German
Background and history

- **Early German**
  - Dates back to as early as the 6th century
  - Started with old high German, as it split from Old Saxon

- **Modern German**
  - Started out as a purely written language
  - Became standardized in both written and spoken forms in 1901
  - Written spelling was reformed in 1996

- Most widely spoken first language in the EU
Nations where German is spoken

- Official language
- Widely spoken and understood and/or National language
- Spoken in a regional level
German Dialects

- Complex variation between dialects
- All German dialects belong to the dialect continuum of High German and Low Saxon

Low German
- Less common, not often spoken
- Maintained by some newspapers in northern Germany

High German
- Various different branches
- Yiddish is the only one that doesn’t use Latin script
Orthography

- 26 standard Latin letters
- German has three vowels with Umlaut
  - ä, ö and ü,
  - (ae, oe and ue)
- Eszett or scharfes s (sharp s), ß
Grammatical elements

- Declension of nouns and adjectives
  - nominative (Nominativ/Werfall)
  - genitive (Genitiv/Wesfall)
  - dative (Dativ/Wemfall)
  - accusative (Akkusativ/Wenfall)

- Gender
  - Masculine
  - Feminine
  - Neuter

- Numerical representations
  - Singular and plural

- Verb tenses
  - Present and present perfect
  - Simple past and past perfect
  - Future and future perfect

- Lack of auxiliary verb in negation and question

Ich kenne dich nicht.  Willst du ein Eis?
I know you not.  Want you ice-cream?
Compound words

- Long concatenation of words
- The last word that determines the word type
- The last noun of the compound word determines its gender
- Types:
  - Noun + noun *(der Mülleimer/ the garbage pail)*
  - Adjective + noun *(die Großeltern/ grandparents)*
  - Noun + adjective *(luftleer/ airless)*
  - Verb stem + noun *(die Waschmaschine/ washing machine)*
  - Preposition + noun *(der Vorort/ suburb)*
  - Preposition + verb *(runterspringen/ to jump down)*
  - Adjective + adjective *(hellblau/ light blue)*
Word order

- Normal affirmative sentences
  - The inflected verb always has position 2
- Polar questions, exclamations, and wishes
  - The inflected verb always has position 1
- Subordinate clauses
  - The verb occurs at the very end (not always in speech)
- Main verb or auxiliary verb
  - Always appear second in the sentence
    - With auxiliary, main goes to end of sentence
  - Preceded by the topic
  - The element in focus is at the end
Linguistic data/resources

- COSMAS Corpus
  - Over a billion words, online-searchable German and Austrian corpora

- NEGRA Corpus
  - Saarland University Syntactically Annotated Corpus of German Newspaper Texts

- QTAG Part of speech tagger for German

- 2nd largest number of Wikipedia pages (1355000 pages)

- Europarl German–English parallel corpus
German and MT

- **Difficulties:**
  - Different syntactic structure at the clause level
  - Rich German morphology
  - Specific problems with compound words

- **Progress:**
  - Lots of work on German due to its popularity/relation to English
  - As early as 1972 from UT
  - Many MT systems now exist for German including Google translate
Sources:
- Statistical Machine Translation of German Compound Words, Maja Popovi´c, Daniel Stein, Hermann Ney
- Towards better Machine Translation Quality for the German–English Language Pairs, Philipp Koehn Abhishek Arun Hieu Hoan
- [http://en.wikipedia.org/wiki/German_language](http://en.wikipedia.org/wiki/German_language)
- [http://en.wikipedia.org/wiki/German_grammar](http://en.wikipedia.org/wiki/German_grammar)