Resolving Shell Nouns

Ph.D. Dissertation Defence
Advisor: Graeme Hirst

by
Varada Kolhatkar
Department of Computer Science
University of Toronto
varada@cs.toronto.edu
Shell nouns


Shell noun resolution: Identifying shell content of a shell noun phrase in the given context

The municipal council will have to decide whether to balance the budget by raising revenue or cutting spending. The council will have to come to a resolution by the end of the month. This issue is dividing communities across the country.
Goal of the research

Treatment of shell nouns from a computational linguistics perspective

- Research questions
  - To what extent are speakers of English able to interpret shell nouns?
  - How can we develop a computational system to resolve shell nouns?
  - To what extent can the knowledge derived from the linguistics literature help in this process?
Examples of shell nouns

Schmid provides a list of 670 shell nouns

<table>
<thead>
<tr>
<th>fact</th>
<th>issue</th>
<th>problem</th>
<th>principle</th>
<th>decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>thing</td>
<td>concept</td>
<td>reason</td>
<td>notion</td>
<td>phenomenon</td>
</tr>
<tr>
<td>idea</td>
<td>rumour</td>
<td>legend</td>
<td>message</td>
<td>possibility</td>
</tr>
<tr>
<td>belief</td>
<td>plan</td>
<td>truth</td>
<td>theory</td>
<td>thought</td>
</tr>
<tr>
<td>order</td>
<td>trend</td>
<td>argument</td>
<td>proposal</td>
<td>certainty</td>
</tr>
</tbody>
</table>
Why do we care?

- Ubiquity of shell nouns
  - *fact, idea, problem*: among 100 most frequently occurring nouns in the BNC (*Schmid 2000*)

- Functions in discourse
  - *Characterize* and *label* information in the context

- Potential applications
  - Discourse understanding, text summarization, non-factoid question answering, ESL learning (*Francis 1988; Flowerdew 2003; Hinkel 2004*)
State of the art in CL

- Fair amount of attention in Linguistics

... but largely ignored in CL

Current challenge in anaphora resolution:
Going beyond nominal anaphora
(Byron 2004; Poesio 2011)
Pilot study

Chapter 3
(Kolhatkar and Hirst 2012)
Focus

Resolve *this issue* instances in the Medline domain

- Data
  - 183 *this issue* instances
  - Agreement: **0.86** *(Krippendorff 2013)*
Resolution algorithm

- Medline abstracts containing *this issue*
- Predicted shell content for unseen instances
- Candidate extraction
  - syntactic constituents given by the Stanford parser
- Feature extraction
  - syntactic, semantic, lexical features
- Candidate ranking
  - SVM ranking models
  - Candidate ranking
Pilot study: summary

Feasible to reliably annotate and resolve the shell noun phrase *this issue* in the Medline domain.
Generalizing to other shell nouns

- Goal
  - A variety of shell nouns
  - Broader domain
- Primary challenges
  - Idiosyncrasies
  - No annotated data
  - A variety of constructions
The municipal council will have to decide whether to balance the budget by raising revenue or cutting spending. The council will have to come to a resolution by the end of the month. This issue is ...

The issue that this country and Congress must address is how to provide optimal care for all without limiting access for the many.

A bad idea does not harm until someone acts upon it.

Mathis is the cover subject of this week’s issue of Sports Illustrated.
Different types of usages

The municipal council will have to decide whether to balance the budget by raising revenue or cutting spending. The council will have to come to a resolution by the end of the month. This issue is ...

The issue that this country and Congress must address is how to provide optimal care for all without limiting access for the many.

A bad idea does not harm until someone acts upon it.

Mathis is the cover subject of this week’s issue of Sports Illustrated.
The municipal council will have to decide whether to balance the budget by raising revenue or cutting spending. The council will have to come to a resolution by the end of the month. This issue is ... The issue that this country and Congress must address is how to provide optimal care for all without limiting access for the many. A bad idea does not harm until someone acts upon it. Mathis is the cover subject of this week’s issue of Sports Illustrated.
The municipal council will have to decide whether to balance the budget by raising revenue or cutting spending. The council will have to come to a resolution by the end of the month. This issue is...

The issue that this country and Congress must address is how to provide optimal care for all without limiting access for the many.

A bad idea does not harm until someone acts upon it.

Mathis is the cover subject of this week’s issue of Sports Illustrated.

Different types of usages

anaphoric (ASN)

indefinite shell content

cataphoric (CSN)
Different types of usages

The municipal council will have to decide whether to balance the budget by raising revenue or cutting spending. The council will have to come to a resolution by the end of the month. This issue is...

The issue that this country and Congress must address is how to provide optimal care for all without limiting access for the many.

A bad idea does not harm until someone acts upon it.

Mathis is the cover subject of this week's issue of Sports Illustrated.
Resolving Cataphoric Shell Nouns (CSNs)

Chapter 4
(Kolhatkar and Hirst 2014)
## CSN patterns

*(Schmid, 2000)*

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-be-to</td>
<td>Our plan is to hire and retain the best managers we can.</td>
</tr>
<tr>
<td>N-be-that</td>
<td>The major reason is that doctors are uncomfortable with uncertainty.</td>
</tr>
<tr>
<td>N-be-wh</td>
<td>Of course, the central issue is whether animal testing is cruel.</td>
</tr>
<tr>
<td>N-to</td>
<td>The decision to disconnect the ventilator came after doctors found no brain activity.</td>
</tr>
<tr>
<td>N-that</td>
<td>These challenges do not undermine the fact that museums are on a high.</td>
</tr>
<tr>
<td>N-wh</td>
<td>If there ever is any doubt whether a plant is a poppy or not, break off a stem and squeeze it.</td>
</tr>
<tr>
<td>N-of</td>
<td>The concept of having an outsider as Prime Minister is outdated.</td>
</tr>
</tbody>
</table>
CSNs: a semantic phenomenon

One reason that 60 percent of New York City public-school children read below grade level is that many elementary schools don’t have libraries.

- Identify that *reason* expects two arguments: cause and effect
- Identify that the shell content is given in the cause argument
- Identify the syntactic constituent representing cause
Shell noun families

(Schmid, 2000)

- Where can we find this kind of semantic knowledge?
  
  Answer: Schmid’s semantic families
  
- Schmid groups together different usages of 670 shell nouns into 79 semantic families
# Shell noun families

*(Schmid, 2000)*

**Idea**

<table>
<thead>
<tr>
<th>Shared semantic features</th>
<th>[MENTAL], [CONCEPTUAL]</th>
</tr>
</thead>
<tbody>
<tr>
<td>frame</td>
<td>Mental; focus on propositional content of IDEA</td>
</tr>
<tr>
<td>Nouns</td>
<td><em>Point, idea, position, issue, theory, notion, thought, principle, rule, subject, image, myth, law, theme, concept, secret, scenario, wisdom, hypothesis, thesis,...</em></td>
</tr>
<tr>
<td>Pattern</td>
<td><em>N-be-that, N-that</em></td>
</tr>
</tbody>
</table>
There is now some question whether the country was ever really in a recession.
The question that Japanese movie people most frequently ask the American visitor is why ...
Evaluation data

**Idea family**

- **Semantic features**: [mental], [conceptual]
- **Frame**: mental
- **Nouns**: idea, issue, concept
- **Patterns**: N-be-that/of, N-that/of

** Trouble family**

- **Semantic features**: [eventive], [attitudinal], [manner], [deontic]
- **Frame**: general eventive
- **Nouns**: problem, trouble, difficulty
- **Patterns**: N-be-to

**Plan family**

- **Semantic features**: [mental], [volitional], [manner]
- **Frame**: decision, plan, policy
- **Nouns**: decision, plan, policy
- **Patterns**: N-be-to/that, N-to/that

**Problem family**

- **Semantic features**: [factual], [attitudinal], [impeding]
- **Frame**: general factual
- **Nouns**: trouble, difficulty, point, thing, snag
- **Patterns**: N-be-that/of

**Reason family**

- **Semantic features**: [factual], [causal]
- **Frame**: causal, attentional focus on CAUSE
- **Nouns**: reason, thing
- **Patterns**: N-be-that/why, N-be-at/why

<table>
<thead>
<tr>
<th>Frame</th>
<th>Semantic features</th>
<th>Nouns</th>
<th>Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>mental</td>
<td>[mental], [conceptual]</td>
<td>idea, issue, concept</td>
<td>N-be-that/of, N-that/of</td>
</tr>
<tr>
<td>eventive</td>
<td>[eventive], [attitudinal], [manner], [deontic]</td>
<td>problem, trouble, difficulty</td>
<td>N-be-to</td>
</tr>
<tr>
<td>decision</td>
<td>[mental], [volitional], [manner]</td>
<td>decision, plan, policy</td>
<td>N-be-to/that, N-to/that</td>
</tr>
<tr>
<td>general factual</td>
<td>[factual], [attitudinal], [impeding]</td>
<td>trouble, difficulty, point, thing, snag</td>
<td>N-be-that/of</td>
</tr>
<tr>
<td>general factual</td>
<td>[factual], [causal]</td>
<td>reason, thing</td>
<td>N-be-that/why, N-be-at/why</td>
</tr>
</tbody>
</table>

**Table 3: Example families from Schmid (2000). The nouns in boldface are used to evaluate this work.**
CSN resolution results

Semantic knowledge from Schmid’s families is helpful in resolving CSNs
CSN resolution results

Schmid’s framework particularly helps in resolving nouns with strict expectations

![Bar chart showing accuracy comparison between baseline and + Schmid’s cues for 'fact' and 'reason'.]
CSN resolution results

Schmid’s cues were deleterious for more flexible nouns

![Bar chart showing accuracy for 'policy', 'problem', 'trouble', and 'difficulty' with and without Schmid’s cues.](chart.png)
Resolving Anaphoric Shell Nouns

Chapter 5

(Kolhatkar et al. 2013a, 2013b)
The municipal council had to decide whether to balance the budget by raising revenue or cutting spending. The council had to come to a resolution by the end of the month. This issue was dividing communities across the country.

**whether clause in both cases**

Of course, the central, and probably insoluble, issue is whether animal testing is cruel.
Hypothesis

CSN shell content and ASN shell content share some linguistic properties, and hence linguistic knowledge encoded in CSN shell content will help in interpreting ASNs.
Overview

Training

CSN examples

CSN shell content extractor

CSN shell content models

Automatically labeled training data

Testing

ASN examples

Predicted ASN shell content

crowd evaluation
Annotating ASNs to their shell content
(Kolhatkar et al. 2013a)
The ASN corpus

Base corpus: The NYT corpus (Sandhaus 2008)

~475 instances per 6 selected shell nouns
  fact, reason, issue, decision, question, possibility

Total: 2,323 ASN instances
Annotation tasks

ASN instances from the NYT

CrowdFlower Expt. 1

Identify the sentence containing shell content

CrowdFlower Expt. 2

Identify the precise shell content

Annotated ASN Corpus

Crowdsourcing does best with simple tasks
(Madnani et al. 2010; Wang et al. 2012)
New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings.

Some lawmakers worry that cameras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process.

New York's backwardness on this issue hurts public confidence in the judiciary...
New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. Some lawmakers worry that cameras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. New York's backwardness on this issue hurts public confidence in the judiciary...
New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings.

Select one of the options (required)
- None
- one of only three states
- some form of audio-visual coverage
- some form of audio-visual coverage of court proceedings
- that do not allow some form of audio-visual coverage of court proceedings
- audio-visual coverage of court proceedings
- some form
- New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings.
- only three states that do not allow some form of audio-visual coverage of court proceedings
- one of only three states that do not allow some form of audio-visual coverage of court proceedings
- allow some form of audio-visual coverage of court proceedings

10 top-ranked candidates given by Kolhatkar et al. 2013b
How far can we get using CSN models?
Metric and baseline

- Success at \(n\) (\(S@n\))
- Proportion of instances where the crowd’s answers occur within our ranker’s first \(n\) choices
- \(S@1\) is standard precision
- Baseline
  - Consider the crowd-annotated sentence as the correct shell content
Ranker evaluation

- **fact**
  - Baseline: 0.47
  - S@1: 0.70
  - S@2: 0.70
  - S@3: 0.82
  - S@4: 0.94

- **reason**
  - Baseline: 0.52
  - S@1: 0.72
  - S@2: 0.72
  - S@3: 0.84
  - S@4: 0.93

- **question**
  - Baseline: 0.38
  - S@1: 0.70
  - S@2: 0.70
  - S@3: 0.84
  - S@4: 0.90

- **issue**
  - Baseline: 0.26
  - S@1: 0.47
  - S@2: 0.47
  - S@3: 0.75
  - S@4: 0.78

- **decision**
  - Baseline: 0.29
  - S@1: 0.35
  - S@2: 0.35
  - S@3: 0.68
  - S@4: 0.76

- **possibility**
  - Baseline: 0.44
  - S@1: 0.56
  - S@2: 0.56
  - S@3: 0.63
  - S@4: 0.81
Summary of contributions

• First work that sheds light on shell nouns from a computational linguistic perspective

• First step towards resolving abstract anaphora

• Three resolution systems and four reliably-annotated corpora to further pursue this line of research
Future directions

• Short-term future directions
  • Shell noun resolution for other languages (with Heike Zinsmeister and Stefanie Dipper)
  • One SVM ranker for all shell nouns (with Alexander Schwing)

• Long-term future directions
  • Clustering shell nouns with similar semantic expectations similar to verb clustering
  • Identifying shell chains
End-to-end shell noun resolution

- Resolving six CSNs with rules (Kolhatkar et al. 2013a)
- Generalizing CSN resolution (Kolhatkar and Hirst 2014)
- Resolving the same six ASNs using CSN shell content as training data (Kolhatkar et al. 2013a)
- Annotating ASNs (Kolhatkar et al. 2013b)
Schmid’s definition of shell nouns

- Characterization
- Concept-formation
  - Allow speakers to encapsulate the complex chunks of information in temporary nominal concepts with clear-cut conceptual boundaries
- Linking
  - Interpret two groups of linguistic elements together, as being related to and even dependent on each other
Schmid’s definition of shell nouns

Shell nouns

- Full content nouns
  - e.g., teacher, cat, journey

  - stable and rich denotation
  - relatively constant relationship to the experience they encapsulate as a concept
  - suited for exophoric reference

- Pronouns with anaphoric function
  - e.g., she, it, this, that

  - limited potential for characterization
  - no concept-forming effects
  - create links of referential identity or co-reference
Schmid’s definition of shell nouns

- Combine the three functions of characterization, concept-formation and linking, which are otherwise performed separately, each by different types of linguistic elements.
- They perform these functions in a fine-tuned balance between conceptual stability and informational flexibility.
Schmid’s definition of shell nouns
Ideas regarding proper evaluation of shell nouns

• Evaluation metric based on the success of the system at identifying the type of the objects that are really referred to instead of the words that evoke those objects

• This problem is challenging for anaphoric demonstratives, as the type of the referent has to be identified from the predicative context

• For shell nouns, the type of the referent is encoded in the shell nouns themselves
Identifying just the type of the referent is not enough

The teacher erased the solutions before John had time to copy them out, as he had momentarily been distracted by a band playing outside.

A. This fact infuriated him, as the teacher always erased the board quickly and John suspected it was just to punish anyone who was lost in thought, even for a moment.

B. This fact infuriated the teacher, who had already told John several times to focus on class work.
The teacher erased the solutions before John had time to copy them out, as he had momentarily been distracted by a band playing outside.

A. This fact infuriated him, as the teacher always erased the board quickly and John suspected it was just to punish anyone who was lost in thought, even for a moment.

B. This fact infuriated the teacher, who had already told John several times to focus on class work.
Identifying just the type of the referent is not enough

Shell content for A

The teacher erased the solutions before John had time to copy them out, as he had momentarily been distracted by a band playing outside.

Shell content for B

A. This fact infuriated him, as the teacher always erased the board quickly and John suspected it was just to punish anyone who was lost in thought, even for a moment.

B. This fact infuriated the teacher, who had already told John several times to focus on class work.
Ideas regarding proper evaluation of shell nouns

• Identifying the words, i.e., syntactic constituent, representing the required object type itself is challenging, as there is no one-to-one correspondence between a semantic concept and its syntactic shape

• A concept like *issue* can take many different syntactic forms such as verb phrases, noun phrases, sentences, clauses
Ideas regarding proper evaluation of shell nouns

- Consider first $n$ crowd answers rather than the top answer
- Extrinsic evaluation
- E.g., to what extent shell noun resolution helps ESL learners
Automatically determining whether a noun is flexible

<table>
<thead>
<tr>
<th>Noun</th>
<th>N-be-to</th>
<th>N-be-that</th>
<th>N-be-wh</th>
<th>N-to</th>
<th>N-that</th>
<th>N-wh</th>
<th>N-of</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>problem</td>
<td>2</td>
<td>29</td>
<td>2</td>
<td>6</td>
<td>19</td>
<td>3</td>
<td>38</td>
<td>31,189</td>
</tr>
<tr>
<td>reason</td>
<td>-</td>
<td>9</td>
<td>-</td>
<td>59</td>
<td>15</td>
<td>15</td>
<td>2</td>
<td>41,230</td>
</tr>
<tr>
<td>fact</td>
<td>-</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>90</td>
<td>-</td>
<td>5</td>
<td>81,825</td>
</tr>
</tbody>
</table>

Flexible nouns tend to occur with a variety of patterns
The history of language is the history of a process of abbreviation.

— Friedrich Nietzsche
## CrowdFlower Confidence

<table>
<thead>
<tr>
<th>Annotator</th>
<th>Trust</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.75</td>
<td>“a”</td>
</tr>
<tr>
<td>B</td>
<td>0.75</td>
<td>“a”</td>
</tr>
<tr>
<td>C</td>
<td>1.0</td>
<td>“a”</td>
</tr>
</tbody>
</table>
CrowdFlower Confidence

<table>
<thead>
<tr>
<th>Annotator</th>
<th>Trust</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.75</td>
<td>“a”</td>
</tr>
<tr>
<td>B</td>
<td>0.75</td>
<td>“a”</td>
</tr>
<tr>
<td>C</td>
<td>1.0</td>
<td>“a”</td>
</tr>
</tbody>
</table>

score for “$a_2$” = 1.75
CrowdFlower Confidence

<table>
<thead>
<tr>
<th>Annotator</th>
<th>Trust</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.75</td>
<td>“a”</td>
</tr>
<tr>
<td>B</td>
<td>0.75</td>
<td>“a”</td>
</tr>
<tr>
<td>C</td>
<td>1.0</td>
<td>“a”</td>
</tr>
</tbody>
</table>

Crowd’s answer:
“a₂” with confidence 0.7 = 1.75/(1.75 + 0.75)
Evaluators were generally satisfied with the provided options.
Head agreement

About 94% of the time, at least 4 annotators agreed on the head of the antecedent.
Exact agreement

About 89% of the time, at least 4 annotators agreed on the exact antecedent
Syntactic Type Distribution

- Sentences
- Clauses
- Noun Phrases
- Verb Phrases
- Adjective Phrases
- Prepositional Phrases

Bar chart showing the distribution of syntactic types for different categories: fact, reason, issue, decision, question, and possibility.
Hard Examples

The teacher erased the solutions before John had time to copy them out, as he had momentarily been distracted by a band playing outside.

- This fact infuriated him, as the teacher always erased the board quickly and John suspected it was just to punish anyone who was lost in thought, even for a moment.

- This fact infuriated the teacher, who had already told John several times to focus on class work.
Hard Examples

Several Vatican officials said, however, that any such talk has little meaning because the church does not take sides in elections. But the statements by several American bishops that Catholics who vote for Mr. Kerry would have to go to confession have raised the question in many corners about whether this is an official church position.

The church has not addressed this question publicly and, in fact, seems reluctant to be dragged into the fight...”
Any biography of Thomas More has to answer one fundamental question. Why? Why, out of all the many ambitious politicians of early Tudor England, did only one refuse to acquiesce to a simple piece of religious and politica opportunism? What was it about More that set him apart and doomed him to a spectacularly avoidable execution?

The innovation of Peter Ackroyd’s new biography of More is that he places the answer to this question outside of More himself.