1. Record Nr. UNINA9910819095203321

Autore	Delpech Estelle Maryline
Titolo	Comparable corpora and computer-assisted translation / / Estelle Maryline Delpech ; series editor, Narendra Jussien
Pubbl/distr/stampa	London, England ; : , : Hoboken, New Jersey : , : iSTE : , : Wiley, , 2014 ©2014
ISBN	1-119-00265-6 1-119-00252-4
Descrizione fisica	1 online resource (xiv, 287 pages)
Collana	Cognitive Science and Knowledge Management Series
Disciplina	410.285
Soggetti	Computational linguistics Corpora (Linguistics) Translators (Computer programs)
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Description based upon print version of record.
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Cover; Title Page; Copyright; Contents; Acknowledgments; Introduction; PART 1: Applicative and Scientific Context; Chapter 1: Leveraging Comparable Corpora for Computer-assisted Translation ; 1.1. Introduction; 1.2. From the beginnings of machine translation to comparable corpora processing; 1.2.1. The dawn of machine translation; 1.2.2. The development of computer-assisted translation; 1.2.3. Drawbacks of parallel corpora and advantages of comparable corpora; 1.2.4. Difficulties of technical translation; 1.2.5. Industrial context 1.3. Term alignment from comparable corpora: a state-of-the-art1.3.1. Distributional approach principle; 1.3.2. Term alignment evaluation; 1.3.2.1. Precision at rank N or TopN; 1.3.2.2. MRR; 1.3.2.3. MAP; 1.3.3. Improvement and variants of the distributional approach; 1.3.3.1. Favoring distributional symmetry; 1.3.3.2. Using syntactic contexts; 1.3.3.3. Relying on trusted elements; 1.3.3.4. Improving semantic information representation; 1.3.3.5. Using second-order semantic affinities; 1.3.3.6. Improving the bilingual resource with semantic classes; 1.3.3.7. Translating polylexical units 1.3.4. Influence of data and parameters on alignment quality1.3.4.1.

Data; 1.3.4.2. Parameters; 1.3.5. Limits of the distributional approach; 1.4. CAT software prototype for comparable corpora processing; 1.4.1. Implementation of a term alignment method; 1.4.1.1. Implementation and data; 1.4.1.2. Extraction of the terms to be aligned; 1.4.1.3. Collecting context vectors; 1.4.1.3.1. Monolexical term context vectors; 1.4.1.4. Polylexical term context vectors; 1.4.1.5. Translation of the source context vectors; 1.4.1.6. Term alignment; 1.4.2. Terminological records extraction
1.4.3. Lexicon consultation interface1.5. Summary; Chapter 2: User-Centered Evaluation of Lexicons Extracted from Comparable Corpora; 2.1. Introduction; 2.2. Translation quality evaluation methodologies; 2.2.1. Machine translation evaluation; 2.2.1.1. Automatic evaluation measures; 2.2.1.2. Human MT evaluation; 2.2.2. Human translation evaluation; 2.2.2.1. Quantitative models; 2.2.2.2. Non-quantitative models; 2.2.3. Discussion; 2.3. Design and experimentation of a user-centered evaluation; 2.3.1. Methodological aspects; 2.3.1.1. Evaluation criteria and purpose
2.3.1.2. Subject matter expertise2.3.1.3. Basis for comparison; 2.3.2. Experimentation protocol; 2.3.2.1. Data; 2.3.2.1.1. Comparable corpora and extracted lexica; 2.3.2.1.2. Texts to be translated; 2.3.2.1.3. Resources used in the translation situation; 2.3.2.1.4. Translators and judges; 2.3.2.2. Evaluation progress; 2.3.2.2.1. Translation phase; 2.3.2.2.2. Translation quality evaluation phase; 2.3.3. Results; 2.3.3.1. Lexicons usability; 2.3.3.1.1. Translation speed; 2.3.3.1.2. Use of resources; 2.3.3.1.3. Translators' impressions on the lexicons extracted from comparable corpora
2.3.3.2. Quality of the generated translations

Sommario/riassunto

Computer-assisted translation (CAT) has always used translation memories, which require the translator to have a corpus of previous translations that the CAT software can use to generate bilingual lexicons. This can be problematic when the translator does not have such a corpus, for instance, when the text belongs to an emerging field. To solve this issue, CAT research has looked into the leveraging of comparable corpora, i.e. a set of texts, in two or more languages, which deal with the same topic but are not translations of one another. This work had two primary objectives. The first is to