

German and MT

Pengtao Xie

Sample Text

- Die Aufsichtspraxis des BAV verlangt zudem, dass genau wie die Bestellung auch ein Widerruf in öffentlich beglaubigter Form vorgelegt werden muss. ... Gleiches sollte sicherheitshalber auch für die ABC Services erfolgen, auch wenn diese nicht der Versicherungsaufsicht unterliegt.
- Additionally, the Federal Insurance Commissioner (BAV) requires that both the granting and recanting of such authority be submitted in a notarized form. ...To be on the safe side, this procedure should also be followed for ABC Services, even though it is not subject to insurance regulation.

Why is German important?

- ~ 100 million native speakers
- The most widely spoken native language in the European Union
- One of the major languages of the world
- German is the only official language of Germany, Austria, and Liechtenstein
- One of the official languages of Switzerland, Luxembourg, and Belgium
- A recognized minor language in many other countries, such as Italy, Slovenia, Hungary, Namibia, and Poland

Why is German important?

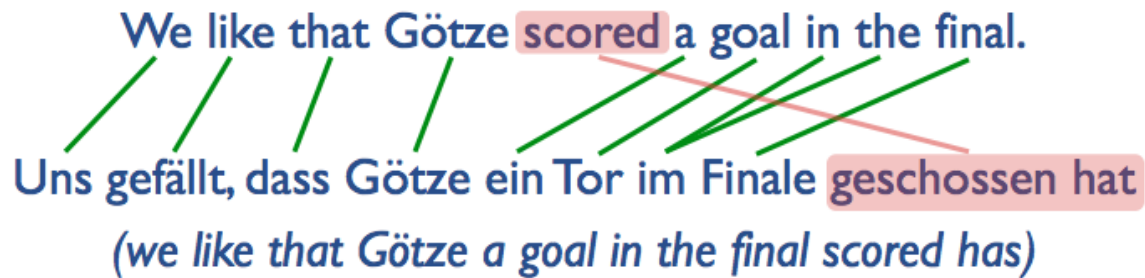
- German is the second most commonly used scientific language
- The third largest contributor to research and development
- An important language in business and culture

Why is German interesting?

- Noun inflection
 - four cases, three genders, two numbers
- Verb inflection
 - two main conjugation classes, three persons, two numbers, three moods, two voices, two tenses
- A large vocabulary:
 - 330,000 headwords in its first edition of The German Dictionary
 - 9 million words in the modern German scientific vocabulary

MT Challenges

- Word order



- Compound words
 - Generalstaatsverordnetenversammlungen (“general states representatives meetings”)
- Agreement between nouns and articles

NLP Resources and MT history

- Parallel data
 - European Parliament Proceedings Parallel Corpus 1996-2011
- NLP tools
 - Dependency parser
 - POS tagger
 - Treebank
- Supported by
 - Google translate
 - Bing translate
- History
 - Research: mid-1960s
 - Commercial: at the end of the 1980s