were detected
- In case of multiple derivational analyses for given valency frame, take only the best-scoring alternation
- 1,553 alternations tokens incorporated into revised lexicon (373 types)
- Predominantly analytical alternations extracted
The top 10 extracted alternations are as follows:

<table>
<thead>
<tr>
<th>Index</th>
<th>Case slot mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$(NP_1{ga} \rightarrow \phi) \quad (NP_2{o} \rightarrow {ga})$</td>
</tr>
<tr>
<td>2</td>
<td>$(NP_1{ga}) \quad (NP_2{o} \rightarrow \phi)$</td>
</tr>
<tr>
<td>3</td>
<td>$(NP_1{ga} \rightarrow \phi) \quad (NP_2{o} \rightarrow {ga})$ $(NP_3{ni})$</td>
</tr>
<tr>
<td>4</td>
<td>$(NP_1{ga} \rightarrow \phi) \quad (NP_2{o} \rightarrow {ga})$ $(NP_3{ni, e})$</td>
</tr>
<tr>
<td>5</td>
<td>$(NP_1{ga}) \quad (NP_2{o} \rightarrow \phi)$ $(NP_3{ni} \rightarrow {o})$</td>
</tr>
<tr>
<td>6</td>
<td>$(NP_1{ga}) \quad (NP_2{o})$ $(NP_3{ni} \rightarrow \phi)$</td>
</tr>
<tr>
<td>7</td>
<td>$(NP_1{ga}) \quad (NP_2{o} \rightarrow {kara, yori})$</td>
</tr>
<tr>
<td>8</td>
<td>$(NP_1{ga} \rightarrow \phi) \quad (NP_2{o} \rightarrow {ga})$ $(NP_3{to, ni})$</td>
</tr>
<tr>
<td>9</td>
<td>$(NP_1{ga}) \quad (NP_2{ni} \rightarrow {o})$</td>
</tr>
<tr>
<td>10</td>
<td>$(NP_1{ga} \rightarrow \phi) \quad (NP_2{o} \rightarrow {ni})$ $(NP_3{de} \rightarrow {o})$</td>
</tr>
</tbody>
</table>
# Breakdown into different alternation types

<table>
<thead>
<tr>
<th>Index</th>
<th>Total score</th>
<th>Analytical</th>
<th>Lexical</th>
<th>Synthetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>417.3</td>
<td>86</td>
<td>34</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>127.8</td>
<td>54</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>110.4</td>
<td>12</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>97.7</td>
<td>12</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>94.7</td>
<td>32</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>93.8</td>
<td>32</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>77.6</td>
<td>16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>74.4</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>71.0</td>
<td>14</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>60.9</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Reflections

• Proposed method shown to be effective in extracting out valid alternations

• Little sense of recall (although not necessarily important for the dictionary reconstruction process)

• Possibility for using translation information to improve the accuracy of the extraction method
A general feature space for automatic verb classification

Eric Joanis and Suzanne Stevenson
EACL 2003
Basic method

- Use alternations and general verbal features to classify verbs according to Levin (1993) classes
- Dodge the issue of alternation detection or subcat acquisition by relying on features which capture alternation effects only indirectly
- Supplement alternation-based features with various weak lexical semantic indicators
### Features

<table>
<thead>
<tr>
<th>Feature Category</th>
<th># Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntactic slots</td>
<td>76</td>
</tr>
<tr>
<td>Slot overlap</td>
<td>40</td>
</tr>
<tr>
<td>“Empty” words</td>
<td>4</td>
</tr>
<tr>
<td>Passive</td>
<td>2</td>
</tr>
<tr>
<td>POS of the verb</td>
<td>6</td>
</tr>
<tr>
<td>Aux, modal, Adv</td>
<td>13</td>
</tr>
<tr>
<td>Derived forms</td>
<td>3</td>
</tr>
<tr>
<td>Animacy of NPs</td>
<td>76</td>
</tr>
</tbody>
</table>
Syntactic slot-based features

• Frequency of different syntactic slots occurring with a verb (includes PPs, conditioned on P)

• Degree of lexical overlap between syntactic slots known to alternate

• Expletive pronouns / there
Tense, voice and aspect features

- Relative frequency of passivisation
- POS (tense) of the verb
- Relative occurrence with modals/adverbials
- Relative occurrence in derived forms
Animacy feature

- Relative occurrence of animate fillers (personal pronouns, person names) in each of the syntactic slots
Task

- 2/3-way classification of a range of verb classes:
  - benefactive vs. recipient verbs
  - admire vs. amuse verbs
  - run vs. sound emission verbs
  - cheat vs. cheat/steal verbs
  - wipe vs. cheat/steal verbs
  - spray/load vs. fill vs. other put verbs
  - run vs. change of state vs. object drop verbs

- Also combined multi-way tasks
Experiments

• Feature values extracted from BNC (parsed with SCOL)

• Focus on verbs which occur $> 100$ times in the BNC in only one of the classes under consideration (with the predominant sense), and which are not excessively polysemous

• C5.0 used as learner (decision tree-based)

• Varied results were obtained
Reflections

• General technique proposed for verbal classification, based partly on alternation behaviour

• Little sense of what works well for what class, or, e.g., whether selectional preferences aid the classifier

• Potential for improvement through subcat frame acquisition (remove independence of syntactic slots), explicit modelling of selectional preferences and a better parser