On grammatical translationese

Diana Santos

Natural Language Group
INESC, Apartado 1306, R.Alves Redol, 9
P-1000 Lisboa, Portugal

and

Department of British and American Studies
University of Oslo, Postboks 1003 Blindern
N-0315 Oslo, Norway

diana.santos@{inesc.pt, iba.uio.no}
## Properties of a translated text

| % Properties common to all translations  | "universals of translation" |
| % Properties language-pair specific    | "translationese"            |
| % Properties text specific            | (quality of a given translation, influence of the style of the author and of the translator) |
Why study translationese?

* I am interested in the contrast and comparison of two language systems in the domain of tense and aspect through the use of real translations

* Translationese constitutes the distance between TL original texts and texts originating from translation

* By identifying translationese one can compare more accurately the two languages by compensating for it
Defining translationese

Translationese = influence of properties of the source language in a translated text in a target language.

This influence can be actualized in various forms:

1) the presence of properties of SL in a TL text
2) the absence of properties of TL in a TL text
3) the excessive use of TL properties in a TL text, due to subconscious awareness of the language differences on the part of the translator ("antitranslationese")
Formalizing translationese:
Motivation

The existence of translationese is generally acknowledged, but to my knowledge it has not been formally related to differences in language systems.

The work by Gellerstam (1986) on English lexical influence on Swedish furnished the starting point for my work, which is centered on grammatical features instead.

Contrarily to Gellerstam, I see the formalization of grammatical translationese as a preliminary step to its systematic detection.
Formalizing translationese...

**Case (a):** A $\rightarrow$ B,C,D
freq(A$\rightarrow$B) $>>$ freq (A$\rightarrow$C), freq (A$\rightarrow$D)
resulting in different distribution of B,C and D in a translated text versus an original text

**Case (b):** A+obligatory B $\rightarrow$ A' + optional B'
freq (A+B $\rightarrow$ A'+B') $>>$ freq (A+B $\rightarrow$ A')
resulting in a larger frequency of A'+B' in translated text (i.e., the value freq(A'+B')/(freq(A'alone)+freq(A'+B')) is larger in translated than in original text)
Formalizing translationese...

Case (c): $C = \text{vague}(A, B) \rightarrow A', B'$

If $A'$ is more marked than $B'$ in the target language, it is possible that $\text{freq}(C \Rightarrow B') > \text{freq}(C \Rightarrow A')$, and that in consequence the distribution of $A'$ and $B'$ differs again from translated texts to original texts.

Case (d) $C = \text{compact}(A, B) \rightarrow A' + B'$

This situation gives origin to three kinds of translation pairs $C \Rightarrow A' + B'$, $C \Rightarrow A'$ and $C \Rightarrow B'$.

Translationese will be reflected by a larger frequency of conjoined cases than in original text, i.e.

$\text{freq}(A' + B')/(\text{freq}(A') + \text{freq}(B') + \text{freq}(A' + B'))$ is greater.
Examples from English and Portuguese tense and aspect

Case (a): progressive -> estar, ir, andar
Case (b): Imperfeito (past + imperfective) -> progressive
já + Perfeito -> (already) + simple past or present perfect
Case (c): be = vague(state, inception) -> MQP = compact
(state, inception)
acquisitions = vague (state, inception)
Case (d) PPC -> pres perf + interval
predicates compact for manner and movement -> manner +
movement
### One example in detail

#### Distribution of the progressive:

<table>
<thead>
<tr>
<th></th>
<th>OE 58</th>
<th>TEP 20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TPE 98</td>
<td>OP 30</td>
</tr>
</tbody>
</table>

freq (progP => progE) = 13
freq (Imperf => progE) = 85
freq (Imperf => simple past) = 623
freq (progTPE) >> freq (progOE) ===> case (b) of translationese

**Hypotheses:**
1) Imperfeito [past + imperfective] -> past + (progressive)
2) All occurrences of Imperfeito can be analysed thus (trivially false...)
2') All occurrences of Imperfeito => prog can be analysed thus
3) past progressive in English is always optional (??)
4) there is no other translation for Imperfeito with that meaning (what about Imperfeito -> progressive + (past), in other words, gerund?) This would suggest (1') Imperfeito = compact (past, imperfective) -> past + prog as the right analysis ===> case (d) of translationese
5) there is no other factor in Portuguese that increases the frequency of progE
Discussion

What one has available are

# competence models of translation A -> B
# frequency of actual translation pairs freq (A=>B)
# distribution of grammatical markers in source, original and translated texts
  freq (A), freq (B trans) and freq (B orig)

One has, in addition, to reckon with

! the fact that several distinct markers translate into the same A,B,C,D -&gt; E
! the fact that meanings of grammatical features are not disjoint, rather they are
  structured in a sort of prototype net
! the fact that translation is often partial: A (A1,A2, A3) -&gt; B (A1')
! the fact that not all translations are optimal (some are wrong, others could be
  better given the way the target language works)
Tentative conclusions

What kind of conclusions can one come up with

? If A is significantly more frequent than B in original texts, freq (A) > freq (B trans) > freq (B orig).

? If freq (B trans) is closer to freq (A), one may conclude that the competence model influences greatly the translation;

? if freq (B trans) is closer to freq (B orig), one may conclude that the distribution of A and B is conditioned rather by language-internal features.

As regards their meanings, one can conclude that the closest freq (A) to freq (B trans) the closest A and B are to have the same meaning (shades of meaning), notwithstanding the different importance that the two languages concede to it.

? If freq (A=>B) is negligible compared to freq (A) and freq (B trans), one may conclude that

? either the way the two languages use the same meaning is different,

? or that they do not convey the same meaning.
General assumptions

- Frequencies of grammatical markers matter
- Semantic contrast of two languages has to be cast in
  - what can be said, and how
  - preferences displayed by the language on what is actually said
- Computational Linguistics should greatly benefit from looking at corpora (real performance) instead of concentrating in competence models