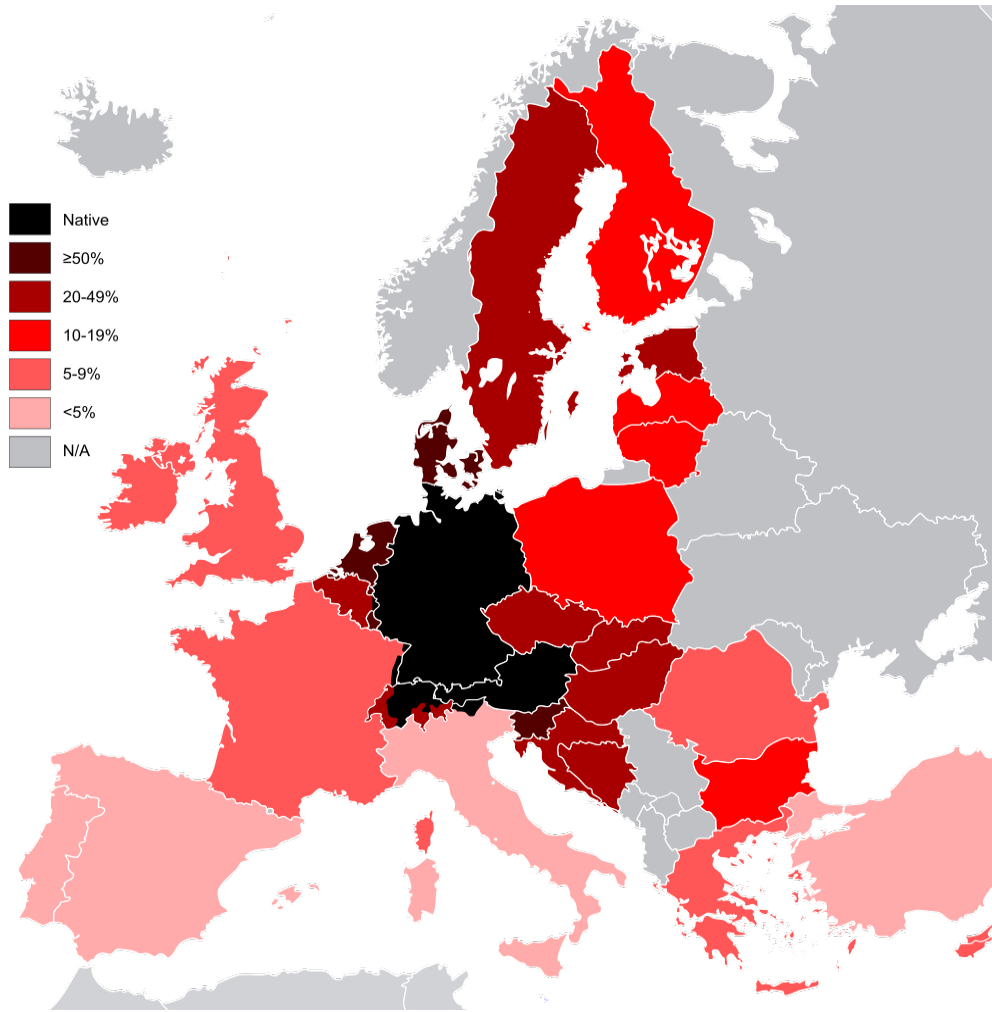


German

Yuan Cao

Demographics



- ~ 101 million native speakers
- The only official language in Germany, Austria, Liechtenstein
- One of the official languages in Switzerland, Belgium, Luxembourg
- 13th biggest language in terms of number of speakers
- 3rd biggest language for websites(5.3%)

Linguistics

- Morphology
 - 4 Cases: nominative, genitive, dative, accusative
 - 3 Genders: masculine, feminine, or neuter
 - 2 Numbers: singular and plural
- Word order
 - Both SVO and OVS are allowed
- Compound words
 - Concatenate words to form a single word
 - Almost unlimited compound words can be created

Examples

- Compound words


Haupt + die Stadt = die Hauptstadt (capital city)

*das Gedank + die Freiheit = die Gedankenfreiheit
(freedom of thought)*

- Separable verbs

*“Now he is **calling** his girlfriend”*

*“Jetzt **ruft** er seine Freundin **an**”*



Challenges for MT

- European language MT BLEU score comparison

		output language				
input language	Czech	15.1	27.1	21.0	18.1	16.2
	German	14.2	28.4			
	English	19.2	20.3	30.4	30.6	19.8
	Spanish	15.2		31.4		
	French	13.7		32.2		
	Russian	14.4		24.6		

German-English

English-French

English-German

French-English

<http://matrix.statmt.org/>

MT Resources

- Parallel corpora
 - Europarl parallel corpus (~2.2 million sentences)
 - Common crawl (~2.4 million sentences)
 - News commentary (~0.2 million sentences)
 - (check <http://www.statmt.org/wmt13/>)
- Tree banks
 - Tübingen, NEGRA, TIGER
- Morphology analyzer
 - Morphisto

Danke!