Semantic factors in the choice between ethnic adjectives and PP counterparts: Quantitative evidence

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

\textsuperscript{1}Codeus\textsuperscript{1}

Universitat Politècnica de Catalunya\textsuperscript{2}

Universitat Pompeu Fabra\textsuperscript{3}

QITL-4, 28-31 March 2011
A contrast and two questions

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 1:** Are these two structures really interchangeable? → This is what is expected according to the prevalent view that EAs are “nouns in disguise”.

**Question 2:** Under what conditions are **French** and similar adjectives used?
Outline

1. Introduction
2. Question 1: Are EAs equivalent to nouns?
3. Question 2: What licenses the use of EAs?
4. Semantic analysis
5. Conclusion
Outline

1. Introduction
2. Question 1: Are EAs equivalent to nouns?
3. Question 2: What licenses the use of EAs?
4. Semantic analysis
5. Conclusion
The adjective variant generally involves so-called relational adjectives (Bally, 1944; Levi, 1978; Bosque and Picallo, 1996):

(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

**Thematic use:**

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

**Classificatory use:**

(6) **French** bread

See e.g. Bosque and Picallo (1996), though most theoretical work deals with the thematic use only.

- Previous work suggests a separate treatment of these two uses, whereas a unified solution might be preferrable.
The theoretical issues these adjectives 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 . . .
  - . . . but adjectives generally modify, they don’t saturate.

We provide quantitative evidence and theoretical discussion to shed light on these issues.
Previous proposals

- Ethnic adjectives are nominal in nature (Levi, 1978; Fábregas, 2007; Alexiadou and Stavrou, to appear)
- For illustration: Alexiadou and Stavrou (A&S)
  - EAs “start out” as nouns in the syntax – the semantics “sees” the EA as a noun
  - They are true syntactic arguments, assigned the agent role by the nominalization they modify
    - EAs are syntactically base-generated in the position where agents are base-generated (Kayne, 1984)
  → EAs are morphologically deficient nouns that become adjectives in the course of the syntactic derivation

Prediction: If EAs are covert nouns, we might expect them to behave like their PP counterparts
In A&S and other work on EAs and relational adjectives more generally,

- **Claim 1**: EAs are semantically equivalent to their nominal (PP) counterparts
  → We will show that EAs modify different types of head nouns than their PP counterparts

- **Claim 2**: EAs are syntactic arguments of nominalizations
  → We will show that EAs avoid modifying relational nouns

We will address these claims to answer Question 1
Data

- British National Corpus
- 49 country names: adjective/noun
  - list from Wikipedia
  - adjective (*French*) and proper noun (*France*) frequencies 1,000–30,000
- EA vs. PP examples, filtered as follows:
  - infrequent head nouns (< 25 occurrences)
  - country-specific head 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 nouns:

<table>
<thead>
<tr>
<th>Concept</th>
<th>Descr.</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1stOrderEntity</td>
<td>object</td>
<td>French wine / wine from France</td>
</tr>
<tr>
<td>2ndOrderEntity</td>
<td>event</td>
<td>French agreement / agreement by France</td>
</tr>
<tr>
<td>3rdOrderEntity</td>
<td>abstract</td>
<td>French idea / idea in France</td>
</tr>
</tbody>
</table>

**Table:** Main categories in the TCO
Type of head noun: general trends

Ratio Distribution over Concepts

- Tokens
- Types

preferences: EA, PP

Types: Tokens

Figure: Ratio of EA vs. PP constructions for nodes in the TCO ontology
Event-denoting nouns prefer PP

- exception:
  - **Modal**: force, authority, power, influence, control, proposal, intelligence, passport

- apart from proposal, these are not eventive words, though they are certainly abstract (except for passport)

- (inclusion in) category Modal is questionable
Type of head noun: exceptions to EA preference

Object-denoting nouns prefer EA
- exceptions:
  - **Part**: border, part, city, province, north, . . .
  - **Place**: territory, country, coast, region, side, . . .
- relational nouns (part-whole, spatial)
  7) border with France, north of Italy, city near Germany
- nouns that function as appositions
  8) country of Spain
- EA version ok in some cases (9), not in others (10):
  9) Spanish border, Italian city
  10) #Italian north, #French part
Type of head noun: nominalizations

Pooled data for main categories (tokens):

<table>
<thead>
<tr>
<th>Concept</th>
<th>EA</th>
<th>PP</th>
<th>Total</th>
<th>%EA</th>
<th>%PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1stOrderEntity</td>
<td>13296</td>
<td>7752</td>
<td>21048</td>
<td>63.2</td>
<td>36.8</td>
</tr>
<tr>
<td>2ndOrderEntity</td>
<td>5667</td>
<td>8048</td>
<td>13716</td>
<td>41.3</td>
<td>58.7</td>
</tr>
<tr>
<td>3rdOrderEntity</td>
<td>1539</td>
<td>1528</td>
<td>3067</td>
<td>50.2</td>
<td>49.8</td>
</tr>
</tbody>
</table>

Effect even stronger if only nominalizations are considered:

<table>
<thead>
<tr>
<th></th>
<th>EA</th>
<th>PP</th>
<th>Total</th>
<th>% EA</th>
<th>% PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>nominalizations</td>
<td>482</td>
<td>1135</td>
<td>1617</td>
<td>29.8</td>
<td>70.2</td>
</tr>
</tbody>
</table>
Discussion: EAs are not equivalent to nouns

- Ethnic adjectives and PP counterparts modify different types of nouns
  - EAs prefer object-denoting >> event- or abstract-denoting head nouns
  - Classificatory >> thematic use

- If the two constructions were interchangeable, we would not expect this difference

→ Therefore, EAs are not semantically equivalent to their nominal (PP) counterparts (against Claim 1)
Discussion: EAs are not arguments

- EAs strongly avoid modifying relational nouns
  - Nominalizations
  - Other (Part, Place)
- If EAs were arguments, why shouldn’t they freely combine with relational nouns?
  - EAs are **not** syntactic arguments of nominalizations (against Claim 2)
    - Adjectives cannot introduce referents, cannot saturate argument positions
    - EAs do not license anaphora (Postal, 1969):

  \[(11) \quad *\text{The French}_i \text{ agreement to } \ldots \text{ They}/\text{It}_i \ldots \]
Outline

1. Introduction
2. Question 1: Are EAs equivalent to nouns?
3. Question 2: What licenses the use of EAs?
4. Semantic analysis
5. Conclusion
EAs are used for “stable relations”

- EAs are less explicit than their PP counterparts with regard to the semantic relation between head and modifier.

  **Hypothesis**: The EA is licensed when the relation is clear either from background knowledge or the discourse context.

(12) Background knowledge:

A war has started in Spain . . .
→ The war in Spain . . .
→ The Spanish war . . .
Discourse context:

(13) ...the prime minister added: “I can certainly confirm that the decision in Denmark is for the Danes and I see no external pressure being put on them; but it is a matter for the Danes and for their Government to decide.” The problem was that in reality it was not so much the Danes who had to decide –they had already made up their minds– as the remainder of the EEC partners who had to face the consequences of the Danish decision. (BNC; McNally (2010))
EAs are used for “stable relations”

Discourse context:

(14) Yeltsin met the prospective Democratic presidential candidate Bill Clinton on June 18. His itinerary also included ??an official Canadian visit

(15) Yeltsin met the prospective Democratic presidential candidate Bill Clinton on June 18. His itinerary also included an official visit to Canada (BNC; Arsenijević et al. (to appear))
The hypothesis predicts that EAs appear with a narrower range of nominalizations than PP counterparts:

<table>
<thead>
<tr>
<th>Country</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 appear with a limited set of nominalizations

Paired $t$-test, $p$-value < $10^{-8}$
EAs appear with a limited set of nouns

What if we look at all head nouns?

A: 49 countries

B: 29 most infrequent countries

Paired t-test, p-value: A < 0.05, B < 0.001
Frequent vs. infrequent countries

- Frequent countries: more relation with UK, more established relations $\rightarrow$ higher use of EA

<table>
<thead>
<tr>
<th>Position</th>
<th>Country</th>
<th>EA</th>
<th>PP</th>
<th>log(EA/PP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Germany</td>
<td>3400</td>
<td>2104</td>
<td>0.48</td>
</tr>
<tr>
<td>3</td>
<td>Ireland</td>
<td>3367</td>
<td>1827</td>
<td>0.61</td>
</tr>
<tr>
<td>28</td>
<td>Mexico</td>
<td>406</td>
<td>382</td>
<td>0.06</td>
</tr>
<tr>
<td>29</td>
<td>Romania</td>
<td>396</td>
<td>368</td>
<td>0.07</td>
</tr>
<tr>
<td>48</td>
<td>Jordan</td>
<td>140</td>
<td>249</td>
<td>-0.58</td>
</tr>
<tr>
<td>49</td>
<td>Thailand</td>
<td>107</td>
<td>206</td>
<td>-0.66</td>
</tr>
</tbody>
</table>
Use of EAs and PPs vs. frequency of the country

Correlation test (Spearman): $\rho=0.46$, $p$-value=0.001
Discussion: EAs for stable semantic relations

- Our data supports the hypothesis that EAs are used for “stable semantic relations”, that is, relations that are established
  - in the background knowledge of the speaker and the hearer; or,
  - in the discourse context previous to the mention of the EA.

- Two factors:
  - Less lexical material to indicate the relation.
  - Ethnic adjectives denote *individual*-level properties, thus stable properties.
Claims 1 and 2 above lead

- Fábregas (2007) to posit that EAs are defective nouns (both thematic and classificatory uses)
- Alexiadou and Stavrou (to appear) to posit that thematic EAs and classificatory EAs are merely homophonous

Challenging them, and examining further the conditions for EA licensing, leads us to posit that EAs are adjectives

→ Uniform analysis for thematic and classificatory uses (Arsenijević et al., to appear)
Unified semantic representation

- Semantic analysis based on McNally and Boleda (2004)
- EAs are intersective modifiers of kinds

(16) $[[\text{French wine}]]: \lambda x_k \lambda y [R(y, x_k) \land \text{wine}(x_k) \land R(x_k, \text{France})]$  

(17) $[[\text{French discovery}]]: \lambda x_k \lambda y [R(y, x_k) \land \text{discovery}(x_k) \land R(x_k, \text{France})]$
The $R$ relation: **Origin**

- In classificatory uses of EAs, a head noun is subclassified according to the **Origin** of the head noun in one country.

(18) *French bread, Italian style, German student*

→ For EAs the $R$ relation is **Origin**:  

(19) $\text{Origin}(x, y)$ iff $x$ comes into existence within the spatial domain of $y$.

- Consistent with use as predicates of individuals:

(20) Guillaume is French
Origin for thematic use

- Event nouns modified by EAs generally describe types of eventualities (Grimshaw, 1990; Van de Velde, 2004).
- Eventualities have as their origins the individuals who cause, initiate, or control them.
- Correct prediction: not only agents (Kayne, 1984; Alexiadou and Stavrou, to appear) (21), but also certain nonagents (22)

(21) the Italian attack on Ethiopia
(22) the British arrival on the American continent in the 17th century (THEME; UNACCUSATIVE)
Origin for thematic use

- Analysis further explains:
  
  (23) the French disappearance from Upper Louisiana
  \((\text{THEME}; \text{CONTROL})\)

  (24) ??the French disappearance from the list of nations that haven’t approved the treaty \((\text{THEME}; \text{NO CONTROL})\)

  (25) the Italian love for opera \((\text{EXPERIENCER}; \text{PSYCH. PREDICATE})\)

- Analysis does not explain:

  (26) the Italian table
  \(=\text{table such that the majority of people sitting at it are Italian}\) (L. McNally, p.c.)
Discussion: Semantic analysis

Semantic analysis explains

- thematic and classificatory uses of EA
- similarities between EA and PP:

\[(27) \quad \text{[French]}: \lambda P_k \lambda x_k \lambda y [R(y, x_k) \land P_k(x_k) \land R(x_k, \text{France})] \]

- many of the phenomena that lead to the nominal account
  - predominant agentive interpretation for nominalization modifiers
- EAs’ “allergy” to nominalizations (not arguments)

It does not explain

- non-\textit{Origin} uses of EAs (\textit{Italian table})
- use of EA $>$ PP for established relations
Outline

1. Introduction
2. Question 1: Are EAs equivalent to nouns?
3. Question 2: What licenses the use of EAs?
4. Semantic analysis
5. Conclusion
Conclusions

- We have shown that EAs
  - modify different types of head nouns than their PP counterparts
  - avoid modifying nominalizations
  - appear with a narrow set of head nouns, compared to their PP counterparts
  - are used more for established semantic relations
- Explanatory hypothesis: EAs are licensed by background knowledge or discourse anaphora
Conclusions

- Our research suggests that an analysis of EAs as “nouns in disguise” is not adequate.
  - supports a unified semantic analysis of the classificatory and thematic uses of EAs which treats them as proper adjectives.

- We propose one such account, in which EAs
  - are properties of kinds involving the Origin relation,
  - are not proper arguments,
    - but the argument-like interpretation arises when EAs modify event nominals as a result of the interaction between the semantics of the adjective and that of the noun.
Future work

- EAs likely to occur in collocations / multi-word expressions
  - analyze the relation between EA constructions and collocations
  - use a collocation index to assess “establishedness of relations”

- Explore relation between attributive and predicative uses

(28) Guillaume is French
(29) This chair is Italian
(30) #This kind of chair is Italian
(31) #The agreement was French
Semantic factors in the choice between ethnic adjectives and PP counterparts: Quantitative evidence

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

Codeus\textsuperscript{1}  
Universitat Politècnica de Catalunya\textsuperscript{2}  
Universitat Pompeu Fabra\textsuperscript{3}

QITL-4, 28-31 March 2011


Details of A&S’s analysis

(32) germaniki epithesi
    German attack

DP

... FP/AGRP

a(sp)P
a(sp)′
a(sp)°

F′
F
nP

DP
n′

vP

\sqrt{EPITH}
A&S’s predictions

- The adjective must denote an agent (Kayne, 1984).
  - 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.
  
  \[(33) \quad *\text{French and strong agreement.}\]

  - EAs are not proper adjectives but rather nouns underlyingly.
- EAs do not license anaphora (Postal, 1969).
  
  \[(34) \quad *\text{The French}_{i} \text{ agreement to . . . They/It}_{i} . . .}\]

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

- 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 (Grimshaw, 1990).

→ Therefore, (1) cannot be a real argument saturating use.
Problems with A&S’s analysis II: coordination, gradability, and anaphora

- There are other adjectives that do not coordinate freely, and many are not gradable (McNally and Boleda, 2004; Gehrke and McNally, to appear, and references therein)
- Lack of anaphora to EAs is expected without assuming that they are underlying nouns.
  → Therefore, the gradability, coordination, and anaphora facts are independent factors