

CS 11-737: Multilingual NLP

Translation, Translation Data, and Evaluation

Graham Neubig

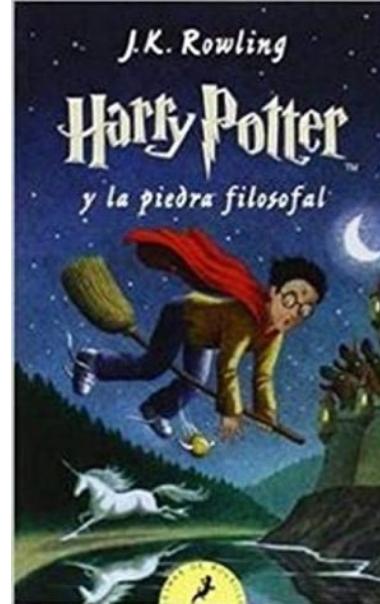
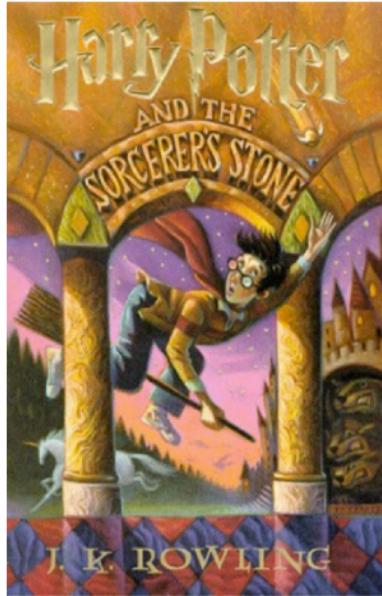
(many slides from Yulia Tsvetkov)



Carnegie Mellon University
Language Technologies Institute



Translation



Mr. and Mrs. Dursley, who lived at number 4 on Privet Drive, were proud to say they were very normal, fortunately.

El señor y la señora Dursley, que vivían en el número 4 de Privet Drive, estaban orgullosos de decir que eran muy normales, afortunadamente.

Plan

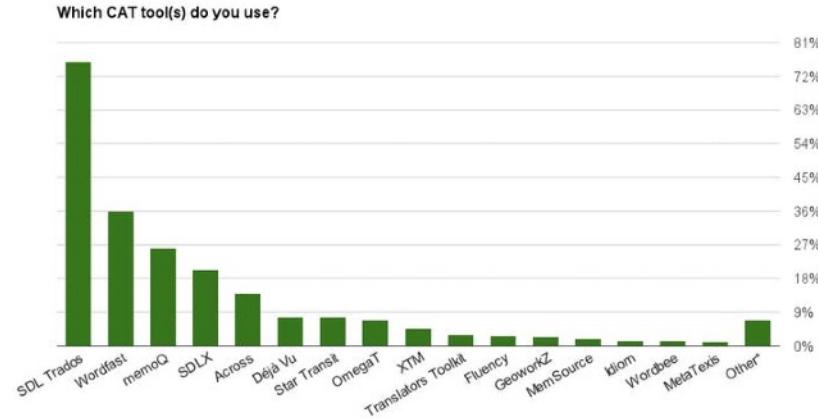
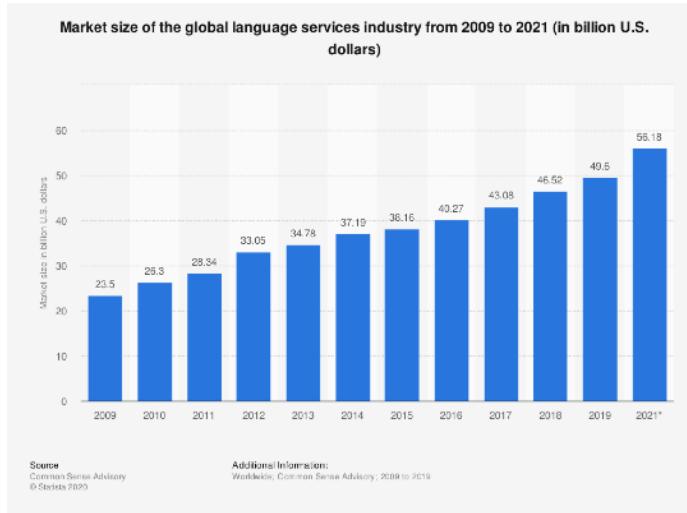
- The practice of translation
- Machine translation (MT)
- MT data sources
- MT evaluation

Translation vs. Interpretation

- **Translation**
 - Conversion of the written word from one language to another
 - Generally less time constrained
 - Translators tend to specialize in one area, work alone
 - High accuracy and fluent output highly valued
- **Interpretation**
 - Conversion of spoken language from one language to another
 - Simultaneous and consecutive interpretation
 - Interpreters tend to be generalists, work in teams
 - Complete accuracy often infeasible, conveying intent most important

How Much Translation is Done?

- Language services market w/ 56.1 billion (translation, interpretation, MT)
- 640,000 translators worldwide (about 75% freelance)
- Europe owns 49% of market share
- 88% of full-time translators use computer aided translation



Difficulty in Translation

Why is it difficult to translate?

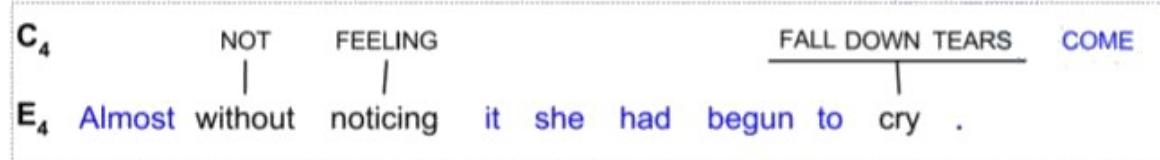
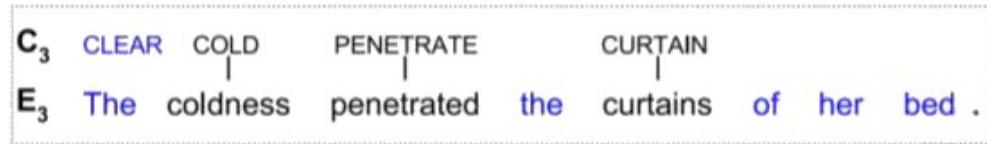
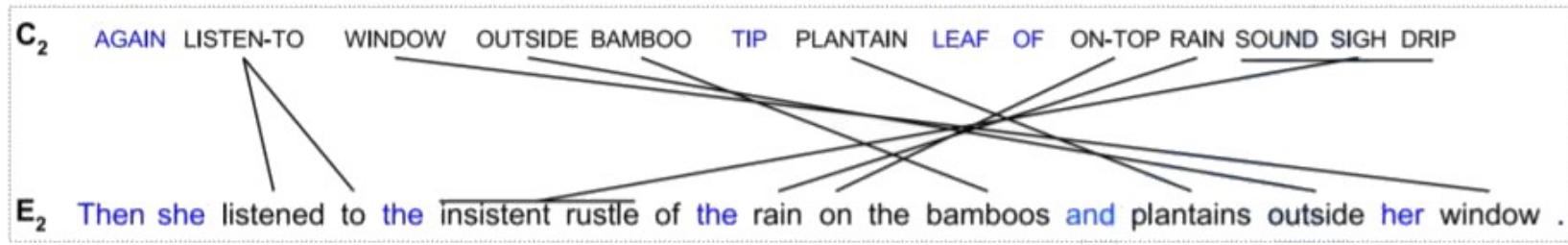
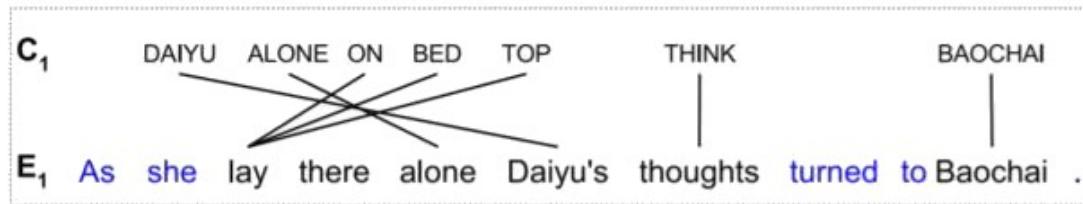
- Cross-lingual lexical and structural divergences

錨玉自在枕上感念寶釵。。。又聽見窗外竹梢焦葉之上，
雨聲漸沂，清寒透幕，不覺又滴下淚來。

dai yu zi zai zhen shang gan nian bao chai...you ting jian chuang wal zhu shao xiang ye
zhe shang, yu sheng xili, qing han tou mu, bu jue you di xia lei lat

From “Dream of the Red Chamber” Cao Xue Qin (1792)

Why is it difficult to translate?

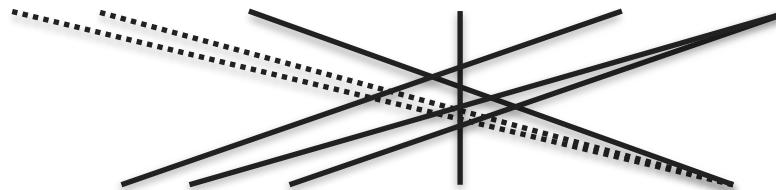


Why is it difficult to translate?

- Structural divergences
 - Morphology
 - Syntax

in the in-city exploded a car-bomb

German: In der Innenstadt explodierte eine Autobombe

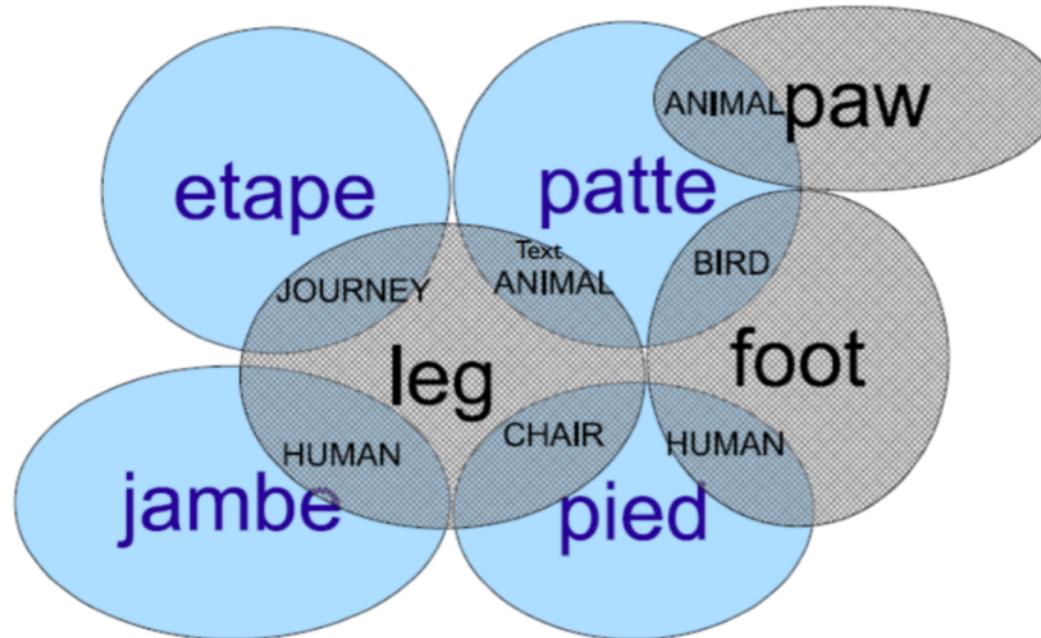


English: A car bomb exploded downtown.

Translationese: In the inner city, there exploded a car bomb.

Why is it difficult to translate?

- Lexical ambiguities and divergences across languages



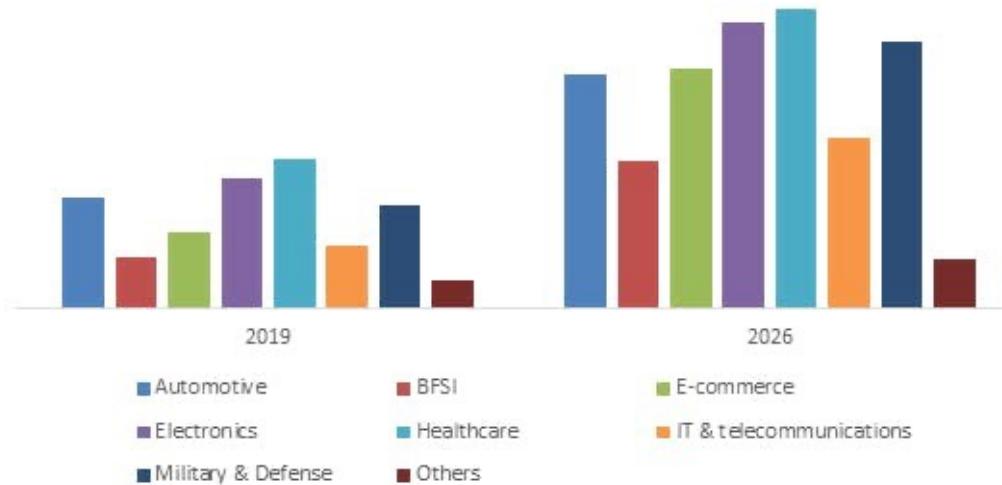
[Examples from Jurafsky & Martin Speech and Language Processing 2nd ed.]

Machine Translation

Machine Translation

- Machine translation is a \$3 billion market
- Top providers: Google, Amazon, DeepL
- Spread across many industries

U.S. Machine Translation Market Size, By Application, 2019 & 2026



Source: www.gminsights.com

Text

Documents

DETECT LANGUAGE

RUSSIAN

ENGLISH

SPANISH



ENGLISH

RUSSIAN

SPANISH



Search languages

Afrikaans	Czech	Hebrew	Latin	Portuguese	Tajik
Albanian	Danish	Hindi	Latvian	Punjabi	Tamil
Amharic	Dutch	Hmong	Lithuanian	Romanian	Telugu
Arabic	<input checked="" type="checkbox"/> English	Hungarian	Luxembourgish	Russian	Thai
Armenian	Esperanto	Icelandic	Macedonian	Samoan	Turkish
Azerbaijani	Estonian	Igbo	Malagasy	Scots Gaelic	Ukrainian
Basque	Filipino	Indonesian	Malay	Serbian	Urdu
Belarusian	Finnish	Irish	Malayalam	Sesotho	Uzbek
Bengali	French	Italian	Maltese	Shona	Vietnamese
Bosnian	Frisian	Japanese	Maori	Sindhi	Welsh
Bulgarian	Galician	Javanese	Marathi	Sinhala	Xhosa
Catalan	Georgian	Kannada	Mongolian	Slovak	Yiddish
Cebuano	German	Kazakh	Myanmar (Burmese)	Slovenian	Yoruba
Chichewa	Greek	Khmer	Nepali	Somali	Zulu

[Explore our content ▾](#)[Journal information ▾](#)

nature > nature communications > articles > article

Article | [Open Access](#) | Published: 01 September 2020

Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals

Martin Popel , Marketa Tomkova, Jakub Tomek, Łukasz Kaiser, Jakob Uszkoreit, Ondřej Bojar & Zdeněk Žabokrtský

Nature Communications **11**, Article number: 4381 (2020) | [Cite this article](#)

翻訳とは、言語において、Aという形で表現されているものを、その意味に対応するBという形に置き換える行為をさす。具体的には、自然言語において、起点言語による文章を、別の目標言語による文章に変換する行為をさす。



a different form

an expression that takes

Translation refers to the act of replacing what is expressed in the form of A in a language with the form of B that corresponds to that meaning. Specifically, in natural language, it refers to the act of converting a sentence in the source language into a sentence in another target language.

Lexical Ambiguity in Translation (e.g. Google Translate)

コード

ko-do

(code, cord, chord)

電気のコードが切れた。

The electric cord has broken.

プログラムのコードを書いた。

I wrote the code for the program.

楽譜にコードを書いた。

I wrote the chord on the score.

プログラマーなのでコードを読むのが得意。

As a programmer, I am good at reading code.

ミュージシャンなのでコードを読むのが得意。

As a musician, I am good at reading codes.

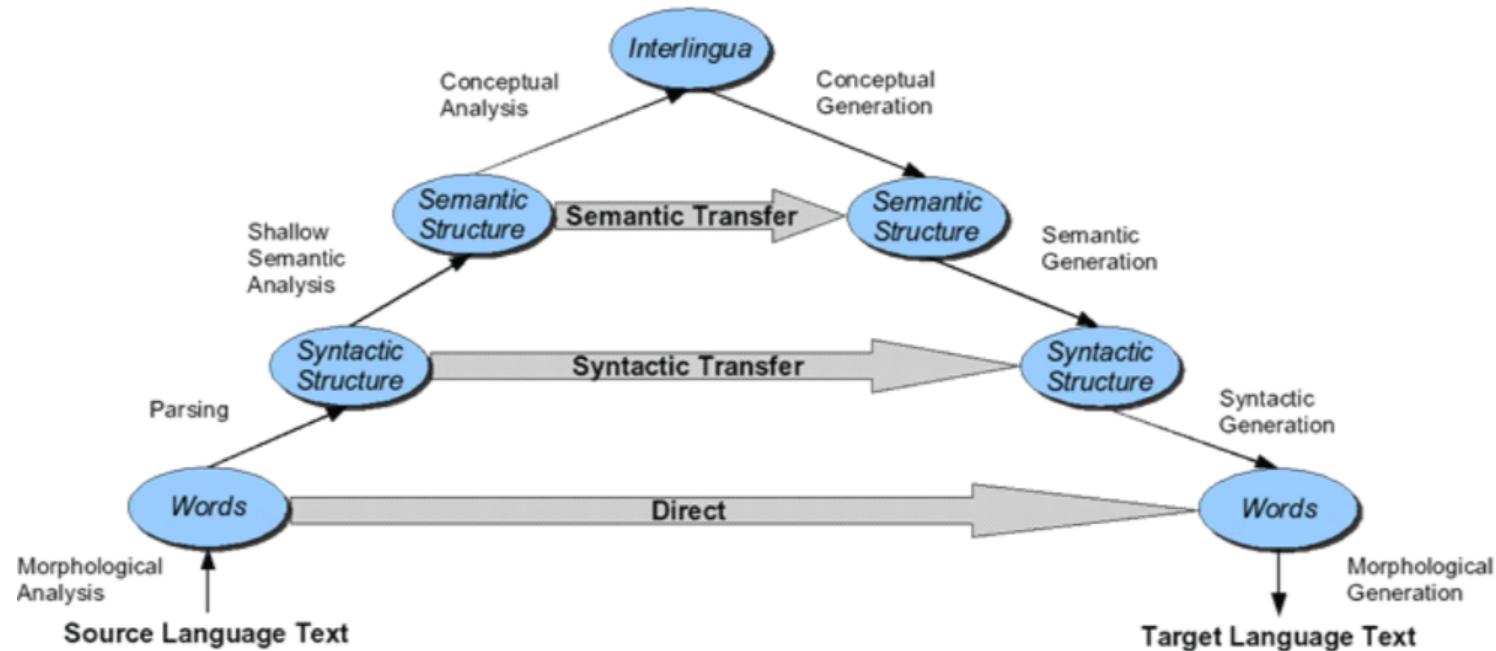
Javaで、ギターのコードを表示するコードを書いた。

In Java, I wrote a chord that displays the chord of the guitar.

3 Classical methods for MT

- Direct
- Transfer
- Interlingua

The Vauquois triangle (1968)



Transfer approaches

- Direct transfer (word-by-word translation)

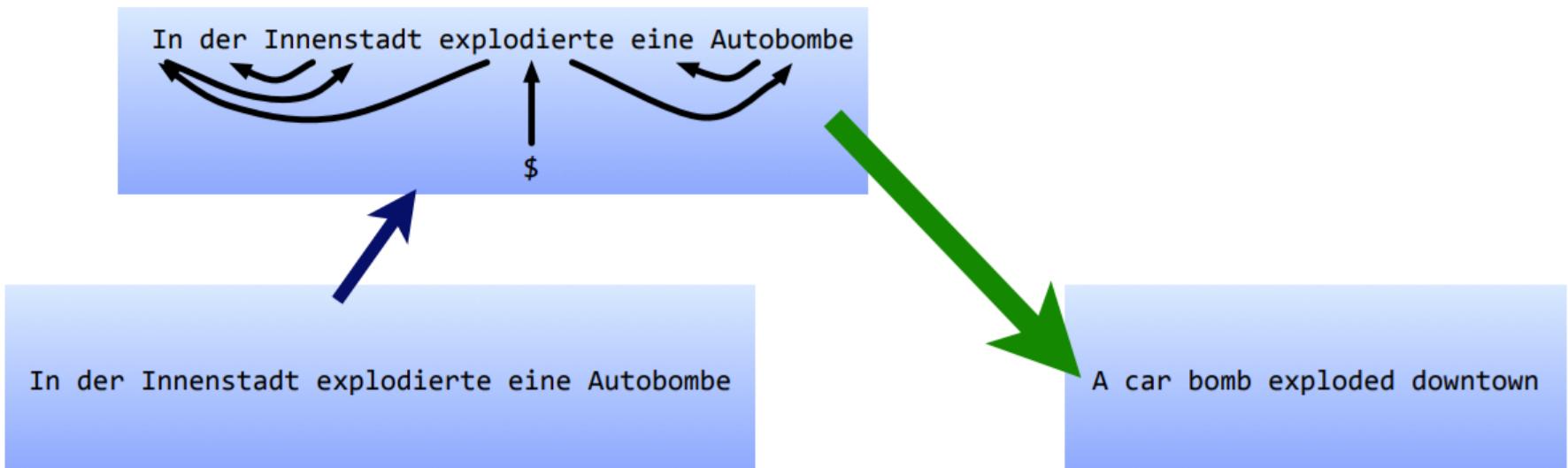
In der Innenstadt explodierte eine Autobombe



A car bomb exploded downtown

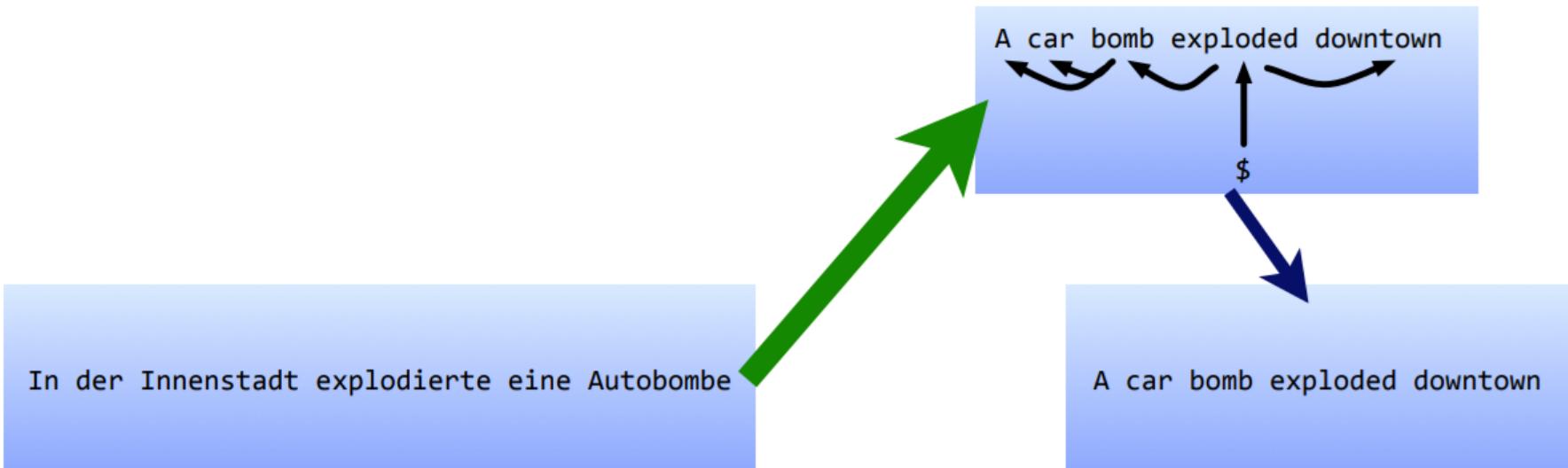
Transfer approaches

- Syntactic transfer



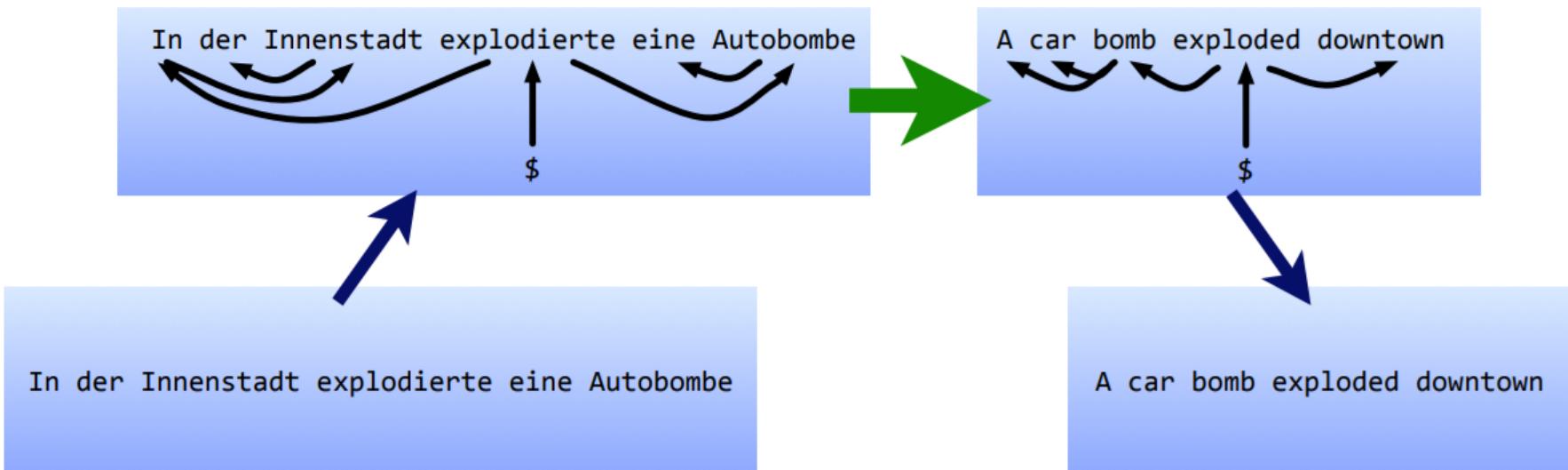
Transfer approaches

- Syntactic transfer



Transfer approaches

- Syntactic transfer



Transfer approaches

- Semantic transfer

Semantics
“logical form”

In der Innenstadt explodierte eine Autobombe

```
detonate
:arg0 bomb
:arg1 car
:loc downtown
:time past
```



Transfer approaches

- Semantic transfer

Semantics
“logical form”

Syntax

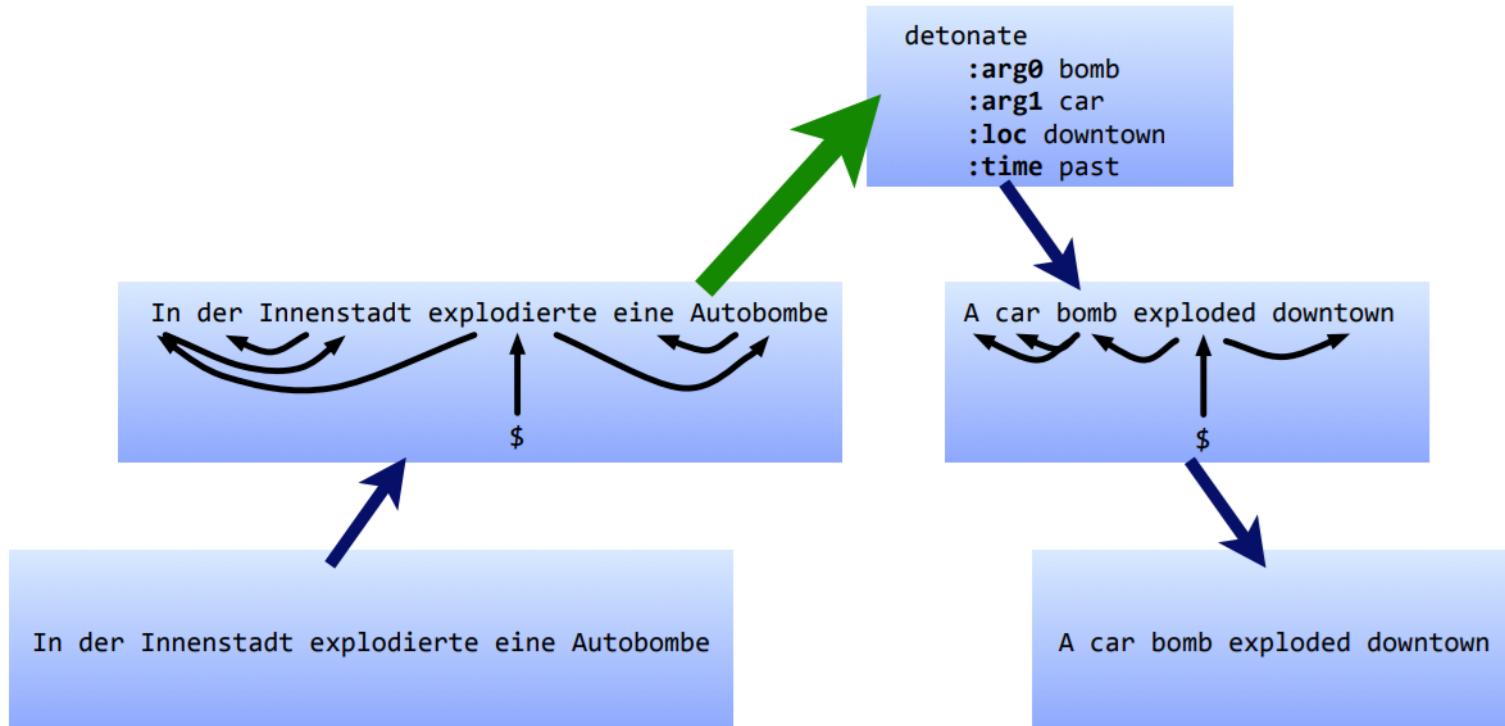
In der Innenstadt explodierte eine Autobombe

detonate
:arg0 bomb
:arg1 car
:loc downtown
:time past

A car bomb exploded downtown

A car bomb exploded downtown

Transfer approaches



Interlingua

Interlingua
“meaning”

```
report_event[  
    factivity=true  
    explode(e, bomb, car)  
    loc(e, downtown)  
]
```

```
explodieren  
:arg0 Bombe  
:arg1 Auto  
:loc Innenstadt  
:tempus imperf
```

```
detonate  
:arg0 bomb  
:arg1 car  
:loc downtown  
:time past
```

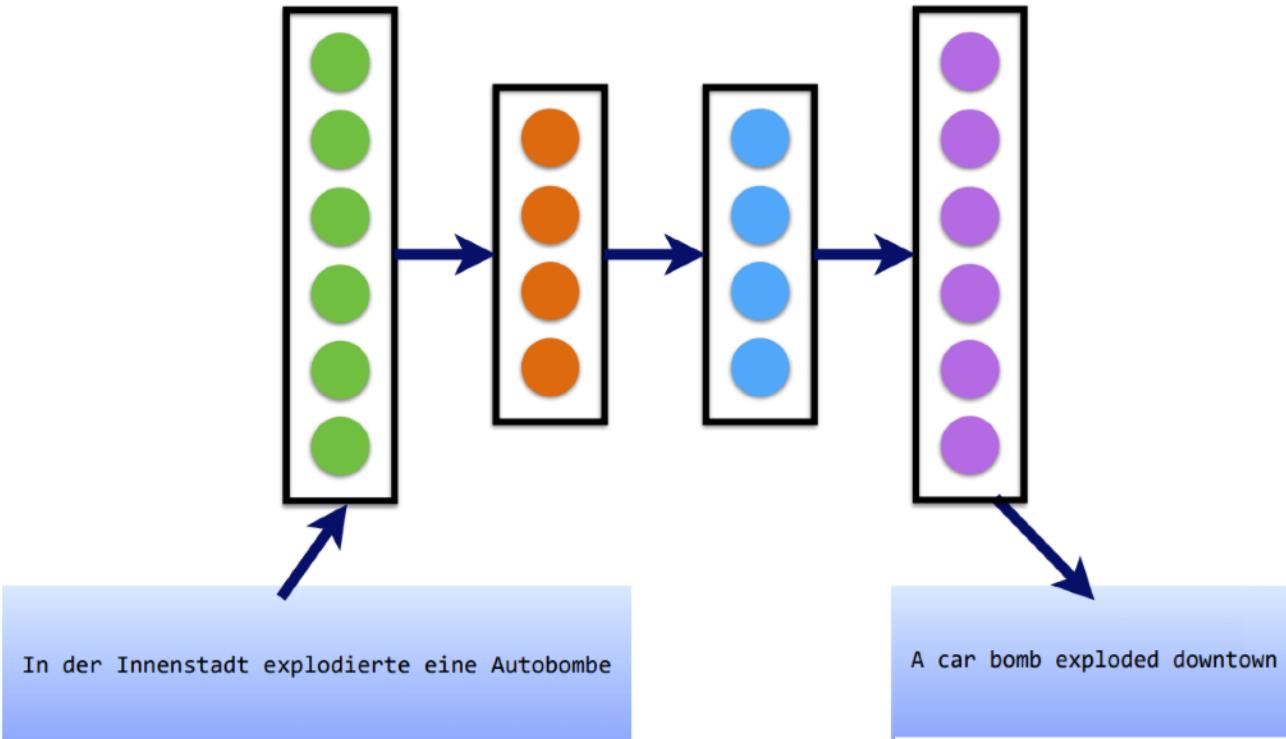
In der Innenstadt explodierte eine Autobombe

A car bomb exploded downtown

In der Innenstadt explodierte eine Autobombe

A car bomb exploded downtown

Interlingua?



MT and Data

Learning from data

1a. ok-voon ororok sprok .

1b. at-voon bichat dat .

2a. ok-drubel ok-voon anok plok sprok

2b. at-drubel at-voon pippat rrat dat .

3a. erok sprok izok hihok ghirok .

3b. totat dat arrat vat hilat .

4a. ok-voon anok drok brok jok .

4b. at-voon krat pippat sat lat .

5a. wiwok farok izok stok .

5b. totat jjat quat cat .

6a. lalok sprok izok jok stok .

6b. wat dat krat quat cat .

7a. lalok farok ororok lalok sprok izok enemok .

7b. wat jjat bichat wat dat vat eneat .

8a. lalok brok anok plok nok .

8b. iat lat pippat rrat nnat .

9a. wiwok nok izok kantok ok-yurp .

9b. totat nnat quat oloat at-yurp .

10a. lalok mok nok yorok ghirok clok .

10b. wat nnat gat mat bat hilat .

11a. lalok nok crrrok hihok yorok zanzanok .

11b. wat nnat arrat mat zanzanat .

12a. lalok rarok nok izok hihok mok .

12b. wat nnat forat arrat vat gat .

Translation challenge: **farok crrrok hihok yorok clok kantok ok-yurp**

(from Knight (1997): Automating Knowledge Acquisition for Machine Translation)

Parallel corpora



Search A-Z Site Index

About Us » Our Work » Events and News Get Involved Coronavirus (COVID-19)



SPORTS

Olympic Truce: to build 'culture of peace' through sport

As the Beijing 2022 Olympic Winter Games will officially open on 4



Naciones
Unidas

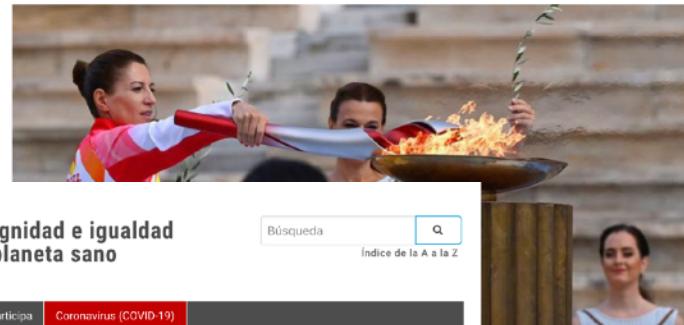
La Organización » Qué hacemos » Eventos y noticias Participa Coronavirus (COVID-19)



健康地球上的和平、尊严与平等

搜索 网址索引

联合国概览 » 行动使命 » 活动与新闻 参与进来 2019冠状病毒病 (COVID-19)



体育

奥林匹克休战：通过体育建立“和平文化”

北京2022年冬季奥林匹克运动会将于2022年2月4日正式开幕，秘书长安东尼奥·古特雷斯敦促世界通过体育的力量“建立和平文化”，并呼吁各国遵守上周通过联合国大会决议批准的奥林匹克休战宣言。

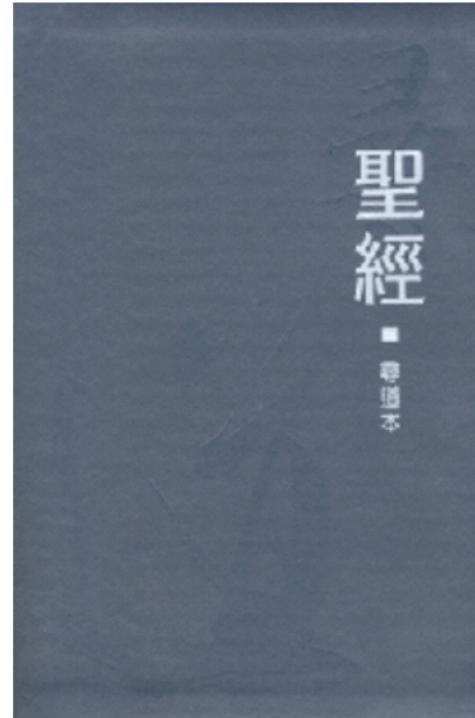
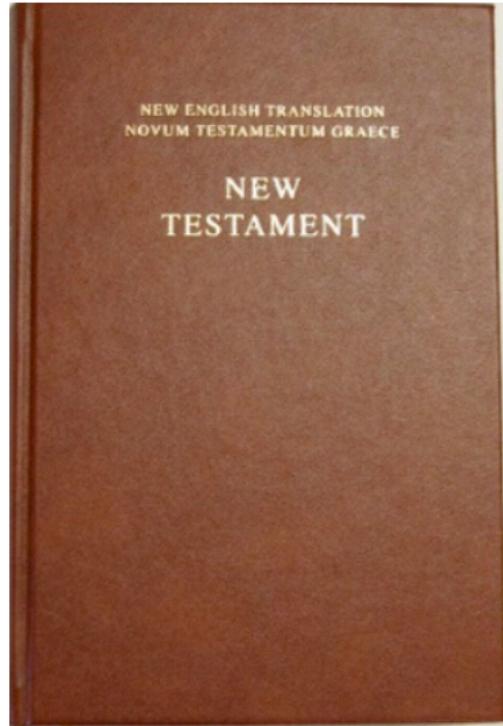


DEPORTES

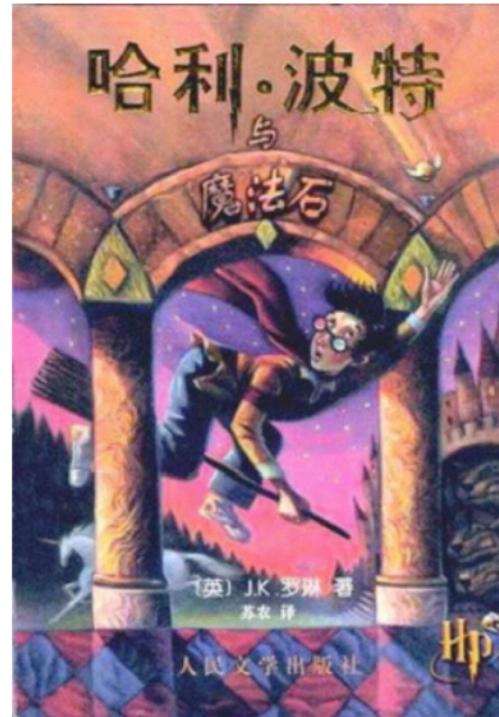
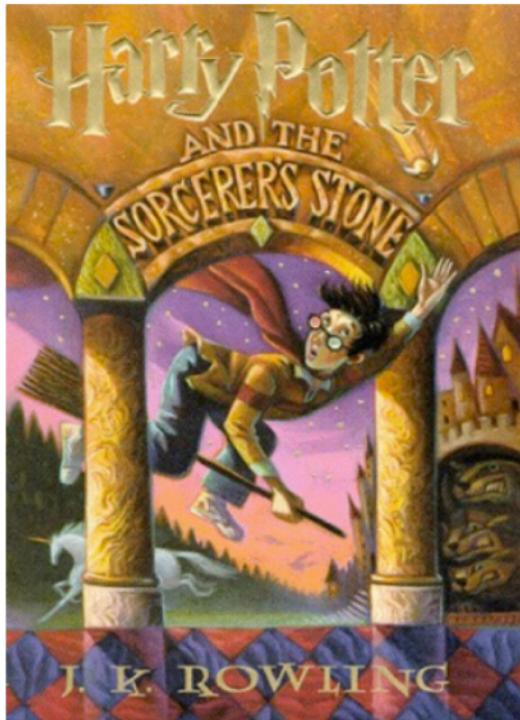
Tregua Olímpica: construir una "cultura de paz" a través del deporte

Con motivo de la inauguración de los Juegos Olímpicos de Invierno de Beijing en el 4 de febrero, el secretario general de la ONU, Antonio Guterres, insta al mundo a "construir una cultura de paz" a través del poder del deporte y ha pedido a las naciones que observen la Tregua Olímpica.

Parallel corpora



Parallel corpora



Parallel corpora

CLASSIC SOUPS

		Sm.	Lg.
清 燉 雞 湯	57.	House Chicken Soup (Chicken, Celery, Potato, Onion, Carrot)	1.50 2.75
雞 飯 湯	58.	Chicken Rice Soup	1.85 3.25
雞 麵 湯	59.	Chicken Noodle Soup	1.85 3.25
廣 東 雲 吞	60.	Cantonese Wonton Soup.....	1.50 2.75
蕃 茄 雞 湯	61.	Tomato Clear Egg Drop Soup	1.65 2.95
雲 吞 湯	62.	Regular Wonton Soup	1.10 2.10
酸 辣 湯	63.	Hot & Sour Soup	1.10 2.10
蛋 花 湯	64.	Egg Drop Soup.....	1.10 2.10
雲 雞 湯	65.	Egg Drop Wonton Mix.....	1.10 2.10
豆 腐 菜 湯	66.	Tofu Vegetable Soup	NA 3.50
雞 玉 米 湯	67.	Chicken Corn Cream Soup	NA 3.50
蟹 肉 玉 米 湯	68.	Crab Meat Corn Cream Soup.....	NA 3.50
海 鮮 湯	69.	Seafood Soup.....	NA 3.50

Parallel corpora

	ENGLISH	MANDARIN
1	i wanna live in a wes anderson world	我想要生活在Wes Anderson的世界里
2	Chicken soup, corn never truly digests. TMI .	鸡汤吧，玉米神马的从来没有真正消化过.恶心
3	To DanielVeuleman yea iknw imma work on that	对DanielVeuleman说，是的我知道，我正在向那方面努力
4	msg 4 Warren G his eday is today 1 yr older.	发信息给Warren G，今天是他的生日，又老了一岁了。
5	Where the hell have you been all these years?	这些年你 TMD 到哪去了
	ENGLISH	ARABIC
6	It's gonna be a warm week!	الاسبوع اليابي حر
7	onni this gift only 4 u	أوني هذه الهدية فقط لك
8	sunset in aqaba :)	غروب الشمس في العقبة:)
9	RT @MARYAMALKHAWAJA: there is a call for widespread protests in #bahrain tmrw	هناك نداء لظاهرات في عدة مناطق غدا

Table 2: Examples of English-Mandarin and English-Arabic sentence pairs. The English-Mandarin sentences were extracted from Sina Weibo and the English-Arabic sentences were extracted from Twitter. Some messages have been shorted to fit into the table. Some interesting aspects of these sentence pairs are marked in bold.

OPUS

... the open parallel corpus

OPUS is a growing collection of translated texts from the web. In the OPUS project we try to convert and align free online data, to add linguistic annotation, and to provide the community with a publicly available parallel corpus. OPUS is based on open source products and the corpus is also delivered as an open content package. We used several tools to compile the current collection. All pre-processing is done automatically. No manual corrections have been carried out.

The OPUS collection is growing! Check this page from time to time to see new data arriving ...
Contributions are very welcome! Please contact <jorg.tiedemann@helsinki.fi>

Search & download resources:

Search & Browse

- OPUS multilingual search interface
- Europarl v7 search interface
- Europarl v3 search interface
- OpenSubtitles 2016 search interface
- EUconst search interface
- Word Alignment Database (old DB)

Tools & Info

- OPUS Wiki
- OPUS API by Yonathan Koren
- Uplug at bitbucket

Some Projects using OPUS

- Let'sMT! - On-line SMT toolkit
- CASMACAT - Computer Aided Translation

Latest News

- 2018-02-15: New corpora: [ParaCrawl](#), [XhosaNavy](#)
- 2017-11-06: New version: [OpenSubtitles2018](#)
- 2017-11-01: New server location: <http://opus.nlpl.eu>
- 2016-10-08: New version: [OpenSubtitles2016](#)
- 2015-10-15: New versions of [TED2013](#), [NCv9](#)
- 2014-10-24: New: [JRC-Acquis](#)
- 2014-10-20: [NCv9](#), TED talks, DGT, [WMT](#)
- 2014-08-21: New: [Ubuntu](#), [GNOME](#)
- 2014-07-30: New: [Translated Books](#)
- 2014-07-27: New: [DOGC](#), [Tanzil](#)
- 2014-05-07: Parallel coref corpus [ParCor](#)

Sub-corpora (downloads & infos):

- [Books](#) - A collection of translated literature ([Books.tar.gz](#) - 535 MB)
- [DGT](#) - A collection of EU Translation Memories provided by the JRC
- [DOGC](#) - Documents from the Catalan Goverment ([DOGC.tar.gz](#) - 2.8 GB)
- [ECB](#) - European Central Bank corpus ([ECB.tar.gz](#) - 3.0 GB)
- [EMEA](#) - European Medicines Agency documents ([EMEA.tar.gz](#) - 13.0 GB)
- The EU bookshop corpus ([EUbookshop.tar.gz](#) - 42 GB)
- [EUconst](#) - The European constitution ([EUconst.tar.gz](#) - 82 MB)
- [EUROPARL](#) v7 - European Parliament Proceedings ([Europarl.tar.gz](#) - 21 GB)
- [GNOME](#) - GNOME localization files ([GNOME.tar.gz](#) - 9 GB)
- [Global Voices](#) - News stories in various languages ([GlobalVoices.tar.gz](#) - 1.2 GB)
- The Croatian - English WaC corpus ([hrenWaC.tar.gz](#) - 59 MB)
- [JRC-Acquis](#)- legislative EU texts ([JRC-Acquis.tar.gz](#) - 11 GB)

MT Evaluation

Is it a good translation?

≡ Google Translate

⋮

Text Documents

ENGLISH - DETECTED HEBREW ENGLISH SWAHILI ↗ KOREAN ENGLISH HINDI ↘

This cat is cute. Her name is Latte. × 이 고양이는 귀엽다. 그녀의 이름은 라떼입니다. ☆

i goyang-inen gwiyebda. geunyeoui ileum-eun lattebnida.

Send feedback

Basic Evaluation Paradigm

- Use parallel test set
- Use system to generate translations
- Compare target translations w/ reference

Human Evaluation

- Ask a human to do evaluation

太郎が花子を訪れた			
	Taro visited Hanako	the Taro visited the Hanako	Hanako visited Taro
Adequate?	Yes	Yes	No
Fluent?	Yes	No	Yes
Better?	1	2	3

- Final goal, but slow, expensive, and sometimes inconsistent

Human Evaluation Shared Tasks

- **Machine Translation**
 - Conference on Machine Translation (WMT) shared tasks
- **Composite Leaderboard**
 - GENIE leadeboard for QA, summarization, MT

BLEU

- Works by comparing n-gram overlap w/ reference

Reference: Taro visited Hanako

System: the Taro visited the Hanako

1-gram: 3/5

2-gram: 1/4

Brevity: $\min(1, |\text{System}|/|\text{Reference}|) = \min(1, 5/3)$ brevity penalty = 1.0

$$\text{BLEU-2} = (3/5 * 1/4)^{1/2} * 1.0 \\ = 0.387$$

- **Pros:** Easy to use, good for measuring system improvement
- **Cons:** Often doesn't match human eval, bad for comparing very different systems

Embedding-based Metrics

- Recently, many metrics based on neural models
 - **BertScore**: Find similarity between BERT embeddings (unsupervised) (Zhang et al. 2020)
 - **BLEURT**: Train BERT to predict human evaluation scores (Sellam et al. 2020)
 - **COMET**: Train model to predict human eval, also using source sentence (Rei et al. 2020)
 - **PRISM**: Model based on training paraphrasing model (Thompson and Post 2020)
 - **BARTScore**: Calculate the probability of source, reference, or system output (Yuan et al. 2021)

Which One to Use?

- **Meta-evaluation** runs human evaluation and automatic evaluation on the same outputs, calculates correlation
- Examples:
 - **WMT Metrics Task** for MT (Mathur et al. 2021)
 - **RealSumm** for summarization (Bhandari et al. 2020)
- Evaluation is hard, especially with good systems!
Most metrics had no correlation w/ human eval over best systems at some WMT 2019 tasks

MT venues and competitions

- MT tracks in *CL conferences
- WMT, IWSLT, AMTA...
- www.statmt.org

- the [NAACL-2006 Workshop on Statistical Machine Translation](#),
- the [ACL-2007 Workshop on Statistical Machine Translation](#),
- the [ACL-2008 Workshop on Statistical Machine Translation](#),
- the [EACL-2009 Workshop on Statistical Machine Translation](#),
- the [ACL-2010 Workshop on Statistical Machine Translation](#)
- the [EMNLP-2011 Workshop on Statistical Machine Translation](#),
- the [NAACL-2012 Workshop on Statistical Machine Translation](#),
- the [ACL-2013 Workshop on Statistical Machine Translation](#),
- the [ACL-2014 Workshop on Statistical Machine Translation](#),
- the [EMNLP-2015 Workshop on Statistical Machine Translation](#),
- the [First Conference on Machine Translation \(at ACL-2016\)](#).
- the [Second Conference on Machine Translation \(at EMNLP-2017\)](#).

		output language							
		Czech	German	English	Estonian	Finnish	Russian	Turkish	Chinese
input language	Czech	33.9							
	German	48.4							
	26.0	48.3	30.9	25.2	18.2	34.8	20.0	43.8	
			24.9						
			34.9						
			28.0						
			29.3						

Discussion

Class discussion

- Pick a 4-line excerpt from a short text (e.g. poem, text message) in English
- Use Google translate to back-translate the text via a pivot language, e.g.,
 - English → Spanish → English
 - English → L1 → L2 → English, where L1 and L2 are typologically different from English and from each other
- Compare the original text and its English back-translation, and share your observations. For example,
 - What information got lost in the process of translation?
 - Are there translation errors associated with linguistic properties of pivot languages and with linguistic divergences across languages?
 - Try different pivot languages: can you provide insights about the quality of MT for those language pairs?