


UPPSALA UNIVERSITET

The Amazing Utility of Parallel Corpora

Jörg Tiedemann
Uppsala University
jorg.tiedemann@lingfil.uu.se



UPPSALA UNIVERSITET

I Came to Portugal by Plane ...

Linguee

English ↔ German ä ö ü ß

plane Search

These are translations of the English term "plane". Click here to see translations of the German term "plane".

Editorial Dictionary

- plane noun
 - Flugzeug nt (aviation)
 - Ebene f (mathematics)
 - Flieger m (aviation)
 - Fläche f Maschine f
 - Hobel m
 - Platane f (botany)
 - Niveau nt Stufe f
- z-plane
 - z-Ebene
- plane adjective
 - plan adj flach adj
 - eben adj
 - zweidimensional adj
- plane verb
 - hobeln vt gleiten vi
 - segeln v planieren vt
 - glätten vt abrichten v

Examples:

- board a plane verb
- ein Flugzeug besteigen v

Detailed information on this dictionary entry

plane noun, aviation

Flugzeug noun, neuter
also translated as: aircraft, airplane, aeroplane

Examples from the Linguee Dictionary:

in time to catch the **plane** rechtzeitig für das Flugzeug

Examples from other sources

[...] obliging companies to lower prices in a bid to keep their customers and, on the other, making potential clients more reluctant to travel long distances **by plane**.

[...] Interesse der Kundenbindung zu Preissenkungen gezwungen waren und zum anderen die potenziellen Kunden weniger geneigt waren, Fernreisen mit dem **Flugzeug** anzutreten.

[...] another man that they are shooting at 1000 metres away, or drop bombs from **a plane** on innocent women and children, that kind of man will not sleep at night, will [...]

[...] blicken, den sie aus 1000 Meter Entfernung erschließen, oder Bomben aus einem **Flugzeug** auf unschuldige Frauen und Kinder werfen, diese Art Mann wird nachts nicht schlafen [...]

[...] you can enter the number (1 to 20), in the fields to the right of it you can enter the name, **plane** type and skill (1=very good to 10=very bad) of each competitor.

[...] Feldern rechts daneben können Sie den einzelnen Konkurrenten Namen geben, ihnen **ein Flugzeug** zuweisen und ihre Geschicklichkeit (1=schwach bis 10=gut)

UPPSALA UNIVERSITET

... but I Didn't Use a Plane

Linguee

English ↔ German ä ö ü ß

plane Search

These are translations of the English term "plane". Click here to see translations of the German term "plane".

Editorial Dictionary

- plane noun
 - Flugzeug nt (aviation)
 - Ebene f (mathematics)
 - Flieger m (aviation)
 - Fläche f Maschine f
 - Hobel m
 - Platane f (botany)
 - Niveau nt Stufe f
- z-plane
 - z-Ebene
- plane adjective
 - plan adj flach adj
 - eben adj
 - zweidimensional adj
- plane verb
 - hobeln vt gleiten vi
 - segeln v planieren vt
 - glätten vt abrichten v

Examples:

- board a plane verb
- ein Flugzeug besteigen v

Detailed information on this dictionary entry

plane noun

Hobel noun, masculine
also translated as: sicer, carpenter's plane

Examples from the Linguee Dictionary:

plane hobeln

carpenter's **plane** Hobel

Examples from other sources

[...] verses of arranging words / and early the pen / and the brush / and then the **plane** hardly understandably hardly durable hardly time / in the eastern the windy dragongreen [...]


[...] gestaltender worte / und früh den stift / und später der pinsel / und dann **der hobel** kaum fassbar kaum haltbar kaum zeit / im östlichen frühlingmorgen das windige drachengrün [...]

One man can take a sharp saw, a square, a good **plane**, and so on, and build a handsome article of furniture; another man can take the same tools and set to work [...]

Ein Mann kann eine scharfe Säge, ein Winkelmaß, einen guten **Hobel** usw. haben und ein attraktives Möbelstück herstellen; ein anderer Mann kann die gleichen Werkzeuge [...]

Examples:

- Plane the workpiece into size before 1. the
- Hobeln Sie das Werkstück vor der 1.



UPPSALA UNIVERSITET

Multi-Lingual Technology on the Web

Translate

From: English To: Swedish Translate

Finnish English Swedish Detect language

attribute

↑ language identification

attribut

speech synthesis

↓

part of speech

- noun
 - egenskap property, attribute, capacity, trait, qualification, quality
 - attribut attribute
 - bestämning determination, qualifier, modifier, adjunct, attribute
 - kännetecken feature, characteristic, attribute, badge, token, earmark
- verb
 - tillskriva ascribe, attribute, accredit, assign to, set down, impute
 - attribuera attribute

↑ synonyms



Google Translate Constantly Grows ...

What languages does Google Translate support?

Google Translate currently supports:

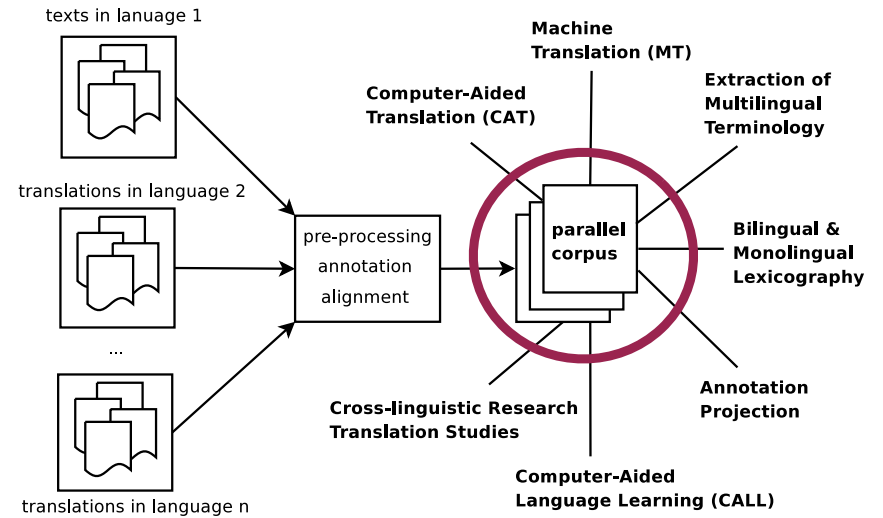
Afrikaans	Dutch	Hindi	Malay	Spanish
Albanian	English	Hungarian	Maltese	Swahili
Arabic	Esperanto	Icelandic	Norwegian	Swedish
Belarusian	Estonian	Indonesian	Persian	Thai
Bosnian	Filipino	Irish	Polish	Turkish
Bulgarian	Finnish	Italian	Portuguese	Ukrainian
Catalan	French	Japanese	Romanian	Vietnamese
Chinese	Galician	Korean	Russian	Welsh
Croatian	German	Latvian	Serbian	Yiddish
Czech	Greek	Lithuanian	Slovak	
Danish	Hebrew	Macedonian	Slovenian	

Current alpha languages are:

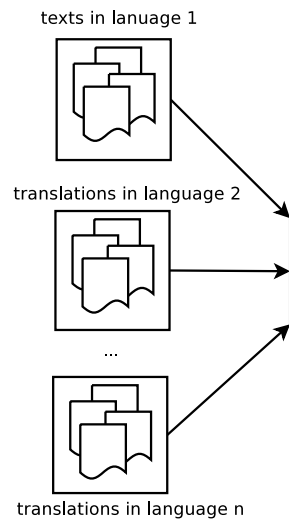
Armenian	Cebuano	Hmong	Lao	Telugu
Azerbaijani	Georgian	Javanese	Latin	Urdu
Basque	Gujarati	Kannada	Marathi	
Bengali	Haitian Creole	Khmer	Tamil	



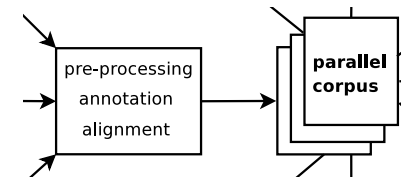
How Is This Possible?



Collect ...

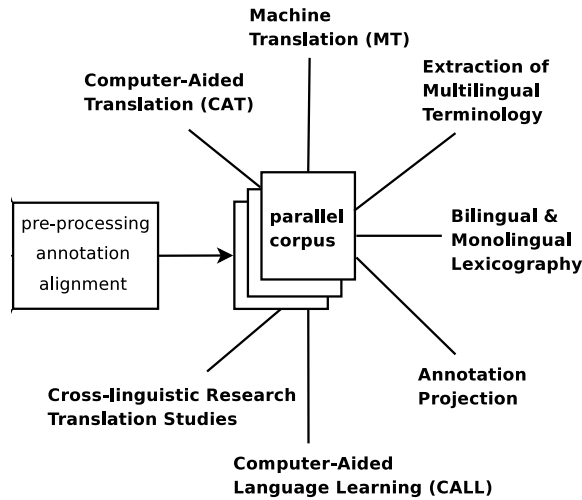


Process ...

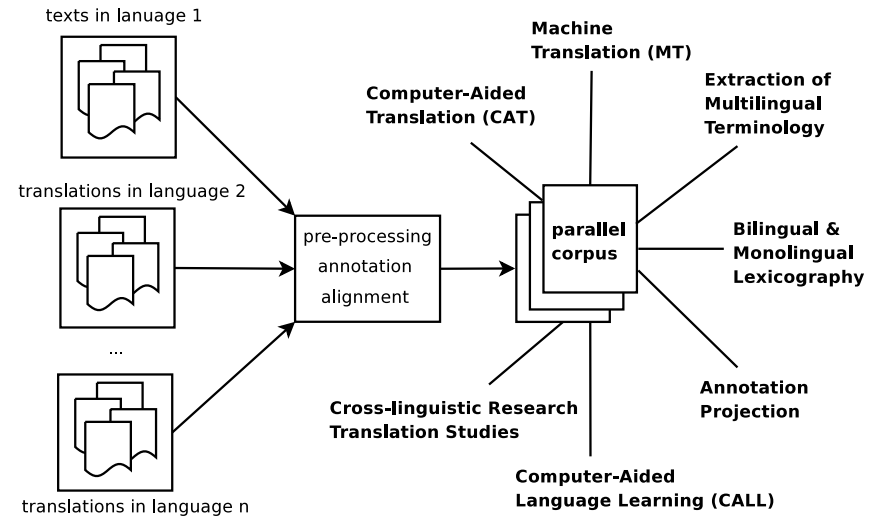




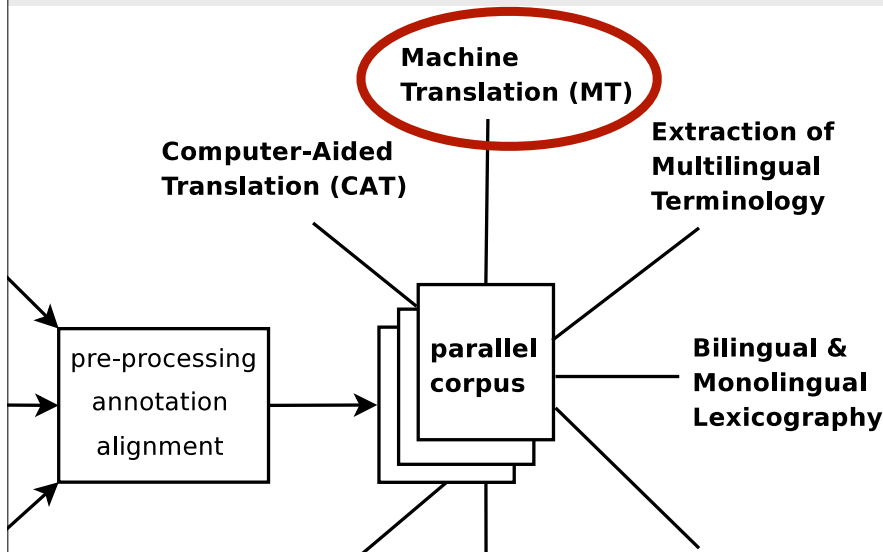
Apply ...



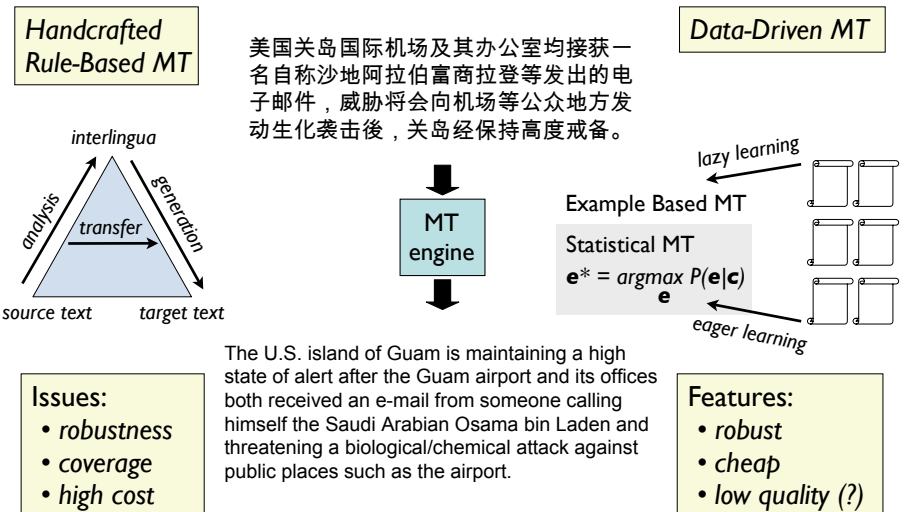
The Amazing Utility of Parallel Corpora



The Amazing Utility of Parallel Corpora



Paradigms in Machine Translation





UPPSALA
UNIVERSITET

Inside Google Translate

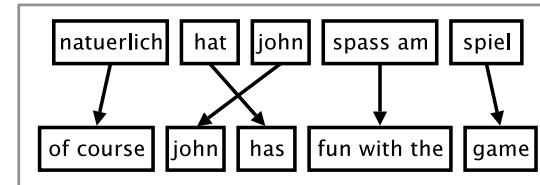


UPPSALA
UNIVERSITET

Phrase-Based SMT in a Nutshell

Training:

millions of
aligned
parallel
sentences



Translation
Model

Reordering
Model

billions of
target
language
sentences

of course john has
course john has fun
john has fun with
.....

Target
Language
Model

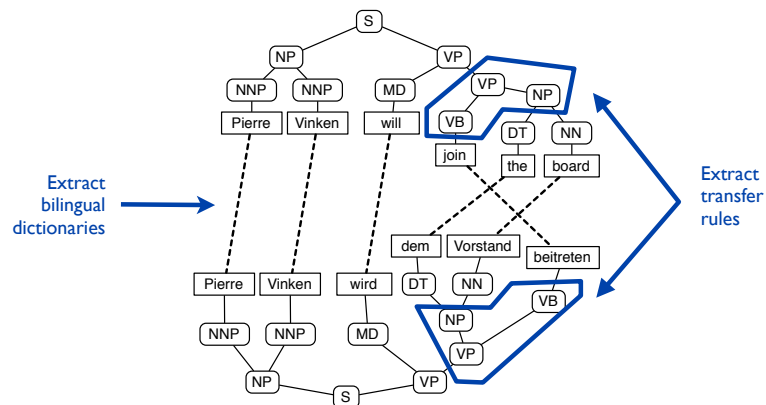
translation = decoding: $e^* = \underset{e}{\operatorname{argmax}} P(e|d)$



UPPSALA
UNIVERSITET

Learning Transfer-Based MT

Possibility: Induce transfer rules from aligned data



UPPSALA
UNIVERSITET

Summary: Learning to Translate

Induce translation knowledge from data

- existing translations
- existing language resources

Learning Transfer Systems

- data-driven parsing (analysis)
- induced transfer rules and dictionaries
- language modeling for generation

Global Data-Driven MT

- example-based MT
- statistical MT (phrase-based, syntax-based)

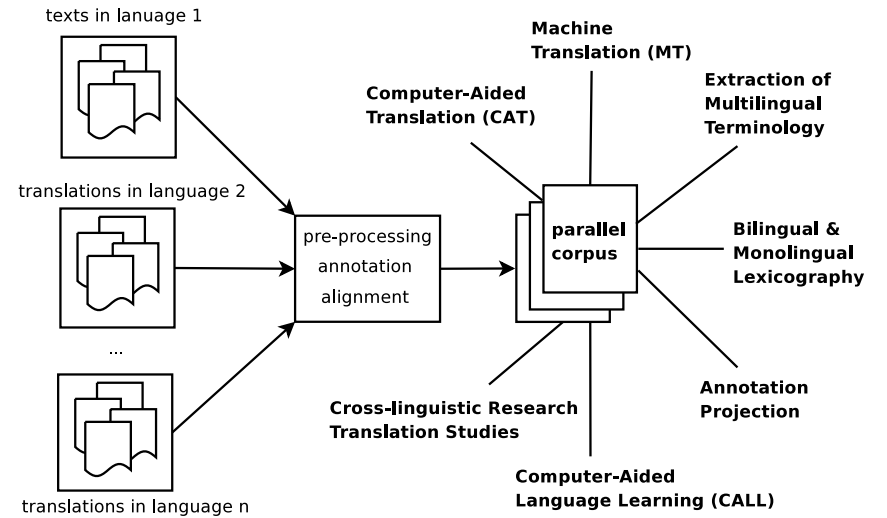


The Amazing Utility of Parallel Corpora

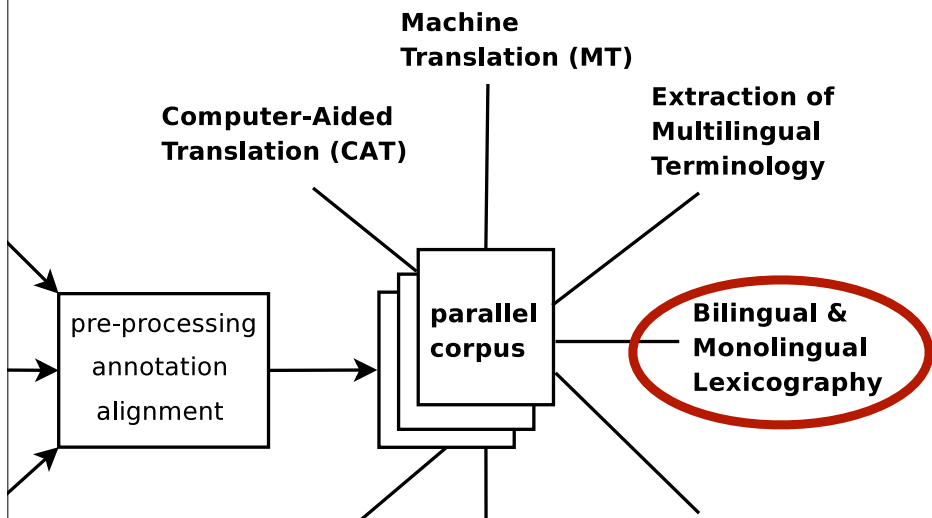
Part II - More Than Machine Translation?



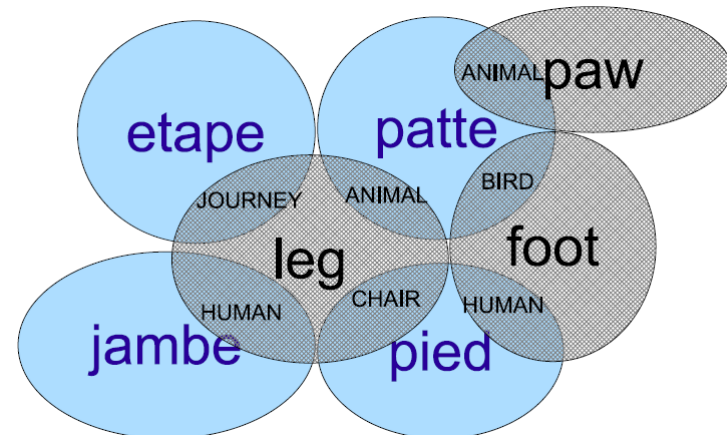
The Amazing Utility of Parallel Corpora



The Amazing Utility of Parallel Corpora



Lexical Ambiguities Across Languages





Translations as “Semantic Mirrors”

Honey/Darling:

- *English:* To some land flowing with milk and **honey!**
- *German:* In ein Land, in dem Milch und **Honig** fließen!
- *Swedish:* Till ett land fullt av mjölk och **honung.**
- *English:* What are you waiting for, **honey?**
- *German:* Worauf wartest du, **Schatz?**
- *Swedish:* Vad väntar du på, **älskling?**

“Back-translations” (uncover word senses & relations)

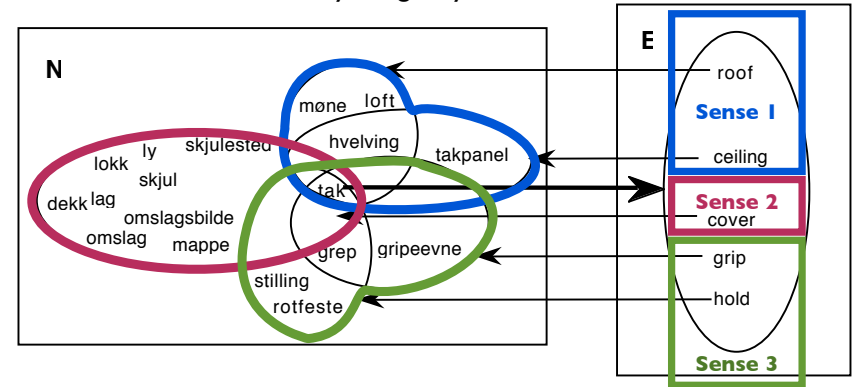
- *Swedish:* Le, **älskling**, le.
- *English:* Smile, **darling**, smile.



Translations as “Semantic Mirrors”

Discover Word Senses from Word Aligned Data

- Work by Helge Dyvik et al:



Synonym Extraction & Paraphrasing

- Hypothesis of distributional similarity (Harris, 1954)
- “*You shall know a word by the company it keeps.*” (Firth, 1957)
- Use aligned translations as context!
- Find lexical items that share translations

Dutch:	lente	voorjaar	jaar	najaar	...
EN: spring	396	1753	0	0	...
EN: early	34	28	0	0	...
EN: year	0	64	67510	0	...
EN: autumn	0	0	0	1296	...
DE: Frühjahr	261	1447	5	0	...
DE: Frühjahrstagung	0	140	0	0	...
SV: vår	42	238	3	0	...
...

→ “lente” & “voorjaar” = synonyms (but not the others)



What is this good for?

Lexical resources

- WordNets
- Ontologies

Query expansion in information retrieval

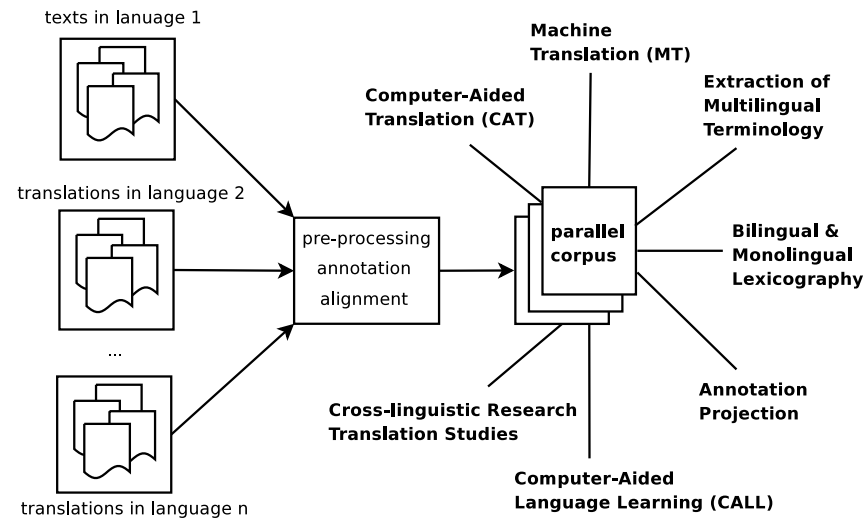
Paraphrasing for machine translation

Computer-Aided Language Learning (CALL)

Cross-linguistic studies



Conclusions: Use Parallel Corpora!



Some Resources (there is much more!)

Parallel corpora

- <http://www.statmt.org/europarl/>
- <http://opus.lingfil.uu.se>

Pre-processing and alignment tools

- <http://traduccionmundolibre.com/wiki>
- <http://opus.lingfil.uu.se/letsmt-trac/>

Search Interfaces

- <http://www.linguee.com>
- <http://bab.la>
- <http://mymemory.translated.net>



Questions?



What is next?

How to build parallel corpora

How to align sentences

How to align words