

Jörg Tiedemann
Department of Linguistics
Uppsala University

Recycling Translations

Extraction of Lexical Data from Parallel Corpora and their Application in Natural Language Processing

ACTA UNIVERSITATIS UPSALIENSIS

Studia Linguistica Upsaliensia 1, 2003

ISBN 91-554-5815-7

Errata

Page 4: In the paragraph about the MATS project: This project was co-ordinated by the Department of Linguistics at Uppsala University.

Page 5: In the first paragraph: The official Swedish name of the joint research program is '*Översättning och tolkning som språk- och kulturmöte*' (which I forgot to include).

Page 9: Figure 2.2 should include information about the first column in the table: In this column the type of alignment is given (e.g. 1:1 refers to a one-to-one sentence alignment).

Page 10: The reference in the first sentence on this page should be moved to the end of the sentence.

Page 14: In the section on co-occurrence measures, second paragraph: add 'some' after 'A common idea behind'.

Page 19: The first sentence of this page which starts already on page 18. Two prepositions ('for') should be changed to 'in' and 'of' : '... than words in the alignment of Dutch and English.'

Page 20: First paragraph, the last but one sentence: '... and the task of the model ...' should be changed to '... and the task in SMT...'

Page 23: First sentence that starts on the previous page: '... and model 4 ...' should be changed to '... and model 4 and 5...'

Page 27: 'Partly' in footnote 16 should be changed to 'partially' in order to be consistent with the terminology used otherwise.

Page 28: Second paragraph, the last but one sentence: '... uses “fuzzy” markers ...' should be changed to '... uses the marker “fuzzy” ...'

- Page 31:** There should be a comma after the first 'e.g.' and a space after 'e.g.' in the first sentence. In the middle of the following paragraph: 'Intersection' should be changed to 'Intersections'
- Page 33:** In the section on PLUG XML, second sentence: 'The initial delivery from' should be changed to 'The initial task in'.
- Page 38:** In the first paragraph: Change preposition 'of' to 'in' ('... over 2 million words in Swedish ...').
- Page 43:** Last paragraph, first sentence: Add '(Uplug systems)' after '... organized in sequences'
- Page 51:** In the first sentence: 'to' should be added after 'three approaches'.
- Page 55:** Last item in the description list (Estimated clue): $A_{Dice}(s, t)$ should be $C_{Dice}(s, t)$ and $A_{LCSR}(s, t)$ should be $C_{LCSR}(s, t)$
- Page 57:** In table 5.1, first column, last row: 'handbagage' should be 'handbagaget'. In figure 5.3: 'baggage' should be 'luggage'.
- Page 58:** The paragraph after the figure of the reference links, first sentence, '... a common source or target language words ...': remove the final s in 'words'
- Page 61:** In the section on 'Association measures as clues': '... LCSR is assumed to be ...' ('is' is missing).
- Page 62:** Last paragraph before the section on 'Learning clues': 'One of the advantages ...' should start a new paragraph.
- Page 65:** Last sentence in the first paragraph: 'associations' should be 'association'.
- Page 71:** Second paragraph: 'refers' should be changed to 'refer' twice in both sentences that start with 'BaseX clues ...'.
- Page 88:** In the middle of the last paragraph: 'similar ties' should be changed to 'similarities'.
- Page 93:** Before figure 6.2: '... the user may continues..' should be changed to '... the user may continue ...'.

The author apologizes for any inconveniences these errors may have caused.

Jörg Tiedemann
December, 2003