National identity and International Language
France, Belgium and the Netherlands

Marc van Oostendorp

Meertens Instituut / KNAW, Amsterdam

Workshop English in Europe: from FL to L2
ICLaVE 3
June 23, 2005
National Identity and International Language

Language and national identity
   Starting point

Linguistic policies
   Similarities and differences

Corpora
   Newspaper archives vs. Google
   Google

Conclusions
   Conclusions
National Identity and International Language

Language and national identity
   Starting point

Linguistic policies
   Similarities and differences

Corpora
   Newspaper archives vs. Google
   Google

Conclusions
   Conclusions
Language and national identity

- It has often been claimed that there is an intimate relation between (national) language and (national) identity.
- Further, it is claimed that this works out differently for different nations.
- If this is true, we expect different nations to react differently to the rise of English as an international (European) language.
An empirical issue

- The issue of the relation between (inter)national language and national identity is an empirical one.
- In particular, how is official policy related to national discourse?
- The approach taken here is an empirical one, based on corpus study:
  - High-brow newspapers
  - The World-wide Web as a corpus
- Do we find differences between European nations and to they correspond to formal policies?
A methodological issue

- The answers may be found in careful analysis of open discussions (next to other means)
- How can we use electronic text corpora to study such questions?
National Identity and International Language

Language and national identity
  Starting point

Linguistic policies
  Similarities and differences

Corpora
  Newspaper archives vs. Google
  Google

Conclusions
  Conclusions
France, Belgium, The Netherlands

• Three countries which can be seen as ‘central’ to Europe and European institutions

• There has been a long history of (monodirectional) influence, both extralinguistic and linguistic:
  • Code Napoleon has strongly influenced Dutch and Belgium civil law. Van Hemel 1998
  • French (still) is the major source of loanwords in Dutch (25% of the lexicon, vs. 10% from English). Van der Sijs 2000
France, Belgium, The Netherlands

- Three countries which can be seen as ‘central’ to Europe and European institutions
- There has been a long history of (monodirectional) influence, both extralinguistic and linguistic:
  - Code Napoleon has strongly influenced Dutch and Belgium civil law. Van Hemel 1998
  - French (still) is the major source of loanwords in Dutch (25% of the lexicon, vs. 10% from English). Van der Sijs 2000
France, Belgium, The Netherlands

- Three countries which can be seen as ‘central’ to Europe and European institutions
- There has been a long history of (monodirectional) influence, both extralinguistic and linguistic:
  - Code Napoleon has strongly influenced Dutch and Belgium civil law. Van Hemel 1998
  - French (still) is the major source of loanwords in Dutch (25% of the lexicon, vs. 10% from English). Van der Sijs 2000
France, Belgium, The Netherlands

- Three countries which can be seen as ‘central’ to Europe and European institutions
- There has been a long history of (monodirectional) influence, both extralinguistic and linguistic:
  - Code Napoleon has strongly influenced Dutch and Belgium civil law. Van Hemel 1998
  - French (still) is the major source of loanwords in Dutch (25% of the lexicon, vs. 10% from English). Van der Sijs 2000
France

- France has an official policy to promote the French language:
  - «la langue de la République est le français» (article 2 of Constitution)
  - Efforts to maintain and extend the international community of *francophonie*
  - Within Europe, and on a global scale, France promotes language diversity; according to some, mainly to promote French as a ‘second foreign language’ (after English)
  - Within France, attempt to ratify the European Charter for Regional or Minority Languages have failed

France

- Les principes d’indivisibilité de la République, d’égalité devant la loi et d’unicité du peuple français, qui « s’opposent à ce que soient reconnus des droits collectifs à quelque groupe que ce soit, défini par une communauté d’origine, de culture, de langue ou de croyance »[...]

Décision du Conseil Constitutionnel sur la Charte européenne des langues régionales ou minoritaires, June 1999
The Netherlands

- Attempts of getting the Dutch language ‘anchored’ (“De bevordering van het gebruik van de Nederlandse taal is voorwerp van zorg van de overheid”) in constitution failed (but new attempt on its way)
- Dutch colonialism has not led to wide-spread use of Dutch in e.g. current Indonesia
- Present Dutch (internal and external) language policies relegated to *Taalunie* (Language Union) with Belgium and Surinam
- Ratified European Charter and recognized Frisian, Low Saxonian and Limburgian in the 1990’s; in recent times, recognition of Zeelandic was blocked, after negative advice by Taalunie.

Belgium

- Three languages recognized in Constitution (German, French, Dutch)
- Belgian colonialism has led to some spread of French, but not of Dutch
- Flanders takes part in Taalunie, Wallonia in Francophonie
- Belgium did not sign European Charter

Geerts 1972; De Caluwe et al. 2003; Hambye 2005
National Identity and International Language

Language and national identity
  Starting point

Linguistic policies
  Similarities and differences

Corpora
  Newspaper archives vs. Google
  Google

Conclusions
  Conclusions
Some hypotheses

- The role of French will be more important in French discourse than the role of Dutch in Dutch discourse
- The role of the constitution should be bigger in French (and Belgian) discourse than in Dutch discourse
Newspaper archives vs. Google

- Assumption: discourse analysis will give us some insight in feelings and opinions on identity and language
- Two corpora:
  - high-brow newspapers (edited; assumed to reflect ‘official’ opinions of elite)
  - Google (unedited; assumed to reflect ‘unofficial’ opinions of general populace)
Newspaper Archives

- We studied the electronic archives of three high-brow newspapers: French *Le Monde*, Dutch *NRC Handelsblad*, Belgian (Flemish) *De Standaard*
- Dates: January 1, 2000 - June 1, 2005
- In every corpus we searched for the terms *language policy, international language, universal language* in French and Dutch The set of all articles containing at least one of these terms was our research corpus
Newspaper archives: Results (I)

- The resulting corpus was fairly small:
  - Monde: 26 articles (8008 words)
  - Standaard: 58 articles (14906 words)
  - NRC: 40 articles (12440 words)
- No firm conclusions can be drawn from these numbers as to the specific interest which is paid in every one of these newspapers, since we cannot compare the relative sizes of each of the corpora
Newspaper archives: Results (II)

<table>
<thead>
<tr>
<th>Language</th>
<th>Newspaper</th>
<th># of times</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Monde</td>
<td>75</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Standaard</td>
<td>252</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>NRC</td>
<td>120</td>
<td>36</td>
</tr>
<tr>
<td>French</td>
<td>Monde</td>
<td>299</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Standaard</td>
<td>120</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>NRC</td>
<td>68</td>
<td>20</td>
</tr>
<tr>
<td>Dutch</td>
<td>Monde</td>
<td>5</td>
<td>01</td>
</tr>
<tr>
<td></td>
<td>Standaard</td>
<td>97</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>NRC</td>
<td>74</td>
<td>22</td>
</tr>
<tr>
<td>Spanish</td>
<td>Monde</td>
<td>16</td>
<td>04</td>
</tr>
<tr>
<td></td>
<td>Standaard</td>
<td>19</td>
<td>04</td>
</tr>
<tr>
<td></td>
<td>NRC</td>
<td>19</td>
<td>06</td>
</tr>
<tr>
<td>German</td>
<td>Monde</td>
<td>3</td>
<td>01</td>
</tr>
<tr>
<td></td>
<td>Standaard</td>
<td>22</td>
<td>04</td>
</tr>
<tr>
<td></td>
<td>NRC</td>
<td>47</td>
<td>14</td>
</tr>
</tbody>
</table>
Newspaper archives: Results (III)

<table>
<thead>
<tr>
<th>Word</th>
<th>Newspaper</th>
<th># of times</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constitution</td>
<td>Monde</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Standaard</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>NRC</td>
<td>1</td>
</tr>
</tbody>
</table>
Newspaper archives: Discussion

- the national language seems more important in discussions in France than in discussions in Belgium or the Netherlands
- The term *constitution* comes up much more often in French discussions than in the Netherlands or Belgium
- *De Standaard* (surprisingly) sides with *NRC* in virtually all respects
National Identity and International Language

Language and national identity
Starting point

Linguistic policies
Similarities and differences

Corpora
Newspaper archives vs. Google
Google

Conclusions
Conclusions
According to a recent assessment, Google indexes appr. 3,000 million Dutch words; the number of French words is unknown (but larger) Van Oostendorp (2005)

The so-called Google API makes automatic large-scale searches possible.

Again we searched for the terms language policy, international language, universal language in French and Dutch But in this case, we combined this with the names of individual languages, and words such as constitution
<table>
<thead>
<tr>
<th>Language</th>
<th>Newspaper</th>
<th># of times</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Fr</td>
<td>163</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Be</td>
<td>577</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Nl</td>
<td>4420</td>
<td>42</td>
</tr>
<tr>
<td>French</td>
<td>Fr</td>
<td>153</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Be</td>
<td>525</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Nl</td>
<td>957</td>
<td>9</td>
</tr>
<tr>
<td>Dutch</td>
<td>Fr</td>
<td>32</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Be</td>
<td>426</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Nl</td>
<td>3740</td>
<td>36</td>
</tr>
<tr>
<td>Spanish</td>
<td>Fr</td>
<td>75</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Be</td>
<td>256</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Nl</td>
<td>688</td>
<td>7</td>
</tr>
<tr>
<td>Esperanto</td>
<td>Fr</td>
<td>91</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Be</td>
<td>259</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Nl</td>
<td>628</td>
<td>6</td>
</tr>
</tbody>
</table>
Google: Results (II)

<table>
<thead>
<tr>
<th>Word</th>
<th>Newspaper</th>
<th># of times</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constitution</td>
<td>Fr</td>
<td>52 (10)</td>
</tr>
<tr>
<td></td>
<td>Be</td>
<td>76 (3)</td>
</tr>
<tr>
<td></td>
<td>Nl</td>
<td>230 (2)</td>
</tr>
</tbody>
</table>
Google: Discussion

- the Google archive gives a different view on the national debate
- in particular, the differences between nations do not seem to be as big
- High-brow newspapers seem closer to general opinion
- One possible explanation is that the European public is less divergent than official policies suggest
National Identity and International Language

Language and national identity
Starting point

Linguistic policies
Similarities and differences

Corpora
Newspaper archives vs. Google
Google

Conclusions
Conclusions
Possible extensions

- Differentiation of the corpus according to period of time, genre, source of the website, etc.
- Other newspapers (in particular French language Belgian Le Soir)
- Comparison of these results with those of carefully designed questionnaires
Possible extensions

- Differentiation of the corpus according to period of time, genre, source of the website, etc.
- Other newspapers (in particular French language Belgian *Le Soir*)
- Comparison of these results with those of carefully designed questionnaires
Possible extensions

- Differentiation of the corpus according to period of time, genre, source of the website, etc.
- Other newspapers (in particular French language Belgian *Le Soir*)
- Comparison of these results with those of carefully designed questionnaires
Conclusions

- Official policies are reflected in national discourse
- Study of electronic corpora may be a useful additional tool for studying public discourse on national identity and international language
- Differences are less strongly reflected in the Google corpus than in (high-brow) newspapers
- It would be interesting to compare this with the results of e.g. questionnaires