

Language Technology in Belgium

Walter Daelemans
Universiteit Antwerpen, CLiPS



- Belgium has three official languages and a number of immigration languages:
 - Dutch (60%, Vlaams)
 - French (33%, Wallon)
 - German (1%)
 - Italian, Spanish, Greek, Arabic, Turkish, ...
- “Language border” dates back to the end of the third century (fall of the West-Roman empire)
 - Roman Road (Boulogne – Köln)
- Multilingualism is essential for Belgium in tourism, trade and commerce, regional collaboration, ...

Languages in Belgium



French is the 3rd most widely-spoken first language in the EU (23%)
Dutch is the 8th most widely-spoken first language in the EU (5%)





Dialogue



Speech transcription



Content access



Multilingual dialogue







Speech transcription and **translation**



Multilingual content access








Google Walter   

Translate Turn off instant translation 

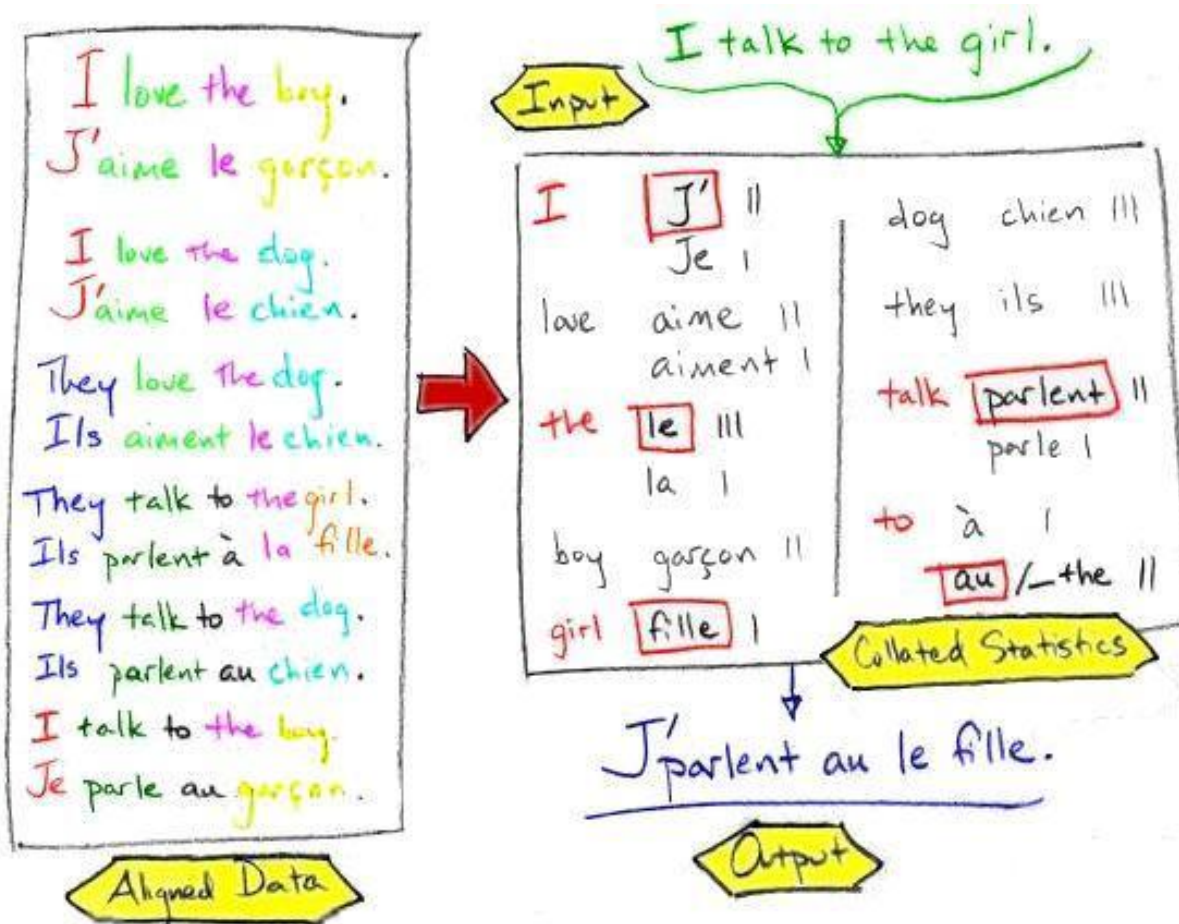
Dutch Chichewa Latin Dutch - detected ↕ English Swahili Latin ▼ Translate

Experts van de Europese Commissie, taaltechnologen, aanbieders van taaldiensten en nationale overheidsdiensten zullen zo samen **in kaart** brengen wat u verwacht van machinevertaling. ✕

Experts from the European Commission, language technologists, providers of language services and national agencies will thus bring together **map out** what you expect from machine translation.

     Suggest an edit

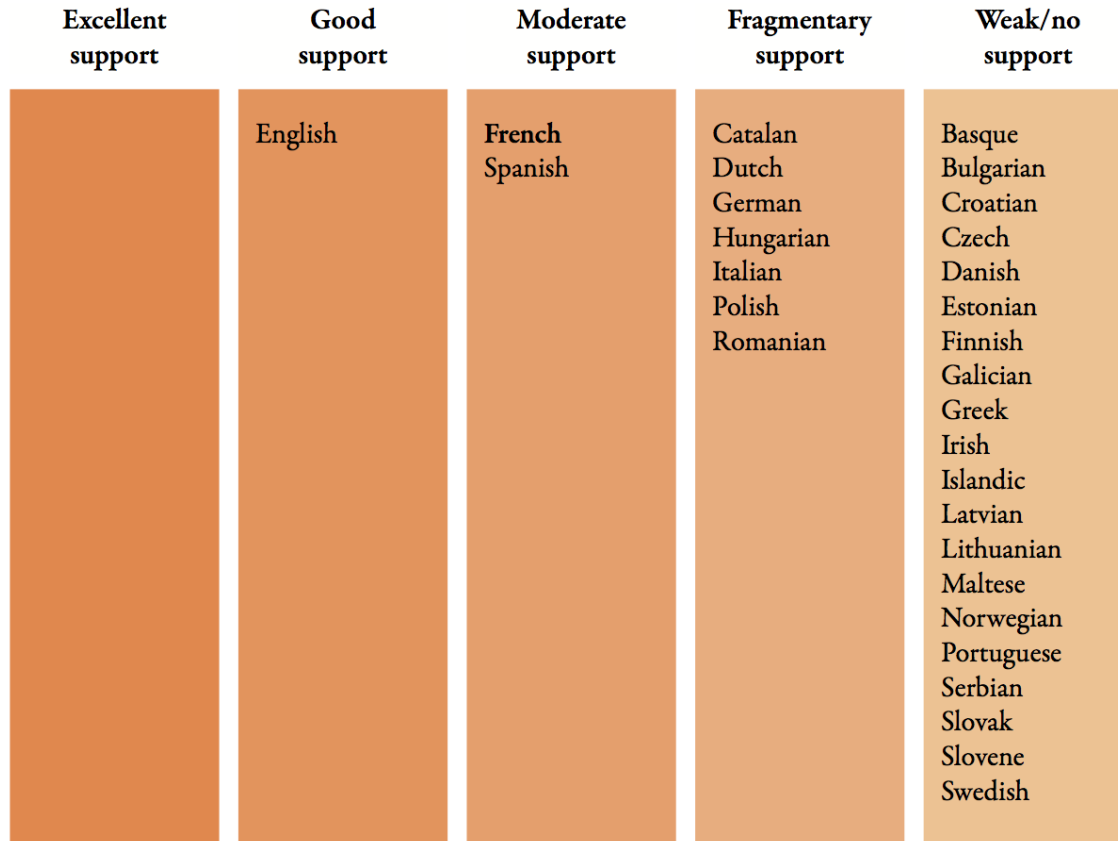
<http://specgram.com/>





- Linguistic Knowledge Based
 - Model the structure and meaning of words, sentences and texts for each language
 - Analyse - Transfer - Generate
- Statistical MT
 - Train a statistical model on large aligned corpora (often annotated with some linguistic information)
- World Knowledge and Common sense Knowledge needed
 - Reliable translation requires understanding

We are not nearly there yet!



Machine Translation: status for 30 European languages

KU Leuven	LIIR	Sien Moens	text, information retrieval
	ESAT-PSI	Patrick Wambacq Dirk Van Compernelle Hugo Van hamme	speech
	CCL	Frank Van Eynde	text
Universiteit Gent	LT ³	Véronique Hoste Lieve Macken	text
	ELIS	Jean-Pierre Martens	speech
Vrije Universiteit Brussel	ETRO	Werner Verhelst	speech
Universiteit Antwerpen	CLiPS	Walter Daelemans	text



UCL	CENTAL	Cédrick Fairon	speech, text
Mons	TCTS	Thierry Dutoit	speech

- Intensive cooperation with LT groups in The Netherlands
 - Yearly CLIN conference
 - Transnational programs such as STEVIN
 - Mobility
- Intensive cooperation between groups in Flanders
- Powerful Open Source tools for research and development
- Strong basic and application-oriented research basis (FWO, IWT SBO projects)



- Nuance (Merelbeke)
- Acapela (Mons)
- Intersystems iKnow (Diepenbeek)
- Yazzoom (Erpe-Mere)
- Textgain (Antwerp)
- Knowbel Technologies (Brussel)
- Crosslang (Gentbrugge)
- Actonomy (Zwijnaarde)
- Intelartes (Waterloo)
- Natlanco (Gent)
- Speechware (Brussel)
- Xplanation (Leuven)
- ...



- All methods for MT rely on the availability of translation data for development, domain adaptation, and testing
- Research and development competence and infrastructure are in place in Belgium
- Availability of more data would lead to better methods, better translation systems, and better resources