Nominalizations and nationality expressions: A corpus analysis

Daniel Berndt\textsuperscript{1}  Gemma Boleda\textsuperscript{2}  Berit Gehrke\textsuperscript{3}
Louise McNally\textsuperscript{3}

University of Osnabrück
Universitat Politècnica de Catalunya
Universitat Pompeu Fabra

CLC 2009
July 21-23, 2009
Two strategies for expressing participant roles with event nominalizations:

(1) **French** agreement to participate in the negotiations

(2) agreement **by France** to participate in the negotiations

**Question:** What determines the choice between these options?

**The goal of this talk:** To present the initial results of a corpus study which will help answer this question.
Outline

1. Introduction
2. More on the data and issues
3. Previous proposals
4. Method
5. Analysis
   - Noun type
   - Concept stability
The adjective variant generally involves so-called relational adjectives (Bally, 1944; Levi, 1978):

(3) presidential, molecular, solar, cerebral, ...

Our study is limited to a subclass of these, the so-called ethnic adjectives, or EAs (Alexiadou and Stavrou, to appear):

(4) French, Spanish, American, South Korean, ...
Two basic uses for EAs

“Argument-saturating” use:
(5) **French** agreement to participate in the negotiations

“Classifying” use:
(6) **French** bread

See e.g. Bosque and Picallo (1996), though most theoretical work deals with the argument-saturating use only.
The theoretical issues these adjective raise

- We must account for the apparent synonymy between *French agreement* and *agreement by France* . . .
  - . . . but it would be desirable to give a single semantic analysis of the EA in its two basic uses.
- The adjective has been claimed to saturate an argument of the nominalization (Alexiadou and Stavrou, to appear) . . .
  - . . . but adjectives generally *modify*, they don’t *saturate*.

An understanding of the distribution of the adjective and prepositional variants should shed light on these issues.

For illustration we will outline Alexiadou and Stavrou (A&S).
Details of A&S’s analysis

EAs are morphologically deficient nouns that become adjectives in the course of the syntactic derivation.

- EAs originate in the specifier of the noun phrase they modify, where A&S assume the agent role is always assigned.

- Since EAs are not valued for case, a morphological requirement forces them to move to the specifier of a higher adjectival phrase; in this position they are spelled out as adjectives.
A&S’s predictions

- The adjective must denote an agent (following Kayne, 1984, and others).
  - It is syntactically base-generated in the position where agents are base-generated.
- EAs are not gradable and cannot be coordinated with ‘normal’ adjectives, but only with other EAs.

(7) *French and strong agreement

- EAs are not proper adjectives but rather nouns underlyingly.
- EAs do not license anaphora (Postal, 1969).

(8) *The French agreement to . . . They/It . . .

- The underlying noun is morphologically deficient and is spelled out as an adjective.
Problems with A&S’s analysis

- Our corpus study shows that EAs do not necessarily relate to agents.
  - Examples: *the Vietnamese arrivals*, *our French investment*, *the Indian debate*

- A&S themselves claim that the nominalizations combining with EAs in Greek lack argument structure (following Grimshaw, 1990).
  - Therefore, (1) cannot be a real argument saturating use.
Problems with A&S’s analysis

- The coordination and gradability facts are not an argument against treating EAs as proper adjectives: there are other adjectives that do not coordinate freely, and many are not gradable (McNally and Boleda, 2004; Gehrke and McNally, to appear)

- Lack of anaphora to EAs is expected without assuming that they are underlying nouns.
Why a corpus study is useful

We provide basic data on the distribution of EAs vs. PPs:
- The sorts of nouns EAs vs. PPs occur with
- The variation in the head nouns that EAs occur with

This will ground our arguments against A&S’s analysis and provide a departure point for a new theoretical analysis.
Data

- British National Corpus
- 49 nationality adjectives/nouns
  - list from Wikipedia
  - adjective *(French)* and proper noun *(France)* frequencies 1,000–30,000
- EA vs. PP examples, filtered as follows:
  - infrequent head nouns (≤ 24 occurrences)
  - nation-specific nouns (e.g., *reunification* for German), using entropy
  - for some purposes 45 manually selected nominalizations were used
Source for noun types: Top Concept Ontology (Álvez et al., 2008), WordNet based resource

- Ontological category of head nouns.
- Roughly equivalent to object-denoting (*wine*), event-denoting (*agreement*), and abstract entity-denoting (*idea*) nouns:
  - *French wine* / *wine from France*
  - *French agreement* / *agreement by France*
  - *French idea* / *idea in France*
### Noun type

<table>
<thead>
<tr>
<th>description</th>
<th>EA</th>
<th>PP</th>
<th>Total</th>
<th>% EA</th>
<th>%PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>object</td>
<td>19,816</td>
<td>10,866</td>
<td>30,682</td>
<td>64.6</td>
<td>35.4</td>
</tr>
<tr>
<td>event</td>
<td>8,813</td>
<td>10,898</td>
<td>19,711</td>
<td>44.7</td>
<td>55.3</td>
</tr>
<tr>
<td>abstract</td>
<td>2,168</td>
<td>1,986</td>
<td>4,154</td>
<td>52.2</td>
<td>47.8</td>
</tr>
</tbody>
</table>

- object-denoting nouns prefer the adjective construction
- event-denoting nouns prefer the prepositional construction
  - effect even stronger if only nominalizations are considered:

<table>
<thead>
<tr>
<th>description</th>
<th>EA</th>
<th>PP</th>
<th>Total</th>
<th>% EA</th>
<th>%PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>nominalizations</td>
<td>492</td>
<td>1,788</td>
<td>2,280</td>
<td>21.6</td>
<td>78.4</td>
</tr>
</tbody>
</table>
object-denoting nouns prefer the adjective construction

- event-denoting nouns prefer the prepositional construction
  - effect even stronger if only nominalizations are considered:

<table>
<thead>
<tr>
<th>description</th>
<th>EA</th>
<th>PP</th>
<th>Total</th>
<th>% EA</th>
<th>% PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>object</td>
<td>19,816</td>
<td>10,866</td>
<td>30,682</td>
<td>64.6</td>
<td>35.4</td>
</tr>
<tr>
<td>event</td>
<td>8,813</td>
<td>10,898</td>
<td>19,711</td>
<td>44.7</td>
<td>55.3</td>
</tr>
<tr>
<td>abstract</td>
<td>2,168</td>
<td>1,986</td>
<td>4,154</td>
<td>52.2</td>
<td>47.8</td>
</tr>
<tr>
<td>nominalizations</td>
<td>492</td>
<td>1,788</td>
<td>2,280</td>
<td>21.6</td>
<td>78.4</td>
</tr>
</tbody>
</table>
### Noun type

<table>
<thead>
<tr>
<th>description</th>
<th>EA</th>
<th>PP</th>
<th>Total</th>
<th>% EA</th>
<th>%PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>object</td>
<td>19,816</td>
<td>10,866</td>
<td>30,682</td>
<td>64.6</td>
<td>35.4</td>
</tr>
<tr>
<td>event</td>
<td>8,813</td>
<td>10,898</td>
<td>19,711</td>
<td>44.7</td>
<td>55.3</td>
</tr>
<tr>
<td>abstract</td>
<td>2,168</td>
<td>1,986</td>
<td>4,154</td>
<td>52.2</td>
<td>47.8</td>
</tr>
</tbody>
</table>

- object-denoting nouns prefer the adjective construction
- event-denoting nouns prefer the prepositional construction
  - effect even stronger if only nominalizations are considered:

<table>
<thead>
<tr>
<th>nominalizations</th>
<th>EA</th>
<th>PP</th>
<th>Total</th>
<th>% EA</th>
<th>%PP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>492</td>
<td>1,788</td>
<td>2,280</td>
<td>21.6</td>
<td>78.4</td>
</tr>
</tbody>
</table>
EAs are used for stable concepts

- The adjective construction is used for stable concepts.
- *A war has started in Spain* ... → *The war in Spain* ... → *The Spanish war* ... 
- We should observe less variety in the head nouns for the adjective construction.

<table>
<thead>
<tr>
<th>Nation</th>
<th>EA</th>
<th>PP</th>
<th>Union</th>
<th>% EA</th>
<th>%PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>15</td>
<td>30</td>
<td>34</td>
<td>44.1</td>
<td>88.2</td>
</tr>
<tr>
<td>Germany</td>
<td>17</td>
<td>26</td>
<td>32</td>
<td>53.1</td>
<td>81.3</td>
</tr>
<tr>
<td>Japan</td>
<td>12</td>
<td>23</td>
<td>28</td>
<td>42.9</td>
<td>82.1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
EAs are used for stable concepts

- The adjective construction is used for stable concepts.
- *A war has started in Spain* . . .
  → *The war in Spain* . . .
  → *The Spanish war* . . .
- We should observe **less variety** in the head nouns for the adjective construction.

<table>
<thead>
<tr>
<th>Nation</th>
<th>EA</th>
<th>PP</th>
<th>Union</th>
<th>% EA</th>
<th>% PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>15</td>
<td>30</td>
<td>34</td>
<td>44.1</td>
<td>88.2</td>
</tr>
<tr>
<td>Germany</td>
<td>17</td>
<td>26</td>
<td>32</td>
<td>53.1</td>
<td>81.3</td>
</tr>
<tr>
<td>Japan</td>
<td>12</td>
<td>23</td>
<td>28</td>
<td>42.9</td>
<td>82.1</td>
</tr>
<tr>
<td>. . .</td>
<td>. .</td>
<td>. .</td>
<td>. .</td>
<td>. .</td>
<td>. .</td>
</tr>
</tbody>
</table>
EAs are used for stable concepts

- The adjective construction is used for stable concepts.
- *A war has started in Spain* . . .
  - *The war in Spain* . . .
  - *The Spanish war* . . .
- We should observe **less variety** in the head nouns for the adjective construction.

<table>
<thead>
<tr>
<th>Nation</th>
<th>EA</th>
<th>PP</th>
<th>Union</th>
<th>% EA</th>
<th>%PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>15</td>
<td>30</td>
<td>34</td>
<td>44.1</td>
<td>88.2</td>
</tr>
<tr>
<td>Germany</td>
<td>17</td>
<td>26</td>
<td>32</td>
<td>53.1</td>
<td>81.3</td>
</tr>
<tr>
<td>Japan</td>
<td>12</td>
<td>23</td>
<td>28</td>
<td>42.9</td>
<td>82.1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Adjective construction is used for stable concepts
Concept stability is an important factor in the choice between EAs and PPs.

- (Classificatory use a special case of concept stability?)
- A priori, EAs+nominalizations are unlikely candidates for stable concepts
  (cp. French bread vs. French agreement)
- This explains why nominalizations strongly prefer PPs.

- If EAs were able to saturate arguments, we would expect them to appear more frequently with, and with a wider variety of, nominalizations.
  - Our data thus argue against analyses like A&S’s.
Nominalizations and nationality expressions: A corpus analysis

Daniel Berndt
Gemma Boleda
Louise McNally

Berit Gehrke

University of Osnabrück
Universitat Politècnica de Catalunya
Universitat Pompeu Fabra

CLC 2009
July 21-23, 2009


