

DETECTION OF COMPLEX TRANSLATION EQUIVALENTS TO IMPROVE SPANISH – GERMAN/ NORWEGIAN WORD ALIGNMENT

Carla Parra Escartín

Supervisor: Prof. Koenraad de Smedt



Problem description

German: tendency to use compound nouns

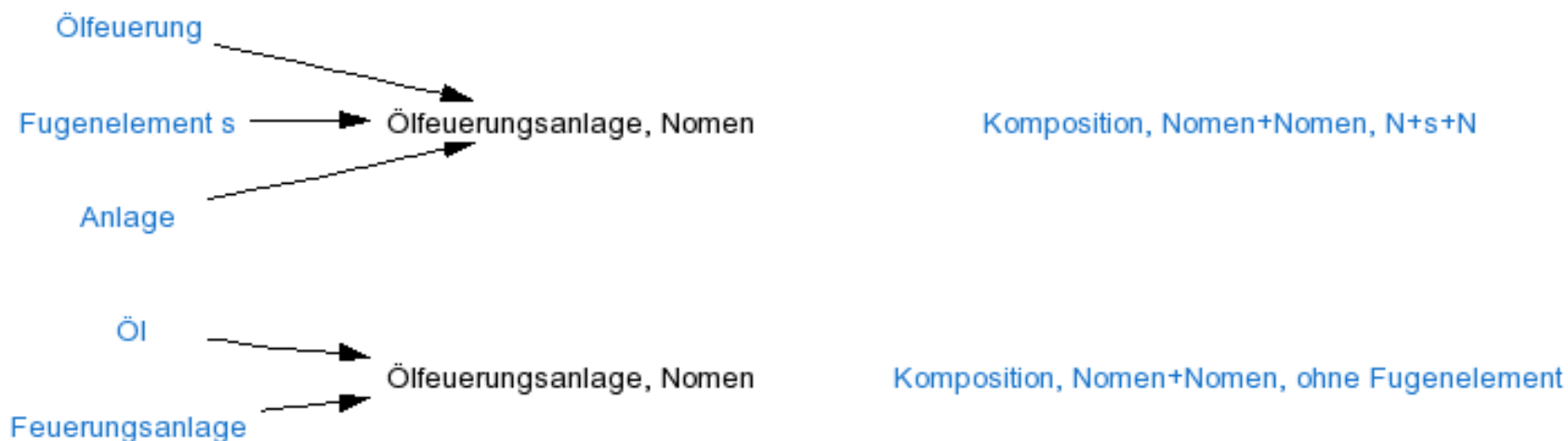
- Translated into Spanish as phraseological expressions
- Need to determine:
 - How to split the compound (DE > ES)
 - How to generate compounds (ES > DE)



Samples

Der gesuchte Begriff ist nicht im canoonet-Wörterbuch enthalten.

Ihre Eingabe lässt sich jedoch mit Hilfe unseres [Unknown Word Analyzers](#) in folgende bekannte Einträge zerlegen:

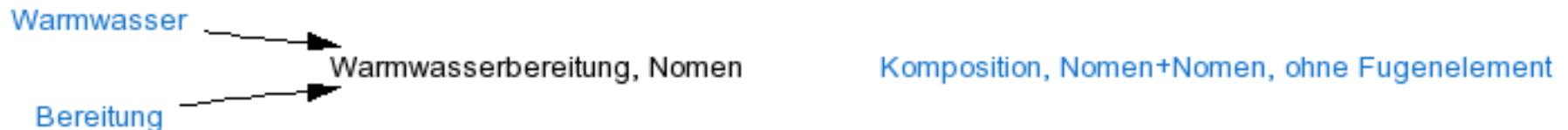


Equipos de combustión de petróleo / equipos de combustión petrolíferos
Google Translate: los hornos de petróleo

Samples

Der gesuchte Begriff ist nicht im canoonet-Wörterbuch enthalten.

Ihre Eingabe lässt sich jedoch mit Hilfe unseres [Unknown Word Analyzers](#) in folgende bekannte Einträge zerlegen:



Producción caliente de agua / Producción de agua caliente

Google Translate: agua caliente

Compounds are a challenge

- Human translators devote a lot of time to find the proper translation correspondences
 - MT Systems do not produce accurate and quality translations & translations do not sound natural
- This is a problem for bilingual dictionary extraction too:
Based on subsentential alignment



Word alignment samples

- (1) *Inverkehrbringen von Kleinfeuerungsanlagen.*
Puesta en el mercado de equipos de combustión de pequeña capacidad.
'Puesta en el mercado de equipos de combustión de pequeña capacidad'
- (2) *Anerkennung von Prüfberichten und Zulassungen.*
Reconocimiento de los informes de revisión y aprobaciones.
'Reconocimiento de los informes de revisión y aprobaciones'
- (3) Der 4. bis 6. Abschnitt sind nur auf \emptyset Feuerungsanlagen anzuwenden [...]
Los apartados 4 a 6 sólo se aplicarán... a los equipos de combustión, [...]
-

Intended impact of the project

- **Improvement of subsentential alignment**
- **Improvement of dictionary induction techniques**
- Computer Assisted Translation
- Cross-lingual Information Retrieval
- Computer Assisted Language Learning
- Other NLP applications involving more than one language



Evaluation issues (1)

1. Evaluation of the extraction of translational correspondences

- Precision in detecting ES MWE units
- Precision in detecting compounds (and splitting them?)
- Precision in the extraction of correspondences
- (Precision and acceptability of newly generated compounds?)

2. Evaluation of the MT output

- Improvement of output of MOSES and JANE using a preprocessed input text

Evaluation issues (2)

3. “Transferability” of the results in DE-ES to other language pairs (family language dependency?)

- Precision when we switch to NO-ES (other Romance languages?)
- To what extent do we need “language tuning”?