Towards an Automatic Classification of Transitive Verbs in English

Effi Georgala
IMS

Motivation

- Linguistic issues of argument structure and transitivity
- NLP and Machine Learning techniques for investigation of linguistic issues
- Automatic acquisition of large-scale lexical resources for NLP applications

Outline

- Transitivity: Background
- Hypothesis & Objective
- Classification Approach
- Experiment
- Future Work
- References

Transitivity – Definition

Transitivity is traditionally understood as a global property of an entire clause, such that an activity is “carried-over” or “transferred” from an agent to a patient

[Hopper & Thompson, 1980]

Transitivity as a primitive notion (1)

Transitivity *continuum* as a composite of scores on 10 morphosyntactic and semantic parameters, e.g. aspect, volitionality, agentivity, affectedness of object etc.

[Hopper & Thompson, 1980]

Transitivity as a primitive notion (2)

- A verb's syntactic behavior can be predicted by its semantic meaning

  [Fillmore, 1970]

- Argument realization reflects event structure complexity

  [Rappaport-Hovav & Levin 1995, Levin 1999]

Transitivity as a primitive notion (3)

Two classes of transitive verbs:

- *Core Transitive Verbs* (CTV)

  e.g. *kill, destroy, open, cut , disappoint*

- *NonCore Transitive Verbs* (NCTV)

  e.g. *eat, hit, rub, sweep, detest*

[Levin, 1999]

Transitivity: 'Derived' Notion

- Transitivity is not a primitive notion that characterizes some verbal entry or not
- Argument realization is associated with structure [Hale & Keyser, 1993]
- Transitivity is a 'derived' notion for a set of structural relationships
  [Alexiadou 2003, Marantz 2003]

Outline

Transitivity: Background

Hypothesis & Objective

Classification Approach

Experiment

Future Work

References

Hypothesis

Syntactically relevant information about verbs is encoded in the lexicon in semantic event templates

- Transitive verbs can be partitioned into CTVs and NCTVs
- The contrast between CTVs and NCTVs is reflected in their argument expression

CTVs

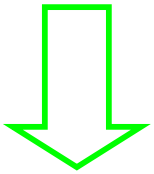
Semantic profile

Agent acts on and *necessarily* causes a change in the patient

CTVs

Complex event structure

[[x ACT <*MANNER*>] CAUSE [BECOME [y <*STATE*>]]]



Restricted argument expression

NCTVs

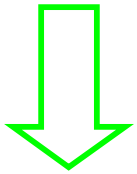
Semantic Profile

Agent acts and causes an effect on patient

NCTVs

Simple event structure

[x ACT <*MANNER*> y]



Varied argument expression

Argument Expression (Diagnostics)

Unlike CTVs, NCTVs allow:

- Unexpressed objects

  Mary scrubbed (the floor) this morning

- Resultative phrases

  She slammed herself inside the bedroom (reflexive resultative)

- Path phrases

  Mary pushed the car to the store

- *out*-prefixation

  I'm no slouch in the food department, but she consistently outordered and outate me

Objective

Classify transitive verbs on the basis of

- their morphosyntactic reflexes

- probability distribution of the reflexes

Outline

Transitivity: Background
Hypothesis & Objective
Classification Approach
Experiment
Future Work
References

Approach

- Use large data samples to automatically extract
  - Levin's morphosyntactic *diagnostics*
  - their *frequency* of occurrence

- *Classify* verbs automatically using standard classification algorithms

Extraction of Diagnostics (1)

Two types of diagnostics:

- Syntactic:
  - Resultative phrases
  - Path phrases
  - Frames
    - Unexpressed objects
    - …
- Morphological:
  - *out*-prefixation

Extraction of Diagnostics (2)

Data

- Annotated Text Corpora
  - BNC (~100 million words)
  - Tipster (~405 million words)
  - WSJ (~43 million words)
- World Wide Web (WWW)

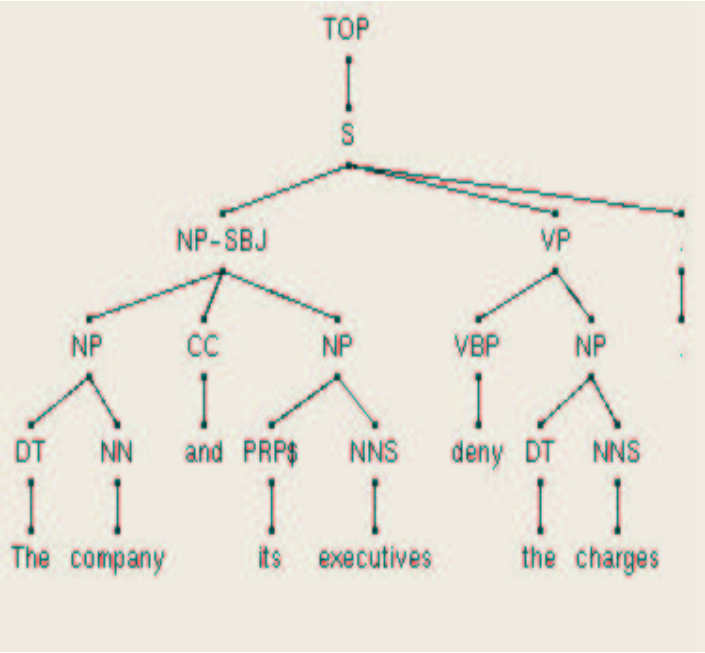
Extraction of Diagnostics (3)

Approach

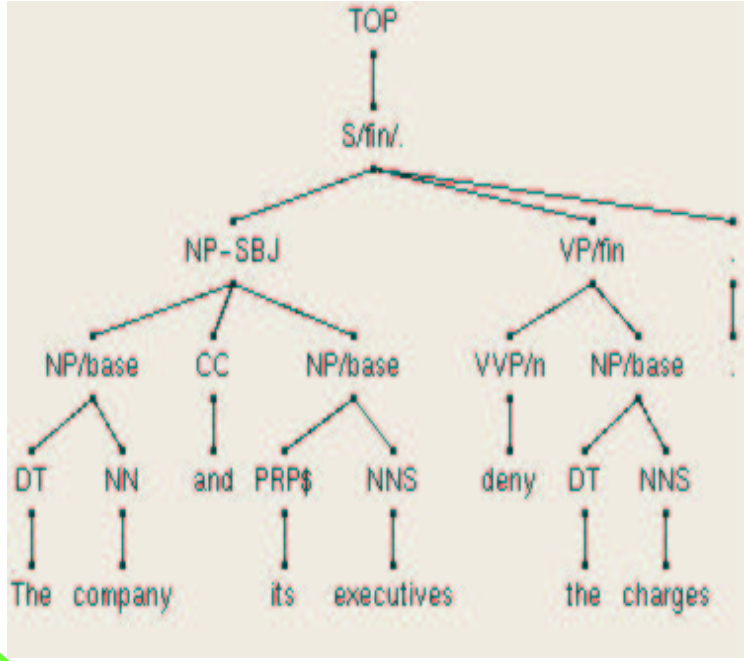
- Extraction of frames
- Acquisition of selectional preferences for identification of causative alternation
- Semi-automatic extraction of resultative constructions and *out*-prefixation

Extraction of Frames (1)

Penn-III Treebank



Penn-III Treebank & Features



Extraction of Frames (2)

Extraction of Probabilistic Context-Free Grammar from Treebank (76,144 rules)

Example:

```
5529 S/fin/. -> NP-SBJ VP/fin .
19   VP/fin  -> VVP/n NP/base
908  NP/base -> DT NNS
2733 NP/base -> PRP$ NNS
```

Extraction of Frames (3)

- Parsing of data with BitPar
- Extraction of frames from viterbi parses

Extraction of Frames (4)

Some Frame Elements:

SUBJ	:	Subject
NP	:	NP object
NP-PRD	:	NP predicative
PP-CLR	:	PP “argument” of VP
PP-LOC	:	PP locative
PRT	:	Particle
S	:	Infinitive
SBAR	:	Subordinate clauses

...

Extraction of Frames (5)

Most frequent frames of “eat”

```
337 SUBJ NP
307 SUBJ
 37  SUBJ NP PP
 25  SUBJ PP-LOC
 20  SUBJ NP PP-LOC
 18  SUBJ PRT NP
…
```

Identification of Causative Alternation

Approach

[McCarthy, 2001]

- Subcategorization frames
- Acquisition of selectional preferences at target slots
- Data similarity at target slots

Semi-Automatic Extraction

Semi-automatic extraction of

- *way*-construction (resultative)
- *out*-prefixation

from:

- corpora using the IMS Corpus Query Processor [Christ et al., 1999]
- WWW using heuristics

Classification

Objective

- Assign each verb to the subset of CTVs or NCTVs
- Identify those attributes which are relevant for the classification

Outline

Transitivity: Background
Hypothesis & Objective
Classification Approach
Experiment
Future Work
References

Experiment (1)

Data

- 58 verbs from literature on causative verbs
- Extraction of frames & their frequency: 1000 sentences of length <= 25 for each verb

Experiment (2)

Classification

- Decision tree learning algorithm: C4.5 [Quinlan, 1992]
- Attributes: 19
- Training data: 48 verbs
- Test data:10 verbs

Experiment (3)

Attributes & their Values

- *Way*-construction & *out*-prefixation (value: boolean):
  - < 50 occurences : no
  - >= 50 occurences: yes
- 17 most frequent frames (value: probability)

Experiment (4)

List of Attributes

- **way**
- **out**
- subj-np
- **subj**
- subj-np-pp
- subj-np-ppclr
- subj-np-pploc
- subj-prt
- subj-prt-np
- subj-pp
- subj-ppclr
- subj-ppdir
- subj-s
- subj-np-prt
- subj-sbar
- subj-np-ppdir
- subj-pploc
- subj-np-np
- subj-np-sbar

Experiment (5)

Results

- Evaluation on test data:
  38% misclassified verbs
- *Out*-prefixation best classifier

Experiment (6)

Discussion

- *out*-prefixation as best classifier
- What about the rest of diagnostics?
  - Too much noise in data?
  - C4.5 not suitable for this task?

Outline

Transitivity: Background
Hypothesis & Objective
Classification Approach
Experiment
Future Work
References

Future Work

- Large scale experiment for English with 1500 verbs (in preparation)
- Other classification algorithms, e.g. Support Vector Machines, Neural Networks etc.
- Experiment for Greek
- Crosslinguistic comparison

Outline

Transitivity: Background
Hypothesis & Objective
Classification Approach
Experiment
Future Work
References

References (1)

Alexiadou, Artemis. 2003. Encoding Transitivity. Paper presented at the Workshop Verb Classes and Alternations, Stuttgart. January 10–11, 2003.

Christ, Oliver, Bruno Schulze and Esther Koenig. 1999. Corpus Query Processor (CQP). User's Manual. Institut fuer Maschinelle Sprachverarbeitung, Universitaet Stuttgart, Germany.

Fillmore, Charles J. 1970. The Grammar of *Hitting* and *Breaking*. In *Readings in English Transformational Grammar*, ed. R. Jacobs and P. Rosenbaum, 120–133. Waltham, MA: Ginn.

Hale, Kenneth and Samuel J. Keyser. 1993. On Argument Structure and the Lexical Expression of Syntactic Relations. In *The View from Building 20: Essays in Honor of Sylvain Bromberger,* ed. Kenneth Hale and Samuel J. Keyser. 53–109. Cambridge, MA: MIT Press.

Hopper, Paul and Sandra A. Thompson. 1980. Transitivity in Grammar and Discourse. *Language* 56, 251–299.

References (2)

Levin, Beth. 1999. Objecthood: An Event Structure Perspective. In *Proceedings of CLS 35, Volume 1: The Main Session*, Chicago Linguistic Society, University of Chicago, Chicago, IL, 223-247.

Marantz, Alec. 2003. Roots and Argument Structure. Paper presented at the Workshop Verb Classes and Alternations, Stuttgart. January 10-11, 2003.

McCarthy, Diana. 2001. Lexical Acquisition at the Syntax-Semantics Interface: Diathesis Alternations, Subcategorization Frames and Selectional Preferences. Ph.D. Thesis. University of Sussex, Great Britain.

Quinlan, J. Ross. 1992. C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann Publishers.

Rappaport-Hovav, Malka and Beth Levin. 1995. Unaccusativity: At the Syntax-Lexical Semantics Interface. Linguistic Inquiry Monograph 26. Cambridge, MA: MIT Press.