Towards deep content extraction from specialized discourse: The case of verbal relations in patent claims

Gabriela Ferraro

Department of Communication and Information Technologies
Universitat Pompeu Fabra

November 2012
Outline

1. Introduction
2. Theoretical Framework
3. Related Work
4. Approach
5. Evaluation
6. Conclusions
Introduction

Theoretical Framework

Related Work

Approach

Evaluation

Conclusions

Problem definition

An automatic focusing device comprising: an objective lens for focusing a light beam emitted by a light source on a track of an information recording medium; a beam splitter for separating a reflected light beam reflected by the information recording medium at a focal spot thereon and through the objective lens from the light beam emitted by the light source; an astigmatic optical system including an optical element capable of causing the astigmatic aberration of the separated reflected light beam; a light detector having a light receiving surface divided, except the central portion thereof, into a plurality of light receiving sections which are arranged symmetrically with respect to a first axis extending in parallel to the axial direction of the optical element and to a second axis extending perpendicularly to the first axis and adapted to receive the reflected beam transmitted through the optical element and to give a light reception output signal corresponding to the shape of the spot of the reflected light beam formed on the light receiving surface; a focal position detecting circuit capable of giving an output signal corresponding to the displacement of the objective lens from the focused position, on the basis of the output signal given by the light detector; and a lens driving circuit which drives the objective lens along the optical axis on the basis of the output signal given by the focal position [...]
Problem definition

An automatic focusing device **comprising**: an objective lens for focusing a light beam emitted by a light source on a track of an information recording medium; a beam splitter for separating a reflected light beam reflected by the information recording medium at a focal spot thereon and through the objective lens from the light beam emitted by the light source; an astigmatic optical system including an optical element capable of causing the astigmatic aberration of the separated reflected light beam; a light detector having a light receiving surface divided, except the central portion thereof, into a plurality of light receiving sections which are arranged symmetrically with respect to a first axis extending in parallel to the axial direction of the optical element and to a second axis extending perpendicularly to the first axis and adapted to receive the reflected beam transmitted through the optical element and to give a light reception output signal corresponding to the shape of the spot of the reflected light beam formed on the light receiving surface; a focal position detecting circuit capable of giving an output signal corresponding to the displacement of the objective lens from the focused position, on the basis of the output signal given by the light detector; and a lens driving circuit which drives the objective lens along the optical axis on the basis of the output signal given by the focal position [...]

Proposal
Why patent claims?

- Patents contain up-to-date scientific and technical information that is usually not published elsewhere.

- Patent claims form the central section of a patent document as they define the boundaries of legal protection.

- Patent claim genre is a challenge to NLP.
Why patent claims?

- Patents contain up-to-date scientific and technical information that is usually not published elsewhere.

- Patent claims form the central section of a patent document as they define the boundaries of legal protection.

- Patent claim genre is a challenge to NLP.
Why patent claims?

- Patents contain up-to-date scientific and technical information that is usually not published elsewhere

- Patent claims form the central section of a patent document as they define the boundaries of legal protection

- Patent claim genre is a challenge to NLP
Why patent claims?

- Patents contain up-to-date scientific and technical information that is usually not published elsewhere.

- Patent claims form the central section of a patent document as they define the boundaries of legal protection.

- Patent claim genre is a challenge to NLP.
Objectives

- Provide a detailed description of the patent claim genre via analysis of the linguistic idiosyncrasies of the claim style
- Provide a framework for flexible, unlexicalized and n-ary verbal relation extraction via deep dependency parsing
- Provide a framework for the generalization of verbal relations via clustering and cluster labeling techniques
Objectives

- Provide a detailed description of the patent claim genre via analysis of the linguistic idiosyncrasies of the claim style
- Provide a framework for flexible, unlexicalized and n-ary verbal relation extraction via deep dependency parsing
- Provide a framework for the generalization of verbal relations via clustering and cluster labeling techniques
Objectives

- Provide a detailed description of the patent claim genre via analysis of the linguistic idiosyncrasies of the claim style
- Provide a framework for flexible, unlexicalized and n-ary verbal relation extraction via deep dependency parsing
- Provide a framework for the generalization of verbal relations via clustering and cluster labeling techniques
Objectives

- Provide a detailed description of the patent claim genre via analysis of the linguistic idiosyncrasies of the claim style
- Provide a framework for flexible, unlexicalized and n-ary verbal relation extraction via deep dependency parsing
- Provide a framework for the generalization of verbal relations via clustering and cluster labeling techniques
Objectives

- Provide a detailed description of the patent claim genre via analysis of the linguistic idiosyncrasies of the claim style
- Provide a framework for flexible, unlexicalized and n-ary verbal relation extraction via deep dependency parsing
- Provide a framework for the generalization of verbal relations via clustering and cluster labeling techniques
Objectives

- Provide a detailed description of the patent claim genre via analysis of the linguistic idiosyncrasies of the claim style
- Provide a framework for flexible, unlexicalized and n-ary verbal relation extraction via deep dependency parsing
- Provide a framework for the generalization of verbal relations via clustering and cluster labeling techniques
Objectives

- Provide a detailed description of the patent claim genre via analysis of the linguistic idiosyncrasies of the claim style
- Provide a framework for flexible, unlexicalized and n-ary verbal relation extraction via deep dependency parsing
- Provide a framework for the generalization of verbal relations via clustering and cluster labeling techniques
Linguistic dependency

Dependency relations are:

- binary directed asymmetrical relations that take two arguments
- one argument is the head
- the other argument is the dependent of the head
The Meaning-Text Theory
The device is composed by three optical units.
Surface syntax vs. Deep syntax

Surface syntax:
- dependencies capture the grammatical relations of lexical units
- comprises around 50 language-dependent relations
- all the lexemes of the sentence are nodes of the tree

Deep syntax:
- dependencies capture the predicate argument structure
- comprises only 9 language-independent relations
- only semantically full lexemes appear in the tree
Relation extraction

- lexico-syntactic patterns
- purely syntactic patterns
- word co-occurrences
Relation clustering deals with grouping relations according to some similarity criteria

- Over a taxonomy or set of predefined relations (*verbs of putting*, *verbs of removing*, *verbs of motion*, etc.)

- Using similarity measures (MI, LRA, etc.)

- Using the WordNet hierarchy
Cluster labeling

- Internal cluster labeling
- Differential cluster labeling
A step further in RE

- **Relation Extraction**
  - open & n-ary
  - use deep-syntactic trees to identify the arguments of relations

- **Relation Classification**
  - unsupervised methods with simple features
Architecture

**raw patent claim**

**Preprocessing: Claim Simplification**
- XML Structuring
- Pos Tagging and Chunking
- Claim Segmentation

**Sentence Reconstruction**

**Clause Structuring**

**Coreference Resolution**

**simplified sentences**

**Syntactic Parsing**
- Minipar Parsing
- SSynt Mapping
- DSynt Mapping

**deep syntactic trees**

**Relation Generalization**
- Tuple Extraction
- Relation Clustering
- Rel. Cluster Labeling

**relation tuples**
Preprocessing: claim simplification

- Linguistic preprocessing: TreeTagger
- Clause segmentation: Machine learning (ML) approach
- Coreference resolution between NPs
- Clause tree building: ML approach
- Clause reconstruction: rule-based approach
An automatic focusing device comprising: an objective lens for focusing a light beam emitted by a light source on a track of an information recording medium; a beam splitter for separating a reflected light beam reflected by the information recording medium at a focal spot thereon and through the objective lens from the light beam emitted by the light source; an astigmatic optical system including an optical element capable of causing the astigmatic aberration of the separated reflected light beam; a light detector having a light receiving surface divided, except the central portion thereof, into a plurality of light receiving sections which are arranged symmetrically with respect to a first axis extending in parallel to the axial direction of the optical element and to a second axis extending perpendicularly to the first axis and adapted to receive the reflected beam transmitted through the optical element and to give a light reception output signal corresponding to the shape of the spot of the reflected light beam formed on the light receiving surface; a focal position detecting circuit capable of giving an output signal corresponding to the displacement of the objective lens from the focused position, on the basis of the output signal given by the light detector; and a lens driving circuit which drives the objective lens along the optical axis on the basis of the output signal given by the focal position [...]

Simplification input example
An automatic focusing device comprises: an objective lens; a beam splitter; an astigmatic optical system; a light detector; a focal position detecting circuit capable of giving an output signal and a lens driving circuit.

The objective lens focusses light beam.

The light source emits a light beam on a track of an information recording medium.

The beam splitter separates the reflected light beam.

The information recording medium reflects the reflected light beam at a focal spot thereon and through the objective lens from the light beam.

The light source emits the light beam.

The astigmatic optical system includes an optical element.
Syntactic parsing

- Minipar Parsing
- SSynt Mapping
- DSynt Mapping

simplified sentences

deep syntactic trees
Syntactic parsing: Minipar to MTT Surface Syntax

Minipar tree

SSynt tree
Syntactic parsing: Surface Syntax to Deep Syntax

SSynt tree

DSynt tree
Relation generalization

deep syntactic trees

Relation Generalization

Tuple Extraction → Relation Clustering → Rel. Cluster Labeling

relation tuples
Obtaining relation tuples

- argument tags are eliminated
- actantial relations are merged
- attributive relations are merged

**Actantial tuples merging rule:**

IF ( I(V_i, A_1) & II(V_i, A_2) ) 
SET ( V_i(A_1, A_2))
FI

**Attributive tuples merging rule:**

IF ( ATTR(A_1, A_2) & II(A_2, A_3) ) 
SET ( A_2(A_1, A_3))
FI
Relation clustering

- Each verbal relation is represented as a vector of its WordNet-synonyms, e.g.:

\[
\begin{align*}
\text{construct1(construct1, build1, make17)} \\
\text{make24(make24, build1, construct1)}
\end{align*}
\]

**Step 1:** calculate the similarity between synonym vectors

\[
\cos(\mathbf{v}_i, \mathbf{w}_j) = \frac{\mathbf{v}_i \cdot \mathbf{w}_j}{|\mathbf{v}_i| \cdot |\mathbf{w}_j|} = \frac{\sum_{s=1}^{N} v_{i,s} w_{j,s}}{\sqrt{\sum_{s=1}^{N} v_{i,s}^2} \sqrt{\sum_{s=1}^{N} w_{j,s}^2}}
\]

**Step 2:** cluster the vectors according to the obtained similarities

- Best clustering algo.: *optimized repeated bisections*; \( K = 60 \)
Relation clustering

- Each verbal relation is represented as a vector of its WordNet-synonyms, e.g.:

  \[
  \text{construct1(construct1, build1, make17)} \\
  \text{make24(make24, build1, construct1)}
  \]

**Step 1:** calculate the similarity between synonym vectors

\[
\cos (\vec{v}_i, \vec{w}_j) = \frac{\vec{v}_i \cdot \vec{w}_j}{|\vec{v}_i| \cdot |\vec{w}_j|} = \frac{\sum_{s=1}^{N} v_{i,s}w_{j,s}}{\sqrt{\sum_{s=1}^{N} v_{i,s}^2} \sqrt{\sum_{s=1}^{N} w_{j,s}^2}}
\]

**Step 2:** cluster the vectors according to the obtained similarities

- Best clustering algo.: *optimized repeated bisections*; \( K = 60 \)
Relation clustering

- Each verbal relation is represented as a vector of its WordNet-synonyms, e.g.:

  \[ \text{construct1(construct1, build1, make17)} \]
  \[ \text{make24(make24, build1, construct1)} \]

**Step 1:** calculate the similarity between synonym vectors

\[
\cos(\vec{v}_i, \vec{w}_j) = \frac{\vec{v}_i \cdot \vec{w}_j}{|\vec{v}_i| \cdot |\vec{w}_j|} = \frac{\sum_{s=1}^{N} v_{i,s} w_{j,s}}{\sqrt{\sum_{s=1}^{N} v_{i,s}^2} \sqrt{\sum_{s=1}^{N} w_{j,s}^2}}
\]

**Step 2:** cluster the vectors according to the obtained similarities

- Best clustering algo.: *optimized repeated bisections; K = 60*
Relation clustering

- Each verbal relation is represented as a vector of its WordNet-synonyms, e.g.:

construct1(construct1, build1, make17)
make24(make24, build1, construct1)

**Step 1:** calculate the similarity between synonym vectors

\[
\cos (\vec{v}_i, \vec{w}_j) = \frac{\vec{v}_i \cdot \vec{w}_j}{|\vec{v}_i| \cdot |\vec{w}_j|} = \frac{\sum_{s=1}^{N} v_{i,s} w_{j,s}}{\sqrt{\sum_{s=1}^{N} v_{i,s}^2} \sqrt{\sum_{s=1}^{N} w_{j,s}^2}}
\]

**Step 2:** cluster the vectors according to the obtained similarities

- Best clustering algo.: optimized repeated bisections; \( K = 60 \)
Relation clustering

- Each verbal relation is represented as a vector of its WordNet-synonyms, e.g.:

\[
\text{construct1(construct1, build1, make17)}
\]
\[
\text{make24(make24, build1, construct1)}
\]

**Step 1**: calculate the similarity between synonym vectors

\[
\cos (\vec{v}_i, \vec{w}_j) = \frac{\vec{v}_i \cdot \vec{w}_j}{|\vec{v}_i| \cdot |\vec{w}_j|} = \frac{\sum_{s=1}^{N} v_{i,s} w_{j,s}}{\sqrt{\sum_{s=1}^{N} v_{i,s}^2} \sqrt{\sum_{s=1}^{N} w_{j,s}^2}}
\]

**Step 2**: cluster the vectors according to the obtained similarities

- Best clustering algo.: *optimized repeated bisections*; \( K = 60 \)
Relation clustering

- Each verbal relation is represented as a vector of its WordNet-synonyms, e.g.:

  construct1(construct1, build1, make17)
  make24(make24, build1, construct1)

**Step 1:** calculate the similarity between synonym vectors

\[
\cos (\mathbf{v}_i, \mathbf{w}_j) = \frac{\mathbf{v}_i \cdot \mathbf{w}_j}{|\mathbf{v}_i| \cdot |\mathbf{w}_j|} = \frac{\sum_{s=1}^{N} v_{i,s} w_{j,s}}{\sqrt{\sum_{s=1}^{N} v_{i,s}^2} \sqrt{\sum_{s=1}^{N} w_{j,s}^2}}
\]

**Step 2:** cluster the vectors according to the obtained similarities

- Best clustering algo.: *optimized repeated bisections*; \(K = 60\)
Examples of relational clusters

- comprise, contain, stop, block
- bound, limit, restrain, inhibit
- curve, cut, reduce, trim
- differentiate, distinguish, separate
- extract, pull-out
- delete, erase
- enter, insert, introduce
- tighten, fasten, secure
- associate, connect, join, link, relate
- become, come, release, turn
Cluster labeling

Internal cluster labeling:
- Frequency oriented labeling
- Verb hyperonym-oriented labeling
- Thesaurus frequency-oriented labeling

Differential cluster labeling:
- VHyper MI-oriented labeling
- VHyper $\chi^2$-oriented labeling
- Thesaurus MI-oriented labeling
- Thesaurus $\chi^2$-oriented labeling
Cluster labeling

Internal cluster labeling:
- Frequency oriented labeling
- Verb hyperonym-oriented labeling
- Thesaurus frequency-oriented labeling

Differential cluster labeling:
- VHyper MI-oriented labeling
- VHyper $\chi^2$-oriented labeling
- Thesaurus MI-oriented labeling
- Thesaurus $\chi^2$-oriented labeling
Cluster labeling

Internal cluster labeling:
- Frequency oriented labeling
- Verb hyperonym-oriented labeling
- Thesaurus frequency-oriented labeling

Differential cluster labeling:
- VHyper MI-oriented labeling
- VHyper $\chi^2$-oriented labeling
- Thesaurus MI-oriented labeling
- Thesaurus $\chi^2$-oriented labeling
Cluster labeling

Internal cluster labeling:
- Frequency oriented labeling
- Verb hyperonym-oriented labeling
- Thesaurus frequency-oriented labeling

Differential cluster labeling:
- VHyper MI-oriented labeling
- VHyper $\chi^2$-oriented labeling
- Thesaurus MI-oriented labeling
- Thesaurus $\chi^2$-oriented labeling
Cluster labeling

Internal cluster labeling:
- Frequency oriented labeling
- Verb hyperonym-oriented labeling
- Thesaurus frequency-oriented labeling

Differential cluster labeling:
- VHyper MI-oriented labeling
- VHyper $\chi^2$-oriented labeling
- Thesaurus MI-oriented labeling
- Thesaurus $\chi^2$-oriented labeling
Internal cluster labeling

\[ C_i = [\text{bound, limit, restrain, inhibit, fasten, fix, secure, lock}] \]

Frequency oriented labeling


Verb hyperonym-oriented labeling

fasten = [\text{fasten1, fix2, fix firmly1, fasten2, link1, put together1, tie3, [...]}]

Thesaurus frequency-oriented labeling

fasten = [\text{fix, secure, attach, tighten, change, alter, modify, [...]}]
Internal cluster labeling

\[ C_i = \{\text{bound, limit, restrain, inhibit, fasten, fix, secure, lock}\} \]

**Frequency oriented labeling**


**Verb hyperonym-oriented labeling**

fasten = \{fasten1, fix2, fix firmly1, fasten2, link1, put together1, tie3, [...]\}

**Thesaurus frequency-oriented labeling**

fasten = \{fix, secure, attach, tighten, change, alter, modify, [...]\}
Internal cluster labeling

\[ C_i = \{\text{bound, limit, restrain, inhibit, fasten, fix, secure, lock}\} \]

**Frequency oriented labeling**


**Verb hyperonym-oriented labeling**

fasten = \{fasten1, fix2, fix firmly1, fasten2, link1, put together1, tie3, [...\}]

**Thesaurus frequency-oriented labeling**

fasten = \{fix, secure, attach, tighten, change, alter, modify, [...\}]


Internal cluster labeling

\[ C_i = \{ \text{bound, limit, restrain, inhibit, fasten, fix, secure, lock} \} \]

Frequency oriented labeling

\[ C_i = \{ \text{bound:63, limit:74, restrain:21, inhibit:101, fasten:49, fix:53, secure:13, lock:28} \} \]

Verb hyperonym-oriented labeling

\[
\text{fasten} = \{ \text{fasten1, fix2, fix firmly1, fasten2, link1, put together1, tie3, [...] } \}
\]

Thesaurus frequency-oriented labeling

\[
\text{fasten} = \{ \text{fix, secure, attach, tighten, change, alter, modify, [...] } \}
\]
Differential cluster labeling

- **Based on Mutual Information:**
  - VHyper MI-oriented labeling
  - Thesaurus MI-oriented labeling

\[
I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 \frac{p(x, y)}{p_1(x)p_2(y)}
\]

- **Based on \(\chi^2\) test:**
  - VHyper \(\chi^2\)-oriented labeling
  - Thesaurus \(\chi^2\)-oriented labeling

\[
\chi^2 = \sum_{a \in A} \sum_{b \in B} \frac{(O_{a,b} - E_{a,b})^2}{E_{a,b}}
\]
Differential cluster labeling

- **Based on Mutual Information:**
  - VHyper MI-oriented labeling
  - Thesaurus MI-oriented labeling

\[ I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 \frac{p(x, y)}{p_1(x)p_2(y)} \]

- **Based on \( \chi^2 \) test:**
  - VHyper \( \chi^2 \)-oriented labeling
  - Thesaurus \( \chi^2 \)-oriented labeling

\[ \chi^2 = \sum\sum \frac{(O_{a,b} - E_{a,b})^2}{E_{a,b}} \]
Examples of the internal cluster labeling strategies

<table>
<thead>
<tr>
<th>Gold Standard Clusters</th>
<th>GS</th>
<th>Freq</th>
<th>VHyper</th>
<th>ThesFreq</th>
</tr>
</thead>
<tbody>
<tr>
<td>{comprise, contain, have, include}</td>
<td>contain</td>
<td>comprise</td>
<td>comprise</td>
<td>get</td>
</tr>
<tr>
<td>{bound, limit, restrain, inhibit}</td>
<td>limit</td>
<td>inhibit</td>
<td>determine</td>
<td>bind</td>
</tr>
<tr>
<td>{tighten, fasten, fix, secure, deposit}</td>
<td>fix</td>
<td>fix</td>
<td>fix</td>
<td>attach</td>
</tr>
<tr>
<td>{compress, trim, reduce, minimize}</td>
<td>reduce</td>
<td>reduce</td>
<td>cut</td>
<td>lessen</td>
</tr>
<tr>
<td>{extract, pull-out}</td>
<td>extract</td>
<td>extract</td>
<td>remove</td>
<td>take-out</td>
</tr>
<tr>
<td>{colored, remove, cut, delete, erase, exclude}</td>
<td>remove</td>
<td>remove</td>
<td>remove</td>
<td>take-out</td>
</tr>
<tr>
<td>{enter, insert, interpose, introduce, enclose}</td>
<td>insert</td>
<td>insert</td>
<td>connect</td>
<td>introduce</td>
</tr>
<tr>
<td>{apply, feed, provide, give, use, supply, render}</td>
<td>produce</td>
<td>provide</td>
<td>provide</td>
<td>give</td>
</tr>
<tr>
<td>{hold, maintain, retain, support, prevent}</td>
<td>keep</td>
<td>support</td>
<td>maintain</td>
<td>hold</td>
</tr>
<tr>
<td>{accord, allow, let, permit}</td>
<td>let</td>
<td>accord</td>
<td>have</td>
<td>permit</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gold Standard Clusters</th>
<th>GS</th>
<th>VHyper-MI</th>
<th>VHyper $\chi^2$</th>
<th>ThesMI</th>
<th>Thes $\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>{comprise, contain, have, include}</td>
<td>contain</td>
<td>comprise</td>
<td>incorporate</td>
<td>incorporate</td>
<td>incorporate</td>
</tr>
<tr>
<td>{bound, limit, restrain, inhibit}</td>
<td>limit</td>
<td>restrict</td>
<td>restrict</td>
<td>throttle</td>
<td>restrict</td>
</tr>
<tr>
<td>{tighten, fasten, fix, secure, deposit}</td>
<td>fix</td>
<td>put</td>
<td>lay</td>
<td>find out</td>
<td>localise</td>
</tr>
<tr>
<td>{compress, trim, reduce, minimize}</td>
<td>reduce</td>
<td>trim down</td>
<td>thin-out</td>
<td>find-out</td>
<td>minify</td>
</tr>
<tr>
<td>{extract, pull-out}</td>
<td>extract</td>
<td>move forcibly</td>
<td>pull-up</td>
<td>pull-up</td>
<td>press-out</td>
</tr>
<tr>
<td>{remove, cut, delete, erase, exclude}</td>
<td>remove</td>
<td>erase</td>
<td>kill</td>
<td>cancel</td>
<td>take-out</td>
</tr>
<tr>
<td>{enter, insert, interpose, introduce, enclose}</td>
<td>insert</td>
<td>shut-it</td>
<td>enclose</td>
<td>pull-in</td>
<td>pull-in</td>
</tr>
<tr>
<td>{apply, feed, provide, give, use, supply, render}</td>
<td>produce</td>
<td>administer</td>
<td>furnish</td>
<td>furnish</td>
<td>furnish</td>
</tr>
<tr>
<td>{hold, maintain, retain, support, prevent}</td>
<td>keep</td>
<td>hold on</td>
<td>hold on to</td>
<td>defend</td>
<td>defend</td>
</tr>
<tr>
<td>{accord, allow, let, permit}</td>
<td>let</td>
<td>grant</td>
<td>grant</td>
<td>consent</td>
<td>consent</td>
</tr>
</tbody>
</table>
Cluster labeling fallback strategies

- Frequency labeling

<table>
<thead>
<tr>
<th>Label candidate</th>
<th>MI</th>
<th>Corpus frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>come near</td>
<td>1390.48</td>
<td>2</td>
</tr>
<tr>
<td><strong>fill</strong></td>
<td>1390.48</td>
<td><strong>87</strong></td>
</tr>
<tr>
<td>get together</td>
<td>1390.48</td>
<td>0</td>
</tr>
</tbody>
</table>

- Random fallback

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Random label</th>
</tr>
</thead>
<tbody>
<tr>
<td>converge, meet, satisfy</td>
<td>get together</td>
</tr>
<tr>
<td>record, file</td>
<td>impeach</td>
</tr>
<tr>
<td>accord, allow, let, permit</td>
<td>grant</td>
</tr>
</tbody>
</table>
Cluster labeling fallback strategies

- Frequency labeling

<table>
<thead>
<tr>
<th>Label candidate</th>
<th>MI</th>
<th>Corpus frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>come near</td>
<td>1390.48</td>
<td>2</td>
</tr>
<tr>
<td>fill</td>
<td>1390.48</td>
<td>87</td>
</tr>
<tr>
<td>get together</td>
<td>1390.48</td>
<td>0</td>
</tr>
</tbody>
</table>

- Random fallback

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Random label</th>
</tr>
</thead>
<tbody>
<tr>
<td>converge, meet, satisfy</td>
<td>get together</td>
</tr>
<tr>
<td>record, file</td>
<td>impeach</td>
</tr>
<tr>
<td>accord, allow, let, permit</td>
<td>grant</td>
</tr>
</tbody>
</table>
Cluster labeling fallback strategies

- Frequency labeling

<table>
<thead>
<tr>
<th>Label candidate</th>
<th>MI</th>
<th>Corpus frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>come near</td>
<td>1390.48</td>
<td>2</td>
</tr>
<tr>
<td>fill</td>
<td>1390.48</td>
<td>87</td>
</tr>
<tr>
<td>get together</td>
<td>1390.48</td>
<td>0</td>
</tr>
</tbody>
</table>

- Random fallback

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Random label</th>
</tr>
</thead>
<tbody>
<tr>
<td>converge, meet, satisfy</td>
<td>get together</td>
</tr>
<tr>
<td>record, file</td>
<td>impeach</td>
</tr>
<tr>
<td>accord, allow, let, permit</td>
<td>grant</td>
</tr>
</tbody>
</table>
Cluster labeling fallback strategies

- Frequency labeling

<table>
<thead>
<tr>
<th>Label candidate</th>
<th>MI</th>
<th>Corpus frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>come near</td>
<td>1390.48</td>
<td>2</td>
</tr>
<tr>
<td>fill</td>
<td>1390.48</td>
<td>87</td>
</tr>
<tr>
<td>get together</td>
<td>1390.48</td>
<td>0</td>
</tr>
</tbody>
</table>

- Random fallback

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Random label</th>
</tr>
</thead>
<tbody>
<tr>
<td>converge, meet, satisfy</td>
<td>get together</td>
</tr>
<tr>
<td>record, file</td>
<td>impeach</td>
</tr>
<tr>
<td>accord, allow, let, permit</td>
<td>grant</td>
</tr>
</tbody>
</table>
### Cluster labeling fallback strategies

**- Frequency labeling**

<table>
<thead>
<tr>
<th>Label candidate</th>
<th>MI</th>
<th>Corpus frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>come near</td>
<td>1390.48</td>
<td>2</td>
</tr>
<tr>
<td>fill</td>
<td>1390.48</td>
<td>87</td>
</tr>
<tr>
<td>get together</td>
<td>1390.48</td>
<td>0</td>
</tr>
</tbody>
</table>

**- Random fallback**

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Random label</th>
</tr>
</thead>
<tbody>
<tr>
<td>converge, meet, satisfy</td>
<td>get together</td>
</tr>
<tr>
<td>record, file</td>
<td>impeach</td>
</tr>
<tr>
<td>accord, allow, let, permit</td>
<td>grant</td>
</tr>
</tbody>
</table>
Illustration of the approach

**comprise** (automatic focusing device, lens driving circuit)

**focus** (objective lens, light beam)

**emit** (light source, light beam, track of an information recording medium)

**reflect** (information recording medium, reflected light beam, focal spot, objective lens)

**emit** (light source, light beam)

**be-capable** (optical element, **cause** (optical element, astigmatic **aberration** (separated reflected light beam)))

**include** (astigmatic optical system, optical element)

**extend** (second axis, **perpendicularly** (first axis))
Illustration of the approach

**comprise** (automatic focusing device, lens driving circuit)

**focus** (objective lens, light beam)

**point** (light source, light beam, track of an information recording medium)

**reflect** (information recording medium, reflected light beam, focal spot, objective lens)

**point** (light source, light beam)

**be-capable** (optical element, **cause** (optical element, astigmatic **aberration** (separated reflected light beam)))

**comprise** (astigmatic optical system, optical element)

**extend** (second axis, **perpendicularly** (first axis))
Outline

1. Introduction
2. Theoretical Framework
3. Related Work
4. Approach
5. Evaluation
6. Conclusions
### Evaluation: simplification

#### Evaluation of the claim segmentation:

<table>
<thead>
<tr>
<th></th>
<th># automatic segments</th>
<th># 1:1 alignments</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>4078</td>
<td>3282</td>
<td>0.52</td>
</tr>
<tr>
<td>Our system</td>
<td>5342</td>
<td>4327</td>
<td>0.73</td>
</tr>
</tbody>
</table>

#### Evaluation of coreference resolution:

<table>
<thead>
<tr>
<th># manual coref.</th>
<th># automatic coref.</th>
<th># automatic correct</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>199</td>
<td>190</td>
<td>159</td>
<td>0.81</td>
</tr>
</tbody>
</table>
### Evaluation of claim structuring:

<table>
<thead>
<tr>
<th></th>
<th># automatic spans</th>
<th># correct spans</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect input</td>
<td>201</td>
<td>114</td>
<td>0.56</td>
</tr>
<tr>
<td>Raw</td>
<td>143</td>
<td>75</td>
<td>0.42</td>
</tr>
<tr>
<td>Baseline</td>
<td>227</td>
<td>79</td>
<td>0.35</td>
</tr>
</tbody>
</table>
Evaluation: relation extraction

- Comparative evaluation
- Syntactic parsing vs. chunking
- Blohm & Cimiano (2007)
- Small evaluation set (5 patents)

<table>
<thead>
<tr>
<th>Manual extraction</th>
<th>Blohm &amp; Cimiano</th>
<th>Our approach</th>
</tr>
</thead>
<tbody>
<tr>
<td># 94</td>
<td># 51 (54%)</td>
<td># 67 (71%)</td>
</tr>
</tbody>
</table>
Evaluation: relational clustering

- **Gold standard:**
  - Corpus of 2076 English patent claims
  - 193 most frequent verbs
  - 54 classes

- **Clustering baselines**
  - (a) Verbs are randomly assigned to a cluster (50 repetitions)
  - (b) The total of manual clusters and their category sizes are preserved and verbs are randomly assigned to them (50 repetitions)

- **Clustering evaluation measures:**
  - Purity
  - Adjusted Rand Index
  - Pair-wise evaluation
## Evaluation: relation clustering results

<table>
<thead>
<tr>
<th>Eval. measures</th>
<th>Baseline (a)</th>
<th>Baseline (b)</th>
<th>Our clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purity</td>
<td>0.33</td>
<td>0.34</td>
<td>0.73</td>
</tr>
<tr>
<td>ARI</td>
<td>0.02</td>
<td>0.01</td>
<td>0.34</td>
</tr>
<tr>
<td>Pair-wise Recall</td>
<td>0.01</td>
<td>0.03</td>
<td>0.32</td>
</tr>
<tr>
<td>Pair-wise Precision</td>
<td>0.02</td>
<td>0.03</td>
<td>0.33</td>
</tr>
<tr>
<td>Pair-wise F-score</td>
<td>0.02</td>
<td>0.03</td>
<td>0.32</td>
</tr>
</tbody>
</table>
### Evaluation: WSD-based similarity vs. automatic similarity

<table>
<thead>
<tr>
<th>Eval. measures</th>
<th>WSD Verbs</th>
<th>WSD Our Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purity</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>ARI</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>Pair wise Recall</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Pair wise Precision</td>
<td>0.06</td>
<td>0.1</td>
</tr>
<tr>
<td>Pair wise F-score</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Evaluation: relational cluster labeling

Results of internal labeling:

<table>
<thead>
<tr>
<th></th>
<th>% Correct</th>
<th>% Partially correct</th>
<th>% Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold standard</td>
<td>77%</td>
<td>17%</td>
<td>7%</td>
</tr>
<tr>
<td>Freq</td>
<td>78%</td>
<td>20%</td>
<td>2%</td>
</tr>
<tr>
<td>VHyper</td>
<td>43%</td>
<td>25%</td>
<td>32%</td>
</tr>
<tr>
<td>ThesFreq</td>
<td>58%</td>
<td>26%</td>
<td>16%</td>
</tr>
</tbody>
</table>

Results of differential labeling:

<table>
<thead>
<tr>
<th></th>
<th>% Correct</th>
<th>% Partially correct</th>
<th>% Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold standard</td>
<td>77%</td>
<td>17%</td>
<td>7%</td>
</tr>
<tr>
<td>VHyper-MI</td>
<td>50%</td>
<td>45%</td>
<td>5%</td>
</tr>
<tr>
<td>VHyper $\chi^2$</td>
<td>60%</td>
<td>27%</td>
<td>13%</td>
</tr>
<tr>
<td>ThesMI</td>
<td>70%</td>
<td>25%</td>
<td>5%</td>
</tr>
<tr>
<td>Thes $\chi^2$</td>
<td>67%</td>
<td>22%</td>
<td>11%</td>
</tr>
</tbody>
</table>
Evaluation: relational cluster labeling

Results of internal labeling:

<table>
<thead>
<tr>
<th></th>
<th>% Correct</th>
<th>% Partially correct</th>
<th>% Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold standard</td>
<td>77%</td>
<td>17%</td>
<td>7%</td>
</tr>
<tr>
<td>Freq</td>
<td>78%</td>
<td>20%</td>
<td>2%</td>
</tr>
<tr>
<td>VHyper</td>
<td>43%</td>
<td>25%</td>
<td>32%</td>
</tr>
<tr>
<td>ThesFreq</td>
<td>58%</td>
<td>26%</td>
<td>16%</td>
</tr>
</tbody>
</table>

Results of differential labeling:

<table>
<thead>
<tr>
<th></th>
<th>% Correct</th>
<th>% Partially correct</th>
<th>% Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold standard</td>
<td>77%</td>
<td>17%</td>
<td>7%</td>
</tr>
<tr>
<td>VHyper-MI</td>
<td>50%</td>
<td>45%</td>
<td>5%</td>
</tr>
<tr>
<td>VHyper $\chi^2$</td>
<td>60%</td>
<td>27%</td>
<td>13%</td>
</tr>
<tr>
<td>ThesMI</td>
<td>70%</td>
<td>25%</td>
<td>5%</td>
</tr>
<tr>
<td>Thes $\chi^2$</td>
<td>67%</td>
<td>22%</td>
<td>11%</td>
</tr>
</tbody>
</table>
Contributions

The first linguistic study about patent claim style
- facilitates the development of NLP technologies
- supports patent users

A framework for flexible, $n$-ary and unlexicalized RE
- distilled from deep syntactic structures
- not limited in type, number or arity

A framework for unsupervised relation generalization
- the approach is simple and independent of the corpus
- relation classes are provided by cluster labeling

A claim paraphrasing application
- relies on surface linguistic analysis
- claims are more readable
Contributions

The first linguistic study about patent claim style
- facilitates the development of NLP technologies
- supports patent users

A framework for flexible, $n$-ary and unlexicalized RE
- distilled from deep syntactic structures
- not limited in type, number or arity

A framework for unsupervised relation generalization
- the approach is simple and independent of the corpus
- relation classes are provided by cluster labeling

A claim paraphrasing application
- relies on surface linguistic analysis
- claims are more readable
Contributions

The first linguistic study about patent claim style
- facilitates the development of NLP technologies
- supports patent users

A framework for flexible, $n$-ary and unlexicalized RE
- distilled from deep syntactic structures
- not limited in type, number or arity

A framework for unsupervised relation generalization
- the approach is simple and independent of the corpus
- relation classes are provided by cluster labeling

A claim paraphrasing application
- relies on surface linguistic analysis
- claims are more readable
Contributions

The first linguistic study about patent claim style
- facilitates the development of NLP technologies
- supports patent users

A framework for flexible, \textit{n}-ary and unlexicalized RE
- distilled from deep syntactic structures
- not limited in type, number or arity

A framework for unsupervised relation generalization
- the approach is simple and independent of the corpus
- relation classes are provided by cluster labeling

A claim paraphrasing application
- relies on surface linguistic analysis
- claims are more readable
Contributions

The first linguistic study about patent claim style
- facilitates the development of NLP technologies
- supports patent users

A framework for flexible, \( n \)-ary and unlexicalized RE
- distilled from deep syntactic structures
- not limited in type, number or arity

A framework for unsupervised relation generalization
- the approach is simple and independent of the corpus
- relation classes are provided by cluster labeling

A claim paraphrasing application
- relies on surface linguistic analysis
- claims are more readable
Limitations

- claims vs. other patent sections

- some claim types have not been covered

- our approach is English-specific
Limitations

- claims vs. other patent sections

- some claim types have not been covered

- our approach is English-specific
Limitations

- claims vs. other patent sections
- some claim types have not been covered
- our approach is English-specific
Applications

- Content visualization
- Ontology learning
- Search of similar documents
- Question answering

* Support patent users

* Support patent technologies development
Future work

- **Improvement of the current approach**
  - Dispense with the claim simplification module
  - Use a more powerful and patent-adapted syntactic dependency parser

- **Extension of the current approach**
  - Apply the RE approach to distil relations expressed by predicative nouns and adjectives
  - Propose an approach to group similar nouns and similar adjectives
Future work

- **Improvement of the current approach**
  - Dispense with the claim simplification module
  - Use a more powerful and patent-adapted syntactic dependency parser

- **Extension of the current approach**
  - Apply the RE approach to distil relations expressed by predicative nouns and adjectives
  - Propose an approach to group similar nouns and similar adjectives
Future work

- **Improvement of the current approach**
  - Dispense with the claim simplification module
  - Use a more powerful and patent-adapted syntactic dependency parser

- **Extension of the current approach**
  - Apply the RE approach to distil relations expressed by predicative nouns and adjectives
  - Propose an approach to group similar nouns and similar adjectives
Thank you!

Questions?