

CLuBS

Cross-Lingual Bibliographic Search

Juliane Stiller & Cristina España-Bonet

Final Workshop
7th June 2019



The CLuBS Project

Contents

1 Introduction

2 Queries in PubPsych

3 Content Translation in PubPsych

4 Evaluation

Introduction

1 Introduction

2 Queries in PubPsych

3 Content Translation in PubPsych

4 Evaluation

CLuBS

Cross-Lingual Bibliographic Search

Cross/multi-lingual search in DE, EN, ES and FR

A retrieval platform —PubPsych— is already online. We want to improve existing translations and add new ones, improve retrieval components with the latest advances in the area.

Introduction

Motivation of the CLuBS project

- academic discourse happens in several languages
- language barrier ⇒ researchers might not understand or even find relevant articles
- basing research solely on results published in dominant languages such as English bears the risk of drawing conclusions on sub-populations [3]

Introduction

Goals of the CLuBS project

- aims to improve multilingual access to academic bibliographic information
- develops, implements & evaluates different approaches to enable Cross-lingual Information Retrieval (CLIR)
- prototypical domain: psychology with PubPsych search engine (<https://pubpsych.eu>)

Introduction

PubPsych – https://pubpsych.eu

The screenshot shows the homepage of the PubPsych website. At the top, there is a dark header bar with the text "PubPsych" and "Startseite". Below this is a browser-style interface showing the URL "https://pubpsych.zpid.de/pubpsych/" in the address bar, along with various icons for back, forward, search, and navigation.

The main content area features a large, rectangular image of a campfire with glowing embers and charred logs. Overlaid on this image is a white search bar with a magnifying glass icon and the word "suchen". To the left of the search bar, there is a dark grey sidebar containing the "PubPsych" logo in white, and two links: "Erweiterte Suche" and "Hilfe", each preceded by a small circular icon.

On the right side of the page, there is a language selection menu with "En | Es | Fr | De" and a red star icon. Below this are two red link buttons: "Startseite" and "Hilfe". At the bottom of the page, there is a footer bar with copyright information ("© PubPsych") and links to "Kontakt", "Nutzungsbedingungen", and "Über PubPsych", each preceded by a small red star icon.

Introduction

PubPsych

- database of psychological literature, treatments, test and research data
- aggregates bibliographic metadata from nine databases produced by several international partners, e.g. MEDLINE, PSYNDEX and NORART
- metadata mainly in English, German, French and Spanish
 - very uneven distribution due to different indexing practices
 - 20% of all content has no English metadata; 5% of content retrievable with Spanish
- ⇒ results on a topic vary depending on query language

Introduction

A typical document in PubPsych

DFK	0272643
Titel	Selbstzuwendung, Selbstakzeptanz, Selbstvertrauen. Psychotherapeutische Interventionen zum Aufbau von Selbstwertgefühl
Person(en)	Potreck-Rose, Friederike; Jacob, Gitta .
Quelle	Stuttgart: Klett-Cotta, 2013 (8., erw. u. aktualis. Aufl.), 245 S., EUR 26,95 [Leben lernen, Nr. 163], ISBN: 978-3-608-89148-5
Jahr	2013
Sprache	German
Abstract	<p>Psychotherapeutische Interventionen zum Aufbau von Selbstwertgefühl werden vermittelt. Im Theorieteil wird das Grundlagenwissen zusammengefasst und die vorhandene Literatur zum Thema Selbstwert wird auf ihren Nutzen für die psychotherapeutische Arbeit hin betrachtet. Im Praxisteil werden methodisch aufeinander aufbauende Interventionen zur Steigerung des Selbstwertgefühls vorgestellt. Das zugrunde liegende therapeutische Konzept definiert den Prozess der Selbstzuwendung als entscheidenden Schritt für die Entwicklung und Förderung von Selbstakzeptanz und Selbstvertrauen. Die konkreten Interventionsvorschläge werden ergänzt durch eine umfangreiche Sammlung von Übungen für Patienten. - Inhalt: (A) Theorie. (1) Zum Stellenwert von Selbstwert, Selbstvertrauen und Selbstakzeptanz in der Psychotherapie. (2) Theoretische Grundlagen (Begriffsbestimmungen; Korrelate des Selbstwertes; Quellen des Selbstwertes; Dynamik des Selbstwertes). (3) Die vier Säulen des Selbstwertes (Selbstakzeptanz; Selbstvertrauen; soziale Kompetenz; soziales Netz). - (B) Praxis. (4) Stufen der Selbstzuwendung: Sensibilisierung des Ichs (Achtsam sein; sich selbst liebenvoll begegnen; für sich sorgen); (5) Schritte zur Selbstakzeptanz: Differenzierung des Wertesystems (sieben Lebensgebiete; Rollenkünster; Energiekreis; Energiesammler). (6) Wege zum Selbstvertrauen: Selbstregulation und Selbstkontrolle (Bedeutung von Selbstregulation und Selbstkontrolle für den Selbstwert; Grundprinzipien der Selbstveränderung; mit Zielen arbeiten). (7) Risiken selbstwertstärkender Interventionen. (8) Interventionen zum Aufbau des Selbstwertgefühls im psychotherapeutischen Gesamtkonzept. - Das Buch wurde für die achte Auflage erweitert und aktualisiert.</p>
Schlagwörter	Selbstwertgefühl - Selbstvertrauen - Psychotherapeutische Techniken - Selbstregulation - Selbsthilfegermögen - Selbstkonzept
Englische Schlagwörter	Self Esteem - Self Confidence - Psychotherapeutic Techniques - Self Regulation - Self Care Skills - Self Concept
Klassifikation	3310 Psychotherapie und psychotherapeutische Beratung 3120 Persönlichkeitseigenschaften und Persönlichkeitsprozesse
Englische Klassifikation	3310 Psychotherapy & Psychotherapeutic Counseling 3120 Personality Traits & Processes
Segment	PSYDEX Therapy - PSYDEX Lit
Methode	1130 intervention method description
Dokumenttyp	Authored Book
Medientyp	Print
Key Phrase	promotion of self-esteem & self-acceptance, theoretical foundations & psychotherapeutic interventions
Datenquelle	PSYDEX © ZPID

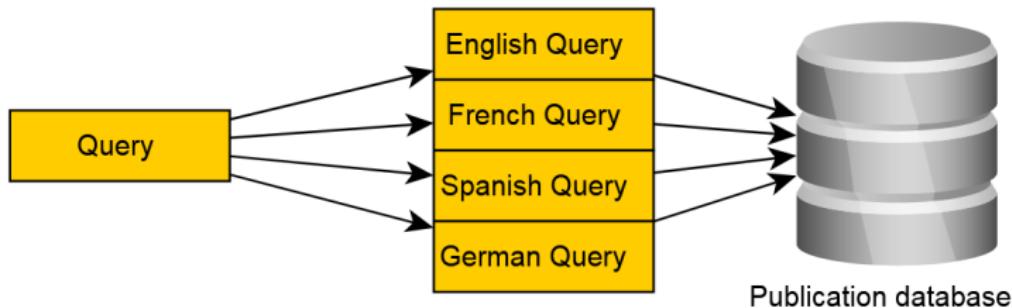
Introduction

Approaches to Cross-lingual Search

- 1** Query translation
- 2** Content translation
- 3** English as a pivot language

Introduction

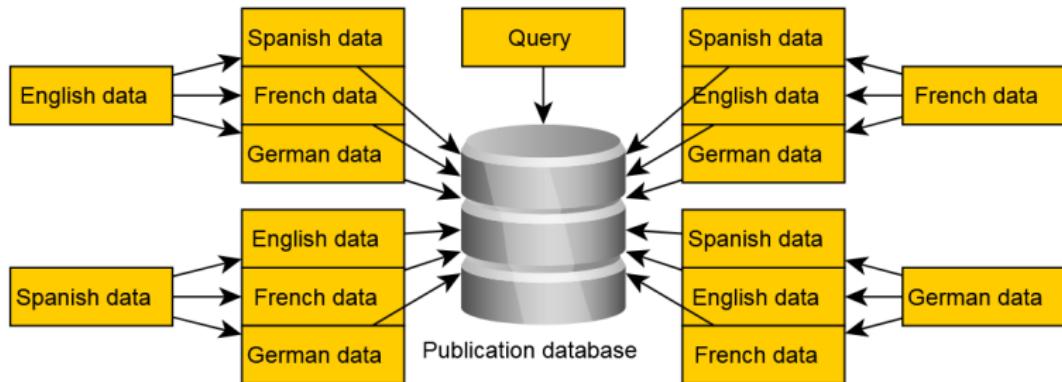
Query Translation



- translation of queries

Introduction

Content Translation



- a translation of content (metadata in this case)

Introduction

English as a Pivot



- a translation of content and query into English

Introduction

Methods employed

- 1 Statistical Machine Translation**
- 2 Neural Machine Translation**
- 3 Vocabulary Mapping**

Queries in PubPsych

Contents

1 Introduction

2 Queries in PubPsych

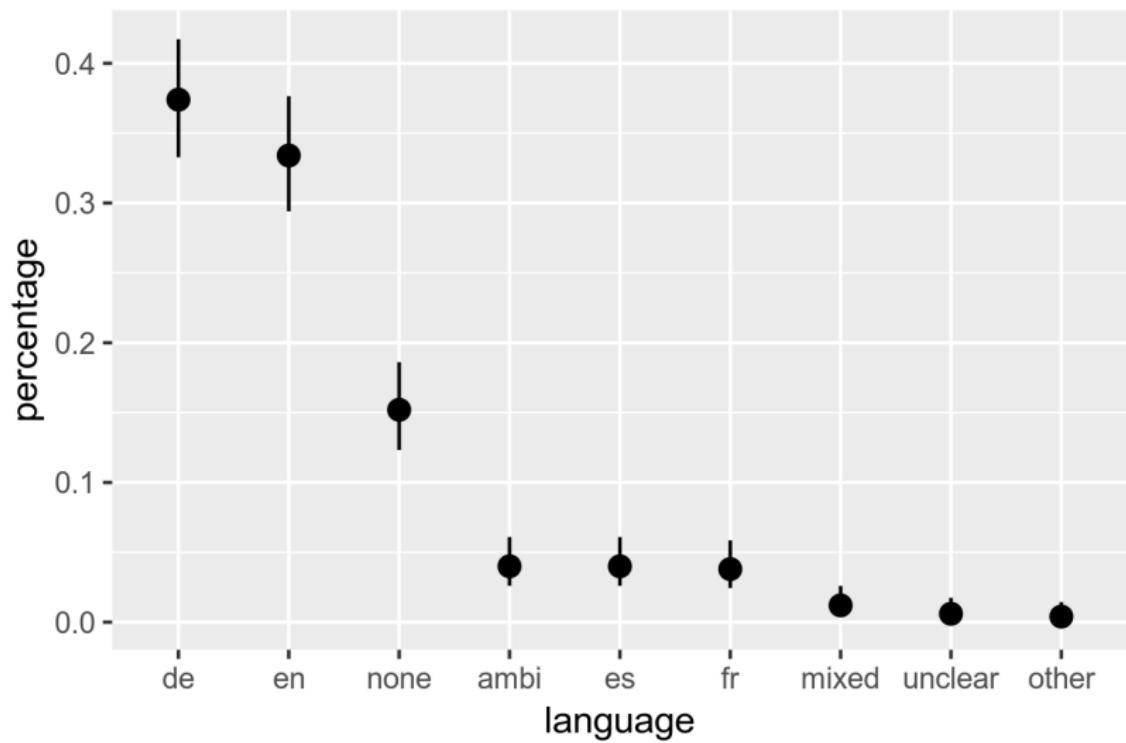
- Queries & Languages
- Keyword & Query Translation
- MeSH-based Quad-lexicon
- Application to Query Translation

3 Content Translation in PubPsych

4 Evaluation

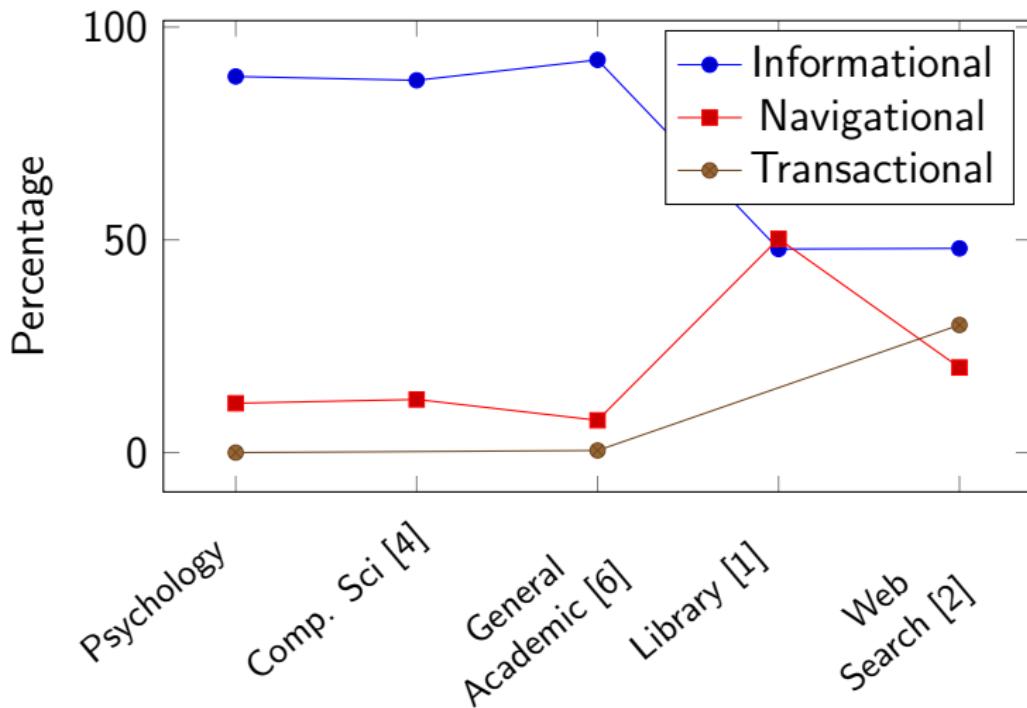
Queries in PubPsych

Query languages



Queries in PubPsych

Query types and distribution in different domains



Queries in PubPsych

Approach

- 88.4% of queries in PubPsych are informational:
mapping approach
- Obtention of **quadrilingual** lexicons
 - Avoid direct language detection
- In-domain but also out-of-domain lexicons
- Seed for query translation
- Seed of the seed: **MeSH** (Medical Subject Headings)

Queries in PubPsych

Multilingual MeSH

```
<concept id="M0000020">
    <term id="T000045" lang="eng" preferred="true">
        <string>Abomasum</string>
        <permutation>Abomasums</permutation>
    </term>
    <term id="spa0000603" lang="spa" preferred="true">
        <string>Abomaso</string>
    </term>
    <term id="spa0049997" lang="spa" preferred="false">
        <string>Cuajar</string>
    </term>
    <term id="ger0000018" lang="ger" preferred="true">
        <string>Labmagen</string>
    </term>
    <term id="fre0063293" lang="fre" preferred="true">
        <string>Abomasum</string>
    </term>
    <term id="fre0000018" lang="fre" preferred="false">
        <string>Caillette</string>
    </term>
</concept>
```

Queries in PubPsych

Lexicon extraction

(key, value) list per language L1 with preferred, non-preferred, and permutations describing a concept mapped to only the preferred term in the other languages L2, L3 and L4.

Queries in PubPsych

Lexicon extraction

(key, value) list per language L1 with preferred, non-preferred, and permutations describing a concept mapped to only the preferred term in the other languages L2, L3 and L4.

mesh.dekey.txt:

Labmagen|||en:Abomasum|||es:Abomaso|||fr:Abomasum|||ID:M0000020

mesh.enkey.txt:

Abomasum|||es:Abomaso|||de:Labmagen|||fr:Abomasum|||ID:M0000020

Abomasums|||es:Abomaso|||de:Labmagen|||fr:Abomasum|||ID:M0000020

mesh.eskey.txt:

Abomaso|||en:Abomasum|||de:Labmagen|||fr:Abomasum|||ID:M0000020

Cuajar|||en:Abomasum|||de:Labmagen|||fr:Abomasum|||ID:M0000020

mesh.frkey.txt:

Abomasum|||en:Abomasum|||es:Abomaso|||de:Labmagen|||ID:M0000020

Caillette|||en:Abomasum|||es:Abomaso|||de:Labmagen|||ID:M0000020

Queries in PubPsych

Is this good enough? As always: Accuracy vs. Recall

```
<concept id="M0025843">
  <term id="T050794" lang="eng" preferred="true">
    <string>Negotiating</string>
  </term>
  <term id="T050795" lang="eng" preferred="false">
    <string>Negotiation</string>
    <permutation>Negotiations</permutation>
  </term>
  <term id="T366631" lang="eng" preferred="false">
    <string>Mediation</string>
  </term>
  <term id="spa0016382" lang="spa" preferred="true">
    <string>Negociación</string>
  </term>
  <term id="ger0016234" lang="ger" preferred="true">
    <string>Verhandeln</string>
  </term>
</concept>
```

mesh.enkey.txt:

Mediation||de:Verhandeln||es:Negociación||fr:Négociation

Our Bet:

Recall

- Retrieve as many related documents as possible
- Enlarge the lexicon with related and out-of-domain multilingual sources

Queries in PubPsych

Multilingual Wikipedia Entries



VIQUIPÈDIA
L'enciclopèdia lliure

Portada
Article a l'atzar
Articles de qualitat

Comunitat
Portal viquipedista
Canvis recents
La taverna
Contacte
Xat
Donatius
Ajuda

En altres idiomes

Boarisch
Deutsch
English
Español
★ Français
Italiano
Nederlands
Русский

Imprimeix/exporta
Crear un llibre
Baixa com a PDF
Versió per a impressora

Pàgina Discussió

Sense sessió iniciada Discussió per aquest IP Contribucions Crea un compte Inicia la sessió

Mostra

Modifica

Mostra l'historial

Més ▾

Cerca a Viquipèdia



Complex d'Èdip

En la **teoria psicoanalítica**, el **complex d'Èdip** planteja la idea que, durant l'**etapa genital** del desenvolupament del nen (una de les **etapes psicosexuals**), que té lloc dels 3 als 6 anys, aquest comença a sentir atracció sexual envers el progenitor del sexe oposat. El nen se sent atret per la mare i comença a odiar el pare, ja que aquest té la mare. Aleshores es desenvolupa un sentiment de **culpa**, ja que el nen sap que no està bé odiar el pare. Aquest conflicte es resol mitjançant la **identificació**, en què el nen adopta característiques del pare. El concepte va ser desenvolupat per **Sigmund Freud**, basant-se en la figura **mitològica d'Èdip**. Freud l'anomena un complex humà universal (**filogenètic**), i la causa de bona part de la culpa inconscient. Postula que existeixen diferents formes d'enfrontar-se al complex: en el cas dels nens, a través del **complex de castració** i en el de les noies en l'**enviada del penis**. Un complex mal resolt, segons Freud, conduceix a la **neurosi, pedofilia o homosexualitat**. Tot això no es pot detectar molt fàcilment però es nota amb el pas del temps.

Tot i que molts cops s'utilitza el nom de "complex d'Èdip" indistintament per a ambdós sexes també es pot anomenar **complex d'Electra** l'etapa de desenvolupament psicosexual corresponent per a la nena per a distingir-la de la del nen.

Crítica [modifica]

Aquesta doctrina psicoanalítica segueix sent controvertida avui en dia, i ha estat criticada o revisada per altres autors com **Otto Rank, Carl Jung, Melanie Klein, Wilfred Bion, Jacques Lacan** o **Hans Keller**, i ha estat especialment figura de polèmica en el **feminisme**.

Els autors **Gilles Deleuze i Félix Guattari** realitzen una crítica del Complex d'Èdip com a estructura autoritària a la seva obra en dos volums **Capitalisme i Esquizofrènia**, en la qual desenvolupen la teoria de l'**esquizoanàlisi**.



Això és un **esborrany** sobre **psicologia**. Amplieu-lo! (citant les fonts)

Categories : Psicoanàlisi | Home

La pàgina va ser modificada per darrera vegada el 17 oct 2017 a les 10:38.

El text està disponible sota la **Llicència de Creative Commons Reconeixement i Compartir-Igual**; es poden aplicar termes addicionals. Vegeu les **Condicions d'ús**. Wikipedia® (Viquipèdia™) és una marca registrada de Wikimedia Foundation, Inc.

Queries in PubPsych

Multilingual Wikidata Entries



Main page
Community portal
Project chat
Create a new item
Create a new Lexeme
Recent changes
Random item
Query Service
Nearby
Help
Donate

Print/export
Create a book
Download as PDF
Printable version

Tools
What links here
Related changes
Special pages
Permanent link
Page information
Concept URI
Cite this page

Item Discussion

Read View history

Search Wikidata



Oedipus complex (Q131246)

concept of psychoanalytic theory; a child's unconscious sexual desire for the opposite-sex parent and hatred for the same [edit](#)

parent

Oedipus Complex | parentcest

▼ in more languages [Configure](#)

Language	Label	Description	Also known as
English	Oedipus complex	concept of psychoanalytic theory; a child's unconscious sexual desire for the opposite-sex parent and hatred for the same-sex parent	Oedipus Complex parentcest
German	Ödipuskonflikt	Konzept der Psychoanalyse	Ödipus-Komplex Ödipal Ödipale Phase Ödipuskomplex
French	complexe d'Edipe	concept central de la psychanalyse	Complexe d'Œdipe Complexe d'Odipe Complexe d'Œdipe LE COMPLEXE D'ŒDIPÉ Complexe oedipien
Bavarian	Edipuskomplex	No description defined	Ödipuskonflikt Ödipuskomplex
Arabic		عقدة أوديب	عقدة أوديب عقدة أوديب عقدة أوديب
Assamese	ঔদিপুস কম্পলেক্স	No description defined	
Asturian	complexu d'Edipu	No description defined	
Azerbaijani	Edip kompleksi	No description defined	
Bulgarian	Едипов комплекс	No description defined	
Bangla	ঔদিপুস কম্পলেক্স	No description defined	
Bosnian	Edipov kompleks	No description defined	
Catalan	complex d'Edip	No description defined	

Wikipedia (64 entries) [edit](#)

ar	عقدة أوديب
ast	Complexu d'Edipu
as	ঔদিপুস কম্পলেক্স
az	Edip kompleksi
bar	Edipskomplex
bg	Едипов комплекс
bn	ঔদিপুস কম্পলেক্স
bs	Edipov kompleks
ca	Complex d'Edip
cs	Oidipovský komplex
da	Ödipuskompleks
de	Ödipuskonflikt
diq	Kompleksé Edipi
el	Οιδίποδειο σύμπλεγμα
en	Oedipus complex
eo	Edipa konflikto
es	Complejo de Edipo
et	Oidipuse kompleks
eu	Edipo konplexu
fa	عقدة ادیب
fi	Oidipuskompleksi
fr	Complexe d'Œdipe
ga	Coimhléasc Eideapúis
gl	Complexo de Edipo
gu	ઔડિપેસ જ્ઞાની
he	ଓଦିପା
hi	ঔদিপুস কম্পলেক্স
hr	Edipov kompleks
hu	Odipusz-komplexus

Queries in PubPsych

Multilingual Wikipedia Entries II

Wikipedia (*key, value*) dictionaries:

wp.dekey.txt:

Wahrnehmung|||en:Perception|||es:Percepción|||fr:Perception

Echoisches Gedächtnis|||en:Echoic memory|||es:Memoria ecoica|||fr:Mémoire auditive

wp.enkey.txt:

Perception|||es:Percepción|||de:Wahrnehmung|||fr:Perception

Echoic memory|||es:Memoria ecoica|||de:Echoisches Gedächtnis|||fr:Mémoire auditive

wp.eskey.txt:

Percepción|||en:Perception|||de:Wahrnehmung|||fr:Perception

Memoria ecoica|||en:Echoic memory|||de:Echoisches Gedächtnis|||fr:Mémoire auditive

wp.frkey.txt:

Perception|||en:Perception|||es:Percepción|||de:Wahrnehmung

Mémoire auditive|||en:Echoic memory|||es:Memoria ecoica|||de:Echoisches Gedächtnis

Queries in PubPsych

Multilingual Wikipedia Entries II

- 1 Quad-lingual in-domain titles from aligned articles in the psychology and health domains
81,369 entries (union among languages)
- 2 Aligned category names
38,038 entries
- 3 Wikidata aligned entries
5,576,686 entries

Queries in PubPsych

Apertium Dictionaries & Post-edited Translations

Apertium dictionaries

- We use three bilingual dictionaries (*en-de*, *en-es*, *es-fr*)
- Extraction of a quadrilingual dictionary with the overlapping entries

Post-edited automatic translations

- We use DeepL to translate a set of tokens within our controlled terms not covered by the previous resources
- Post-edited with the help of the APA thesaurus...
...Wikipedia & Google 

Queries in PubPsych

Extended Quad-Lexicon

Number of **key entries** per language, quadlingual components

	German	English	French	Spanish
MeSH	70,694	175,004	96,333	66,828
WPtitles (h+ph.)		81,369		
WPcategories		38,038		
Wikidata		5,576,686S		
Apertium	7,792	5,935	6,020	5,846
Manual	4,262	4,142	4,047	4,081

Queries in PubPsych

Extended Quad-Lexicon

Number of **key entries** per language, quadlingual components

	German	English	French	Spanish
MeSH	70,694	175,004	96,333	66,828
WPtitles (h+ph.)		81,369		
WPcategories		38,038		
Wikidata		5,576,686S		
Apertium	7,792	5,935	6,020	5,846
Manual	4,262	4,142	4,047	4,081
<hr/>				
<i>S0: MeSH</i>	70,694	175,004	96,333	66,828
<i>S1: QuadLex</i>	202,128	304,277	225,607	195,937
<i>S2: QuadLex</i>	5,778,814	5,880,963	5,802,293	5,772,623

Queries in PubPsych

Coverage on Controlled Terms (English)

Full descriptors, classes

Source	translated (%)
--------	----------------

S0 ACCNO	344,453 (30.4%)
----------	-----------------

S0 DFK	544,275 (33.3%)
--------	-----------------

S0 NORART	5,630 (24.6%)
-----------	---------------

S0 PDID	197 (43.9%)
---------	-------------

S0 PMID	2,987,945 (86.9%)
---------	-------------------

Queries in PubPsych

Coverage on Controlled Terms (English)

Source	Full descriptors, classes translated (%)	Tokens translated (%)
S0 ACCNO	344,453 (30.4%)	1,325,648 (70.9%)
S0 DFK	544,275 (33.3%)	2,043,618 (77.2%)
S0 NORART	5,630 (24.6%)	34,048 (86.9%)
S0 PDID	197 (43.9%)	623 (80.5%)
S0 PMID	2,987,945 (86.9%)	5,007,120 (97.1%)

Queries in PubPsych

Coverage on Controlled Terms (English)

Source	Full descriptors, classes	Tokens
	translated (%)	translated (%)
S0 ACCNO	344,453 (30.4%)	1,325,648 (70.9%)
S2 ACCNO	598,348 (52.9%)	1,867,350 (99.8%)
S0 DFK	544,275 (33.3%)	2,043,618 (77.2%)
S2 DFK	917,681 (56.1%)	2,645,782 (99.9%)
S0 NORART	5,630 (24.6%)	34,048 (86.9%)
S2 NORART	7,263 (31.8%)	39,167 (100.0%)
S0 PDID	197 (43.9%)	623 (80.5%)
S2 PDID	287 (63.9%)	774 (100.0%)
S0 PMID	2,987,945 (86.9%)	5,007,120 (97.1%)
S2 PMID	3,096,000 (90.0%)	5,160,435 (100.0%)

Queries in PubPsych

Comments

- Using the quadrilingual lexicon to translate an entry does not assure a **correct translation**:
 - Noise (remember *mediation!*)
 - The concatenation of word translations does not need to correspond to the term translation

Queries in PubPsych

Comments

- Using the quadrilingual lexicon to translate an entry does not assure a **correct translation**:
 - Noise (remember *mediation!*)
 - The concatenation of word translations does not need to correspond to the term translation

Let's see more analyses when applied to
query translation!

Queries in PubPsych

Contents

1 Introduction

2 Queries in PubPsych

- Queries & Languages
- Keyword & Query Translation
- MeSH-based Quad-lexicon
- Application to Query Translation

3 Content Translation in PubPsych

- CLuBS Settings
- Neural Machine Translation

4 Evaluation

- Experiment Design
- Experiment in Progress ...

Queries in PubPsych

Application to Query Translation

On-line translation of the queries using the
language-independent version of the quadrilingual lexicon

- Contrary to CT translation, we do not know the language of a query

Queries in PubPsych

Application to Query Translation

On-line translation of the queries using the *language-independent* version of the quadrilingual lexicon

- Contrary to CT translation, we do not know the language of a query

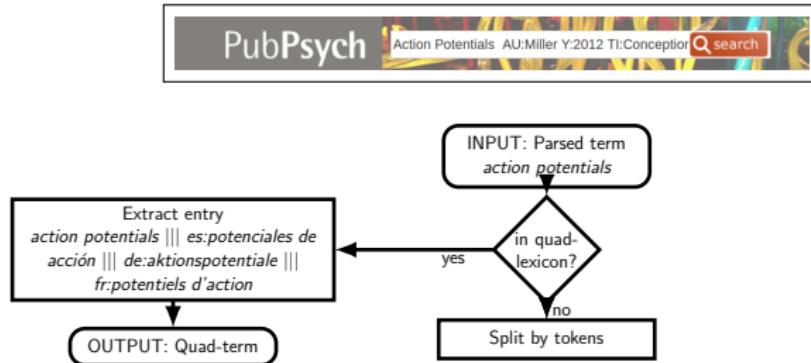
Duplicates (*multiplicates!*) among languages do exist.

Priorities:

- 1 query language frequency: $de \rightarrow en \rightarrow es \rightarrow fr$
- 2 lexicon *domainess*: MeSH → Rest

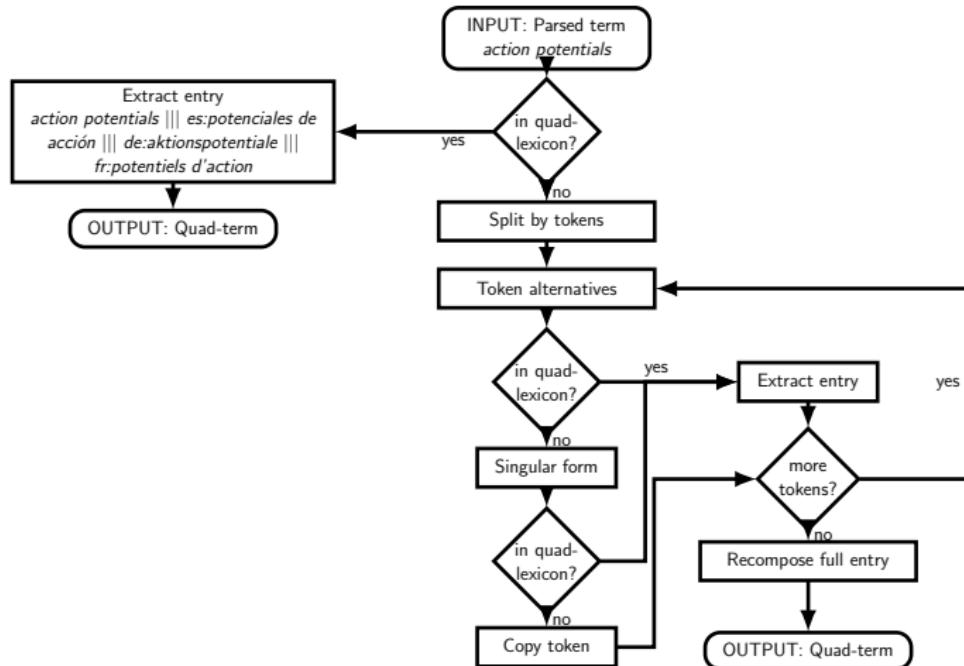
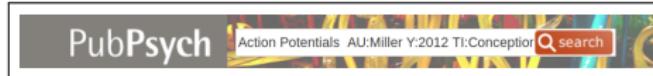
Query Translation

Application to Query Translation



Query Translation

Application to Query Translation



Query Translation

Coverage

- Analysis of 2 million system logs (queries)

Source	Terms		Tokens	
	trad (%)	untrad (%)	trad (%)	untrad (%)
MeSH	167,152 (7.7%)	2,010,469 (92.3%)	2,225,598 (64.2%)	1,240,800 (35.8%)
QuadLex	324,033 (14.9%)	1,853,588 (85.1%)	2,945,959 (85.0%)	520,439 (15.0%)

- **85%** of tokens are translated

Query Translation

Comments

- Using the quadrilingual lexicon to translate an entry does not assure a **correct translation**:
 - Noise
 - The concatenation of word translations does not need to correspond to the term translation
- Copying the source word into the output does not necessarily correspond to a **wrong translation** (e.g. named entities)

Query Translation

Human Evaluation, Adequacy

Annotations	Category description	Example
<i>de, en, fr, es, other</i>	unique language	"Brain"
<i>mixed</i>	at least two languages	"schreiben older adults"
<i>ambi</i>	identical in at least two languages	"Psychologie"
<i>none</i>	not a specific language (e.g. ISSN)	"1869-7712"
<i>unclear</i>	query language and content unclear	"ey Lim"

- 500 queries (100/language + none)

Query Translation

Human Evaluation, Adequacy

Annotations	Category description	Example
<i>de, en, fr, es, other</i>	unique language	"Brain"
<i>mixed</i>	at least two languages	"schreiben older adults"
<i>ambi</i>	identical in at least two languages	"Psychologie"
<i>none</i>	not a specific language (e.g. ISSN)	"1869-7712"
<i>unclear</i>	query language and content unclear	"ey Lim"

- 500 queries (100/language + none)
- 3 raters
- 3-point scale [0, 1, 2]
- Adequacy: **1.4**

Query Translation

Automatic Evaluation, Accuracy

Annotations	Category description	Example
<i>de, en, fr, es, other</i>	unique language	" Brain"
<i>mixed</i>	at least two languages	" schreiben older adults"
<i>ambi</i>	identical in at least two languages	" Psychologie"
<i>none</i>	not a specific language (e.g. ISSN)	" 1869-7712"
<i>unclear</i>	query language and content unclear	" ey Lim"

- 500 queries (100/language)

Query Translation

Automatic Evaluation, Accuracy

Annotations	Category description	Example
<i>de, en, fr, es, other</i>	unique language	" Brain"
<i>mixed</i>	at least two languages	" schreiben older adults"
<i>ambi</i>	identical in at least two languages	" Psychologie"
<i>none</i>	not a specific language (e.g. ISSN)	" 1869-7712"
<i>unclear</i>	query language and content unclear	" ey Lim"

- 500 queries (100/language)
- Manual translation of this test set
- Automatic *MT-like* evaluation with BLEU

Query Translation

Automatic Evaluation, "Accuracy"

Evaluation of query translation quality on the previous test set according to BLEU

- BLEU is the *de facto* MT evaluation metric
- BLEU averages n -gram matches between translation and gold standard
- BLEU-1 similar to accuracy at token level
(but BoW in the full query)
- BLEU-2, BLEU-3... similar to accuracy for longer strings

Query Translation

Query Translation Evaluation

BLEU scores for different lexicons, average of the 5 languages and without considering stopwords:

Lexicon	BLEU	BLEU-1	BLEU-2	BLEU-3	BLEU-4
MeSH	5.34	40.10	20.18	7.64	2.96
QuadLex4	13.30	54.58	31.78	14.46	5.90
MeSH+QuadLex4	17.74	58.44	36.42	16.78	7.62
Wikidata	2.47	39.84	21.08	10.96	0.88
MeSH+QuadLex5	18.75	59.82	37.76	18.96	8.78

- Accuracy? Synonymy, ambiguity...

Query Translation

Query Translation Evaluation

BLEU scores for different lexicons, average of the 5 languages and without considering stopwords:

Lexicon	BLEU	BLEU-1	BLEU-2	BLEU-3	BLEU-4
MeSH	5.34	40.10	20.18	7.64	2.96
QuadLex4	13.30	54.58	31.78	14.46	5.90
MeSH+QuadLex4	17.74	58.44	36.42	16.78	7.62
Wikidata	2.47	39.84	21.08	10.96	0.88
MeSH+QuadLex5	18.75	59.82	37.76	18.96	8.78
Google	10.16	56.66	26.10	10.82	3.32
DeepL	13.52	53.58	26.64	14.64	5.74

Query Translation

Comments



Parsed query: +(annorexie mentale)

String components of query: annorexie mentale

Query Translation

Comments



Parsed query: +(annorexie mentale)

String components of query: **annorexie mentale**

Translations:

en::anorexia nervosa de::anorexia nervosa

es::anorexia nerviosa

Query Translation

Comments



Parsed query: +(+annorexie) +(mentale))

String components of query: annorexie, mentale

Query Translation

Comments



Parsed query: +(+annorexie) +(mentale))

String components of query: **annorexie, mentale**

Translations:

en::anorexia de::anorexia es::anorexia

en::mental de::mental es::mental

What we have seen:

Rules and lexicons to translate queries

But... can we use distributional semantics?

(Gareth Jones talk)

- Multilingual BPEed embeddings
- Easily obtained in multilingual NMT systems
- Simultaneous translation and query expansion

Cool research!!

What we have seen:

Rules and lexicons to translate queries

But... can we use distributional semantics?

(Gareth Jones talk)

- Multilingual BPEed embeddings
- Easily obtained in multilingual NMT systems
- Simultaneous translation and query expansion

Cool research!! As today... The lexicon approach seems to be better for us (but not giving up!)

We've been talking for a while...

before Deep Learning in CLuBS,

questions?

Content Translation in PubPsych

Contents

1 Introduction

2 Queries in PubPsych

3 Content Translation in PubPsych

- CLuBS Settings
- Neural Machine Translation

4 Evaluation

Content Translation in PubPsych

Remember... A typical document in PubPsych

DFK	0272643
Titel	Selbstzuwendung, Selbstakzeptanz, Selbstvertrauen. Psychotherapeutische Interventionen zum Aufbau von Selbstwertgefühl
Person(en)	Potreck-Rose, Friederike; Jacob, Gitta .
Quelle	Stuttgart: Klett-Cotta, 2013 (8., erw. u. aktualis. Aufl.), 245 S., EUR 26,95 [Leben lernen, Nr. 163], ISBN: 978-3-608-89148-5
Jahr	2013
Sprache	German
Abstract	<p>Psychotherapeutische Interventionen zum Aufbau von Selbstwertgefühl werden vermittelt. Im Theorieteil wird das Grundlagenwissen zusammengefasst und die vorhandene Literatur zum Thema Selbstwert wird auf ihren Nutzen für die psychotherapeutische Arbeit hin betrachtet. Im Praxisteil werden methodisch aufeinander aufbauende Interventionen zur Steigerung des Selbstwertgefühls vorgestellt. Das zugrunde liegende therapeutische Konzept definiert den Prozess der Selbstzuwendung als entscheidenden Schritt für die Entwicklung und Förderung von Selbstakzeptanz und Selbstvertrauen. Die konkreten Interventionsvorschläge werden ergänzt durch eine umfangreiche Sammlung von Übungen für Patienten. - Inhalt: (A) Theorie. (1) Zum Stellenwert von Selbstwert, Selbstvertrauen und Selbstakzeptanz in der Psychotherapie. (2) Theoretische Grundlagen (Begriffsbestimmungen; Korrelate des Selbstwertes; Quellen des Selbstwertes; Dynamik des Selbstwertes). (3) Die vier Säulen des Selbstwertes (Selbstakzeptanz; Selbstvertrauen; soziale Kompetenz; soziales Netz). - (B) Praxis. (4) Stufen der Selbstzuwendung: Sensibilisierung des Ichs (Achtsam sein; sich selbst liebenvoll begegnen; für sich sorgen); (5) Schritte zur Selbstakzeptanz: Differenzierung des Wertesystems (sieben Lebensgebiete; Rollenkünster; Energiekreis; Energiesammler). (6) Wege zum Selbstvertrauen: Selbstregulation und Selbstkontrolle @Bedeutung von Selbstregulation und Selbstkontrolle für den Selbstwert; Grundprinzipien der Selbstveränderung; mit Zielen arbeiten). (7) Risiken selbstwertstärkender Interventionen. (8) Interventionen zum Aufbau des Selbstwertgefühls im psychotherapeutischen Gesamtkonzept. - Das Buch wurde für die achte Auflage erweitert und aktualisiert.</p>
Schlagwörter	Selbstwertgefühl - Selbstvertrauen - Psychotherapeutische Techniken - Selbstregulation - Selbsthilfegermögen - Selbstkonzept
Englische Schlagwörter	Self Esteem - Self Confidence - Psychotherapeutic Techniques - Self Regulation - Self Care Skills - Self Concept
Klassifikation	3310 Psychotherapie und psychotherapeutische Beratung 3120 Persönlichkeitseigenschaften und Persönlichkeitsprozesse
Englische Klassifikation	3310 Psychotherapy & Psychotherapeutic Counseling 3120 Personality Traits & Processes
Segment	PSYDEX Therapy - PSYDEX Lit
Methode	1130 intervention method description
Dokumenttyp	Authored Book
Medientyp	Print
Key Phrase	promotion of self-esteem & self-acceptance, theoretical foundations & psychotherapeutic interventions
Datenquelle	PSYDEX © ZPID

Content Translation in PubPsych

Approaches to Metadata Translation

1 Controlled Terms

- Quadrilingual lexicons
- Relation to Query Translation

2 Titles & Abstracts

- SMT, NMT, HMT...
- Goal: Improve retrieval, not high quality translation!

Content Translation in PubPsych

CLuBS Languages



Content Translation in PubPsych

CLuBS Language Pairs



Content Translation in PubPsych

In-domain CLuBS Language Pairs



Content Translation in PubPsych

In-domain CLuBS Language Pairs



Content Translation in PubPsych

Data-Based Machine Translation

Data!

Data!

- More data better results. For us...



German–English



German–Spanish

Content Translation in PubPsych

Data-Based Systems: SMT vs. NMT in CLuBS

- SMT needs a system per language pair
- SMT has a translator for titles and another one for abstracts
- SMT uses pivot translation for unknown language pairs

Content Translation in PubPsych

Data-Based Systems: SMT vs. NMT in CLuBS

- SMT needs a system per language pair
- SMT has a translator for titles and another one for abstracts
- SMT uses pivot translation for unknown language pairs

- NMT has a single system for multiple language pairs and structure
- NMT allows zero-shot translation
- NMT shows higher performance (BLEU) for the evaluated pairs

Content Translation in PubPsych

Basic Architectures

Per language pair:

	SMT	NMT
Generic (GEN)	✓	✓
PubPsych (PP)	✓	✓
GEN+PP	✓	✓

Content Translation in PubPsych

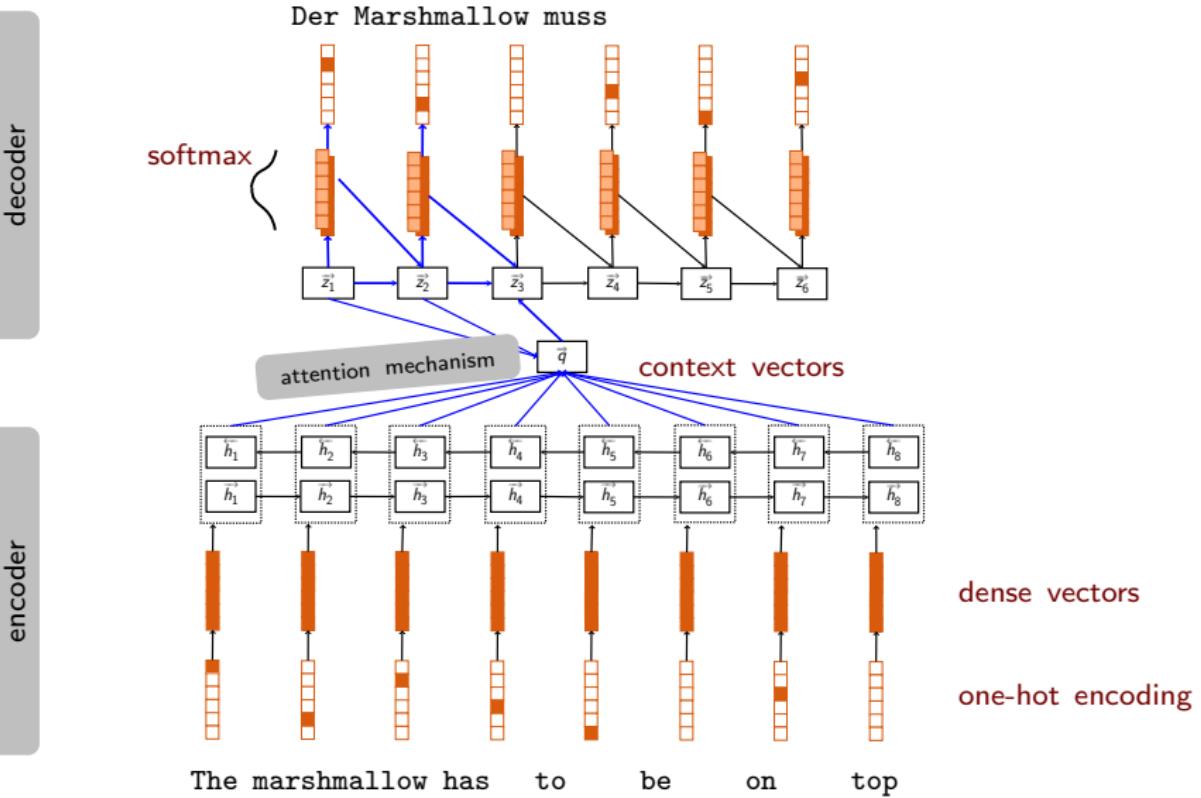
Basic Architectures

Per language pair:

	SMT	NMT	NMT_2	NMT_3	NMT_i
Generic (GEN)	✓	✓	✓	✓	...
PubPsych (PP)	✓	✓	✓	✓	...
GEN+PP	✓	✓	✓	✓	...

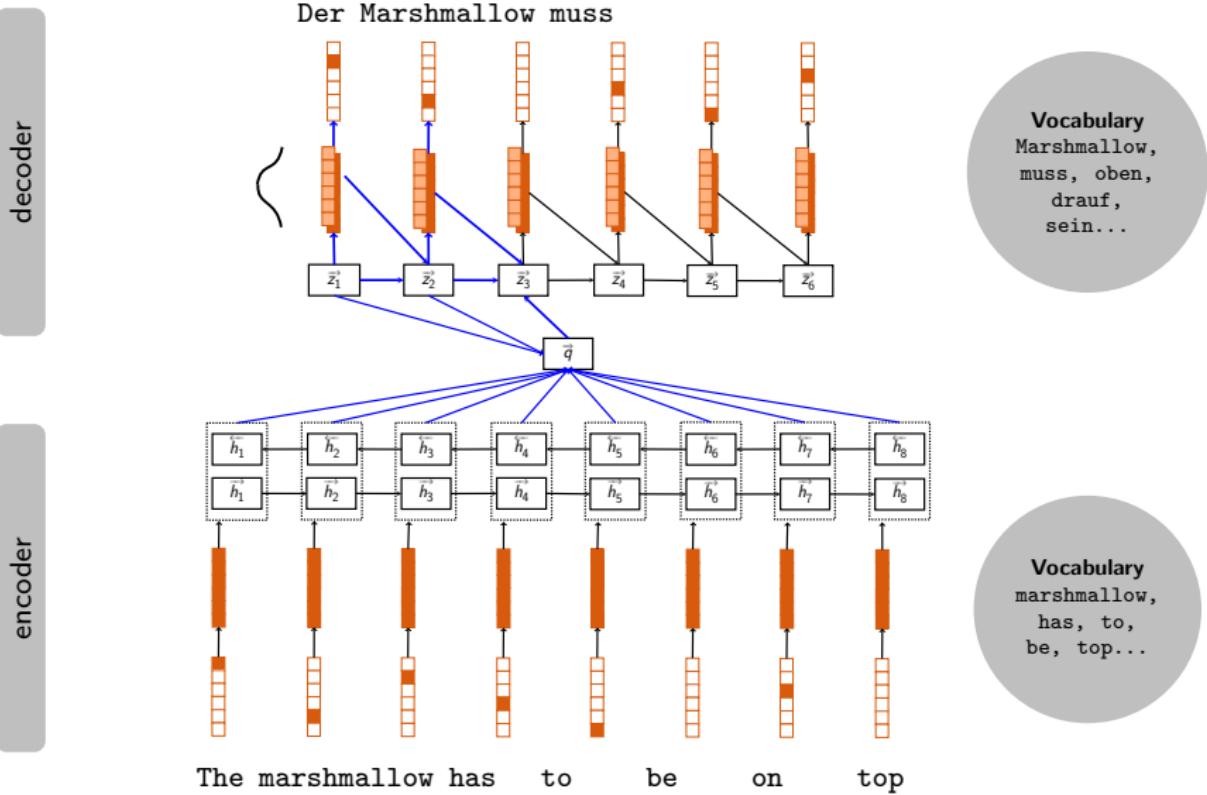
Content Translation in PubPsych

Neural Machine Translation



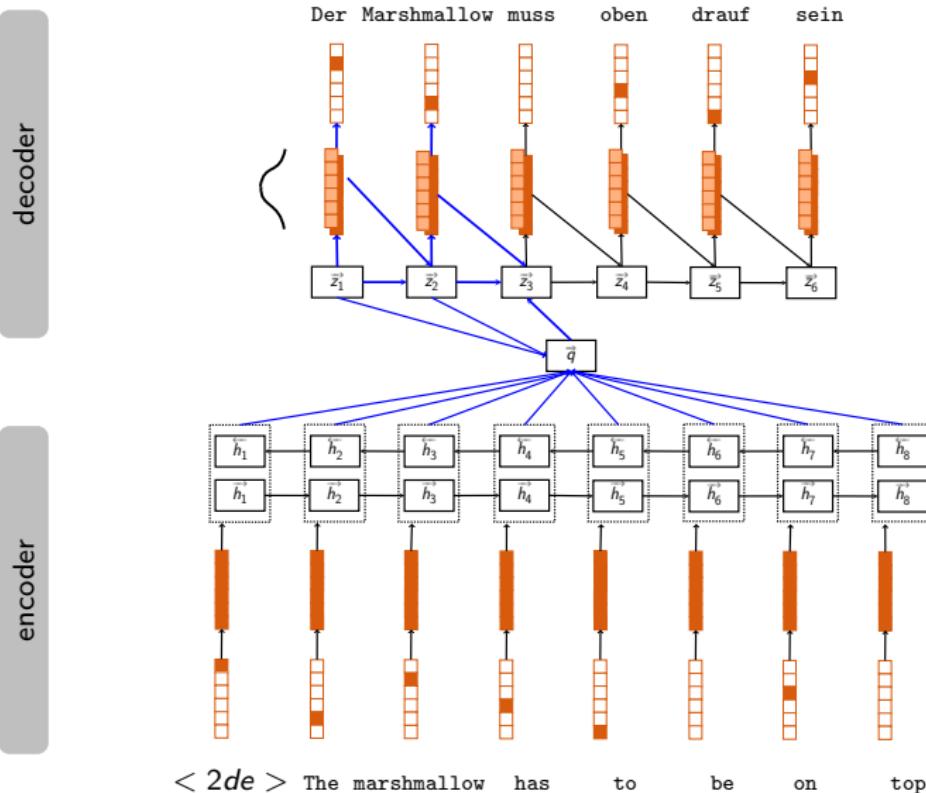
Content Translation in PubPsych

Neural Machine Translation



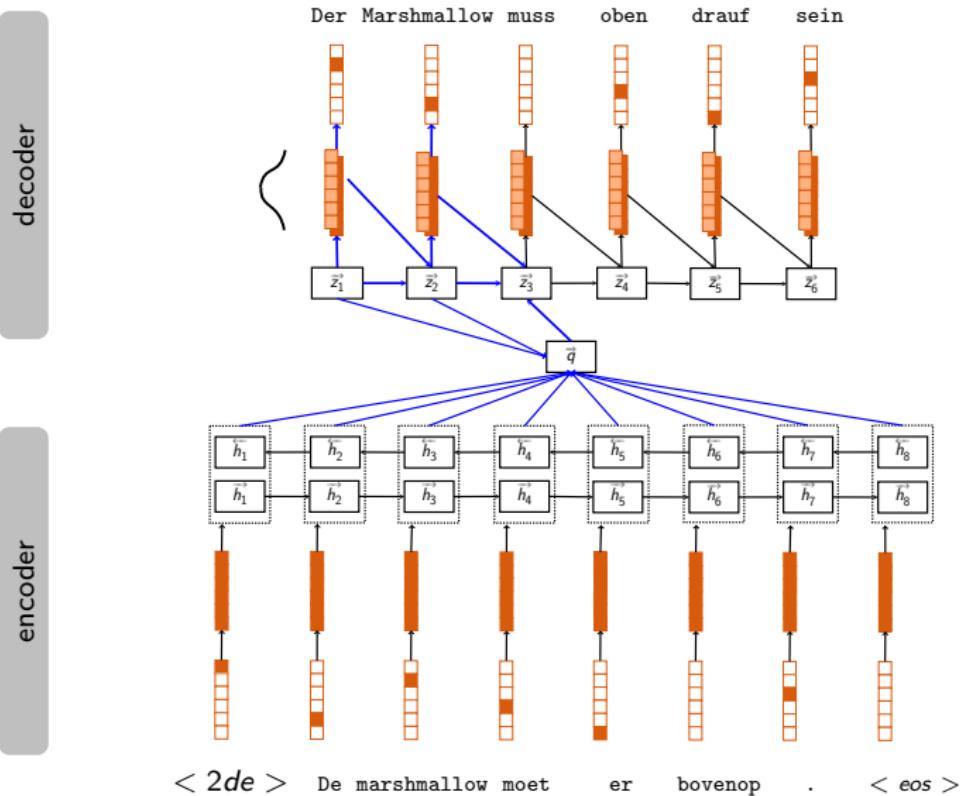
Content Translation in PubPsych

Multilingual Neural Machine Translation



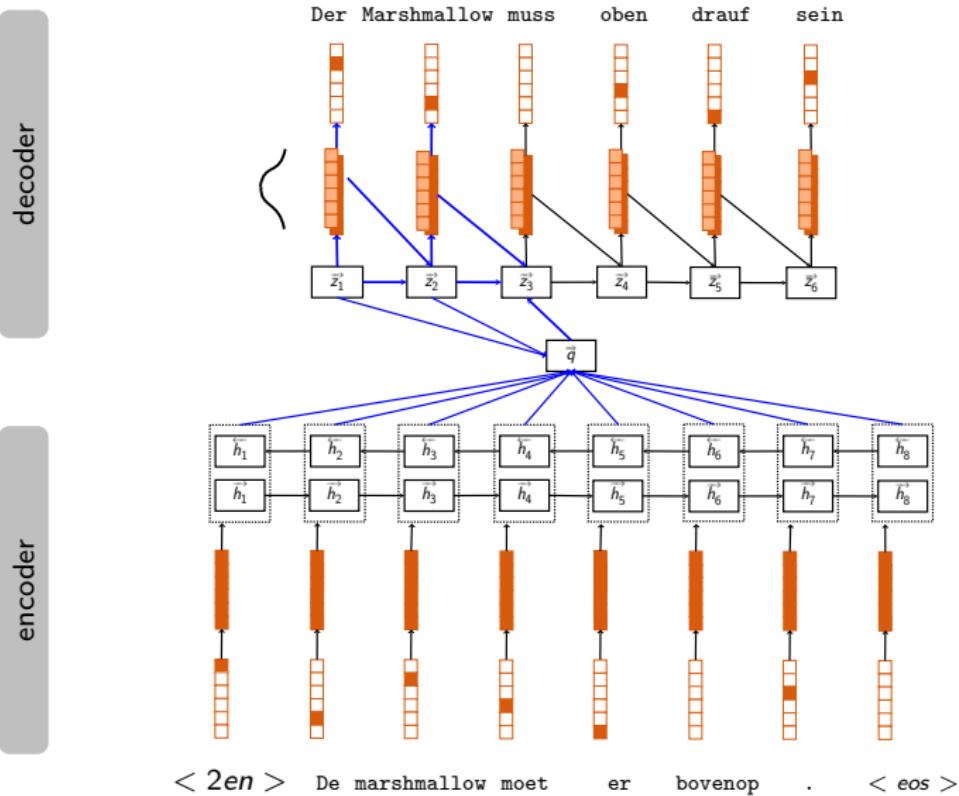
Content Translation in PubPsych

Multilingual Neural Machine Translation



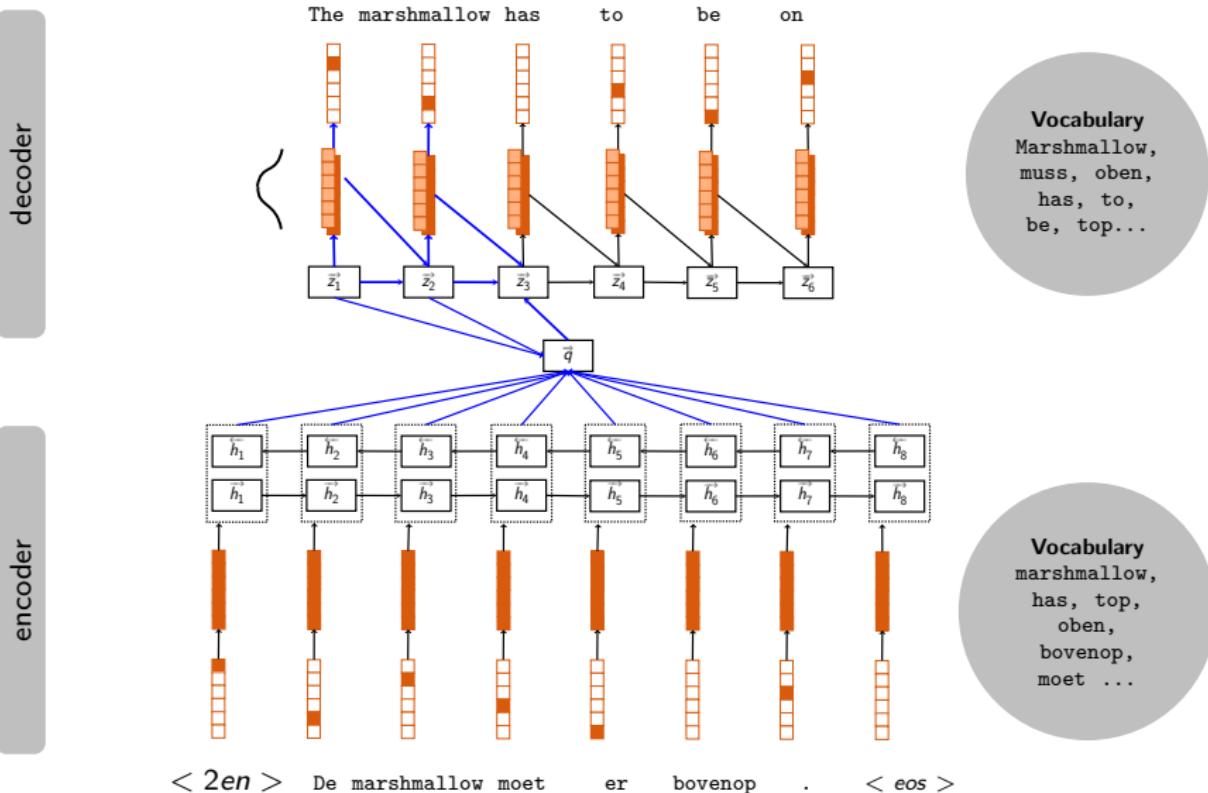
Content Translation in PubPsych

Multilingual Neural Machine Translation



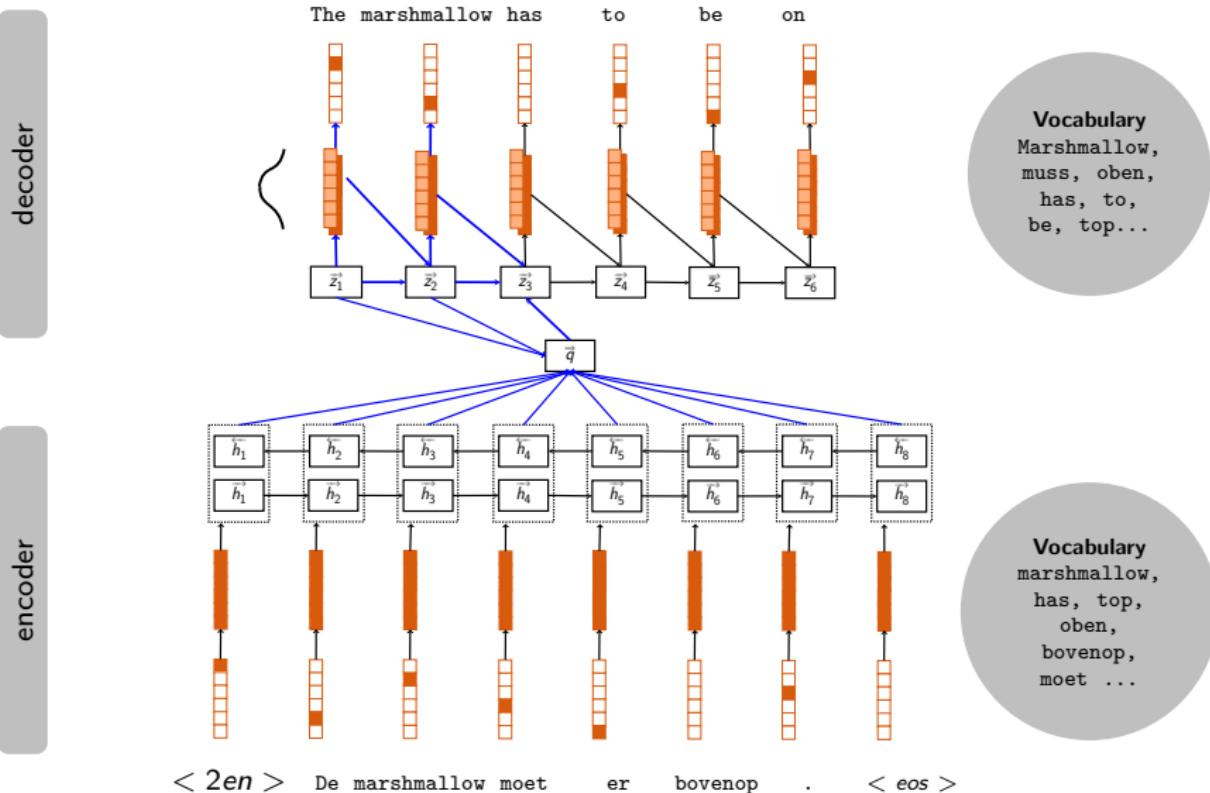
Content Translation in PubPsych

Multilingual Neural Machine Translation



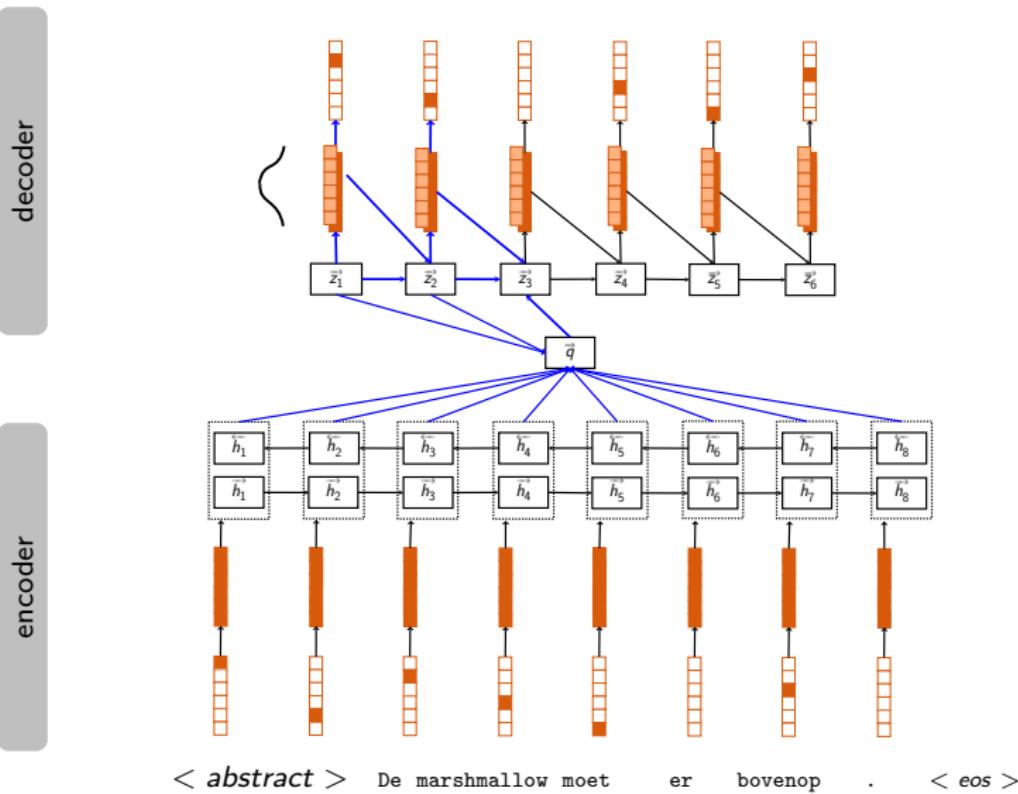
Content Translation in PubPsych

Multilingual Neural Machine Translation



Content Translation in PubPsych

Multilingual Multigenre Neural Machine Translation



Content Translation in PubPsych

Always NMT but several Architectures, Software, Data...

Test set: newstest14 (22k sentences, multilingual)

Software & Architecture (BLEU scores)

- Nematus Seq2Seq 1st: **24.94**
- Nematus Seq2Seq 2nd: **27.76**
- OpenNMT Seq2Seq: **23.64**
- OpenNMT Transformer: **29.65**
- Marian Transformer: **32.10**

Content Translation in PubPsych

Prototype 3: Base system

<i>Software</i>	Marian
<i>Architecture</i>	Transformer Big
Heads	8
Layers	6 (enc) + 6 (dec)
WE dimension:	1,024
FF dimension:	4,096
<i>Corpus size:</i>	96,000,000 parallel sentences (balanced)
<i>Vocabulary size:</i>	150,000 + 2,000 BPE units
<i>Batch size:</i>	5000 tokens

Content Translation in PubPsych

Prototype 3: Domain Adapted NMT

- Base system
- Transfer learning for domain adaptation
 - Continue training with only PubPsych data
 - But no balanced PubPsych corpus!
- Back-translation to improve transfer learning
 - Synthetic corpus by translation with previous system
 - Translations only used as source
 - If bad quality, bad adaptation
 - Iterative process

Content Translation in PubPsych

Prototype 3: Back-translation, the Extent of the Tragedy

Sentences:	Real corpus	Back-translated corpus
<i>en2es</i> titles	25,105	25,000
<i>en2de</i> titles	50,000	—
<i>en2fr</i> titles	45,137	—
<i>de2es</i> titles	—	52,399
<i>de2en</i> titles	50,000	—
<i>de2fr</i> titles	—	47,529
<i>es2de</i> titles	—	50,000
<i>es2en</i> titles	25,105	25,000
<i>es2fr</i> titles	—	47,529
<i>fr2es</i> titles	—	52,399
<i>fr2en</i> titles	45,137	—
<i>fr2de</i> titles	—	50,000

Content Translation in PubPsych

Prototype 3: Back-translation, the Extent of the Tragedy II

Sentences:	Real corpus	Back-translated corpus
<i>en2es</i> abstracts	88,848	100,000
<i>en2de</i> abstracts	241,749	—
<i>en2fr</i> abstracts	—	150,537
<i>de2es</i> abstracts	—	168,419
<i>de2en</i> abstracts	241,749	—
<i>de2fr</i> abstracts	—	150,537
<i>es2de</i> abstracts	—	150,000
<i>es2en</i> abstracts	88,848	100,000
<i>es2fr</i> abstracts	—	150,537
<i>fr2es</i> abstracts	—	168,419
<i>fr2en</i> abstracts	—	150,000
<i>fr2de</i> abstracts	—	150,000

Content Translation in PubPsych

Prototype 3: Automatic Evaluation

German–English: we have real data

Titles	de2en				en2de			
	BLEU	NIST	TER	MTR	BLEU	NIST	TER	MTR
Google	40.91	8.07	43.04	70.55	31.02	6.46	56.32	53.05
DeepL	40.86	8.04	43.57	70.43	30.38	6.34	57.17	52.52
P3	44.61	8.47	40.26	72.39	36.11	7.19	49.94	57.44
Base+DA	45.84	8.55	40.02	72.80	36.44	7.28	49.10	58.20

Abstracts

Google	18.74	5.20	79.25	45.44	13.32	4.19	86.07	32.46
DeepL	19.04	5.24	79.51	46.04	14.23	4.29	85.63	33.27
Base	18.62	5.32	75.46	46.01	13.55	4.37	81.20	33.55
Base+DA	18.95	5.37	74.15	46.45	14.10	4.43	80.12	34.30

Content Translation in PubPsych

Prototype 3: Automatic Evaluation

Spanish–English: we have *some* real data

Titles	es2en				en2es			
	BLEU	NIST	TER	MTR	BLEU	NIST	TER	MTR
Google	42.96	8.02	41.73	73.23	49.92	8.45	39.20	69.97
DeepL	45.08	8.19	39.98	74.61	50.65	8.56	37.19	70.48
Base	46.89	8.40	38.60	74.86	53.44	9.02	33.85	72.62
Base+DA	48.56	8.58	36.75	76.13	56.26	9.27	31.98	74.22

Abstracts

Google	34.43	8.05	53.06	65.16	37.89	8.43	50.85	60.20
DeepL	34.82	8.10	52.82	65.19	39.74	8.63	49.56	61.34
Base	33.93	8.17	52.85	64.71	36.56	8.49	50.17	60.52
Base+DA	34.33	8.22	52.65	64.85	38.21	8.74	48.79	61.75

Content Translation in PubPsych

Prototype 3: Automatic Evaluation

French–English: we have *mostly* back-translations

Titles	fr2en				en2fr			
	BLEU	NIST	TER	MTR	BLEU	NIST	TER	MTR
Google	47.33	7.19	39.70	75.50	42.64	7.08	42.95	65.98
DeepL	45.61	7.09	41.32	73.75	45.88	7.21	42.15	67.14
Base	45.72	7.11	40.90	74.96	43.43	7.06	43.11	66.11
Base+DA	47.17	7.34	39.20	74.97	45.35	7.24	41.21	67.32

Abstracts

Google	29.26	6.70	60.96	59.29	26.42	6.44	62.92	50.77
DeepL	29.69	6.79	59.99	59.31	27.90	6.60	61.70	51.53
Base	27.22	6.62	61.38	58.29	23.66	5.79	63.67	49.19
Base+DA	27.06	6.63	61.11	58.12	24.55	6.03	63.46	49.82

Content Translation in PubPsych

Approaches to Domain Adaptation

Transfer learning with back-translation of in-domain data

Other approaches

- Use of EMEA and Scielo
 -  Not similar enough
- Use synsets to generalise among languages
(tested in a general domain)
 -  Minor improvements

Content Translation in PubPsych

Approaches to Domain Adaptation

Other approaches based on Wikipedia

- Extract parallel sentences in the domain of psychology and health
 - ☒ As a pre-processing, language-dependant, too much noise in general
- Extract multilingual sentences in the domain of psychology and psychiatry
 - ☒ As unsupervised on-line training, language imbalance, still need of back-translation

Translation in CLuBS

Lots of experiments, evaluations and reports
with few but very good hands to do them:

Sophie Henning

Ádám Varga

Dana Ruiter

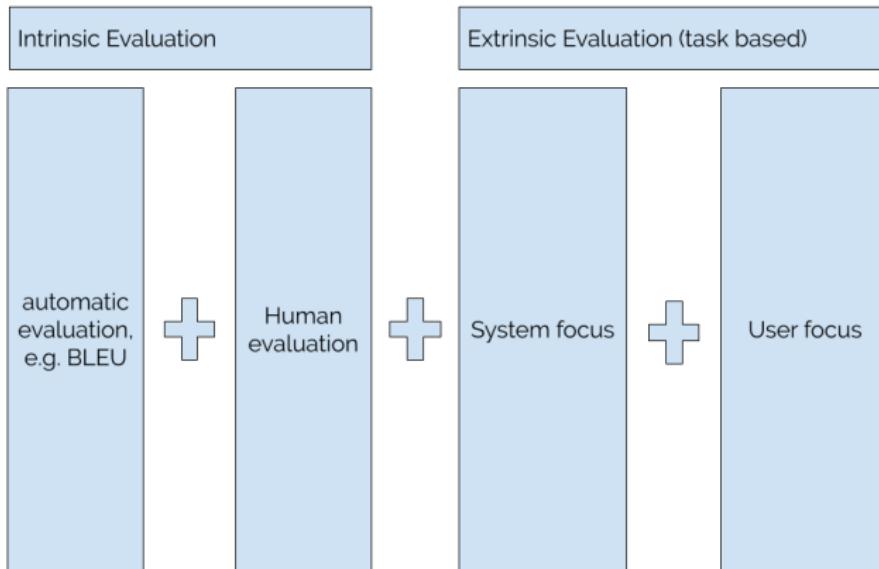
Evaluation

Contents

- 1** Introduction
- 2** Queries in PubPsych
- 3** Content Translation in PubPsych
- 4** Evaluation
 - Experiment Design
 - Experiment in Progress ...

Evaluation

Evaluation Plan



Evaluation

Goals

- Determine the influence of machine translation of different parts of the system on retrieval
- Determine which system performs best with regard to information retrieval
- Compare influence of automatic query translation to human query translation on retrieval

Evaluation

Contents

1 Introduction

2 Queries in PubPsych

- Queries & Languages
- Keyword & Query Translation
- MeSH-based Quad-lexicon
- Application to Query Translation

3 Content Translation in PubPsych

- CLuBS Settings
- Neural Machine Translation

4 Evaluation

- Experiment Design
- Experiment in Progress ...

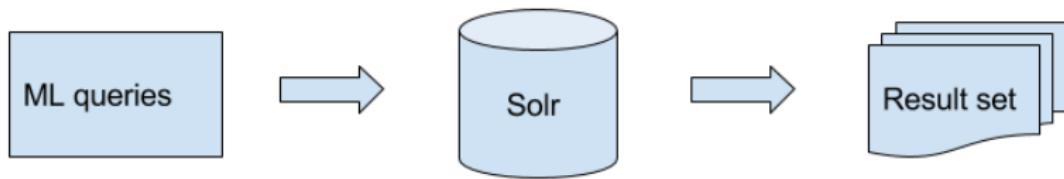
Evaluation

Experiment Design

- 50 queries in 4 languages will be used to query Solr instances with different translated components
- Each query gets a topic description to determine relevant documents
- For 200 searches (4×50 queries) we store up to 10 results
- Remove duplicates in retrieved records
- Pool documents per language
- Assessors assess relevance of documents in their language
- Measuring relevance

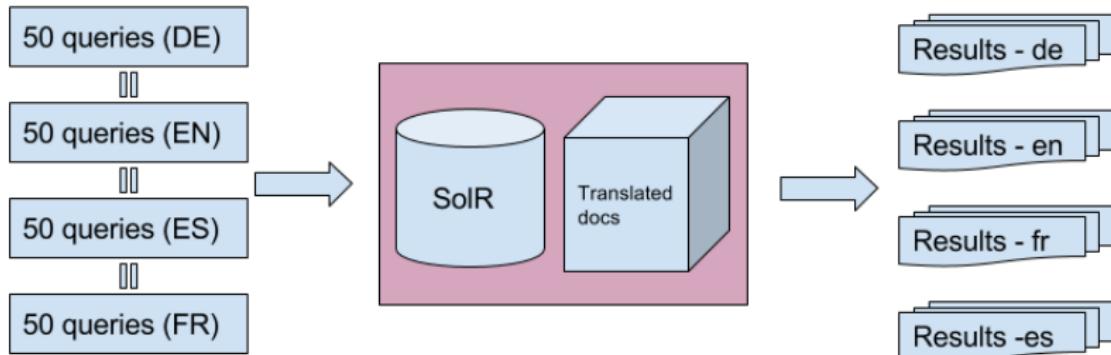
Evaluation

Experiment Design



Evaluation

Solr with translated content



Evaluation

200 original queries

- In English, Spanish, French and German
- Domain experts determined the information need for each query quadruple
- Example: „immigrant depression Canada“ - "immigrant dépression Canada" - "Einwanderer Depression Kanada" - "depresión inmigrantes Canadá"
- Topic description: Highly relevant are documents providing information on the prevalence of depression among immigrants in Canada. Partially relevant are documents containing information on depression of immigrants in other countries.

Evaluation

Relevance Assessment

- 1 **Highly relevant** - the record as presented fulfills the information need and is highly relevant
- 2 **Partially relevant** - the record partially fulfills the information need, but there are some doubts as to whether the whole information need is covered
- 3 **Not relevant** - record does not fulfill the information need

Evaluation

Rating Tool

CLUBS Retrieval Assessment Module

User: juliane

Main language: German

Logout

Query:
children's experiences of hospitalization

Topic:

Highly relevant are documents providing accounts of children,s stays at hospital, as well as information about larger amounts of experiences in the form of statistics.

Partially relevant are accounts of children,s experiences in other care facilities and other age groups,, experiences of hospitalization.

Retrieved Document

Title

[Die Diagnostik belastender Lebensereignisse in der klinischen Routine mittels eines (halb)standardisierten Interviews]

Author

demo_author

Abstract

Bei je 50 stationär und ambulant behandelten kinder- und jugendpsychiatrischen Patienten wurde zusätzlich zur Routinediagnostik ein halbstandardisiertes, blind durchgeführtes Interview zu den psychosozialen Belastungsfaktoren entsprechend der modifizierten Version der Achse V des Multiaxialen Klassifikationsschemas erhoben. Es ergaben sich vergleichbare Informationsmöglichkeiten beim Interview und bei der stationären Routinediagnostik. Vergleichsweise weniger Informationen mit einer entsprechend geringeren Reliabilität zeigten sich hingegen in der ambulanten Routinediagnostik. Es werden Hinweise darauf gegeben, welche psychosozialen Faktoren besser mittels eines Interviews und welche besser mittels der klinischen Einschätzung erhoben werden könnten., Reports on the assessment of 50 inpatients and 50 outpatients in child and adolescent psychiatry using two different methods. Each patient was diagnosed once using routine clinical methods and again in a semistructured interview on psychosocial stressors based on the modified version of Axis V of the Multiaxial Classification Scheme for Psychiatric Disorders in Children and Adolescents. Interviewers were blind to patients' clinical diagnoses. The information acquired during inpatient treatment was comparable to that obtained by interview. However, more information and more reliable information was acquired during the interview than during outpatient treatment. Results indicate that some psychosocial factors can be better assessed by interview, while others are better assessed through clinical diagnosis.

Keywords

eins,zwei,drei

Is this document
relevant?

- Highly relevant
- Partially relevant
- Non relevant

Submit

Skip document

Progress



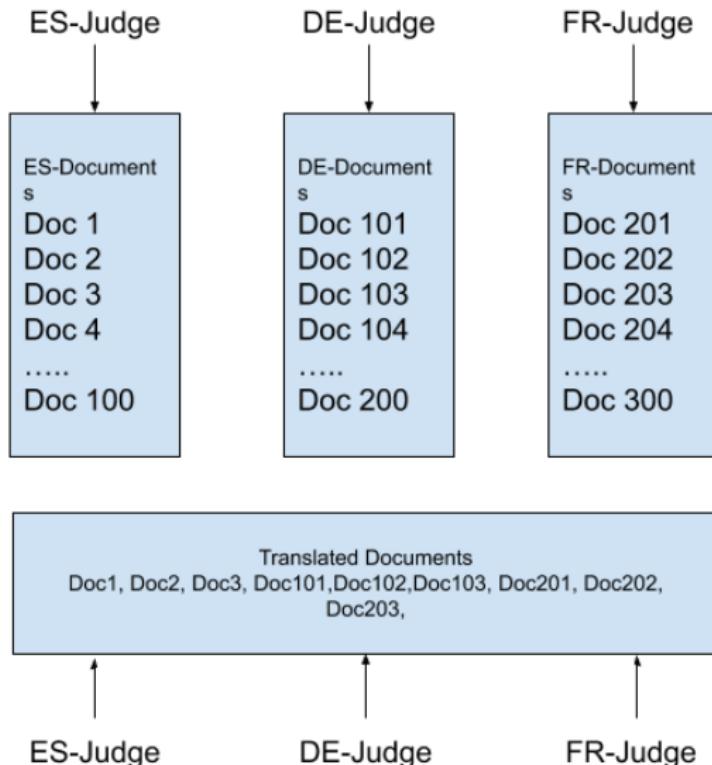
Evaluation

Measures

- R-precision: If the number of relevant documents is $r \leq 10$, we only look at the documents up to the r -th rank of the list. If the number of relevant documents is > 10 , then we calculate R-precision based on 10.
- P@10: We calculate the precision at 10 for the result list.
- Recall(10): How many of the relevant documents were found?
- nDCG (discounted cumulative gain): rank-based measure

Evaluation

Testing the Experiment Design



Evaluation

Results

Assessor	Rating	# of documents
Spanish	Highly relevant	207
	Non relevant	66
	Partially relevant	127
German	Highly relevant	212
	Non relevant	74
	Partially relevant	114

Evaluation

Rating German documents

Rater 1	Highly	Partially	Non	Total
Highly	43	11	0	54
Partially	7	13	1	21
Non	2	10	13	25
Total	52	34	14	100

Evaluation

Interannotator Agreement

2 raters	Translations from Spanish		Translations from German	
	Cohen's Kappa	Scott's Pi	Cohen's Kappa	Scott's Pi
Based on 3 categories highly, partially, non relevant)	0.556	0.554	0.502	0.494
Based on 2 categories (relevant, non)	0.702	0.702	0.664	0.664

According to Landis and Koch [5], the kappa for all categories can be interpreted as ranging from fair (0.21 - 0.40) and moderate (0.41 - 0.60) to substantial (0.61 - 0.80) agreement. The same applies for Scott's Pi.

Evaluation

Intraannotator Agreement

	Rater 1 (German source)		Rater 2 (Spanish source)	
	Cohen's Kappa	Scott's Pi	Cohen's Kappa	Scott's Pi
Based on 3 categories (highly, partially, non relevant)	0.494	0.488	0.6532	0.653
Based on 2 categories (relevant, non)	0.594	0.586	0.554	0.554

According to Landis and Koch [5], the kappa for all categories can be interpreted as ranging from fair (0.21 - 0.40) and moderate (0.41 - 0.60) to substantial (0.61 - 0.80) agreement. The same applies for Scott's Pi.

Evaluation

Learnings from Pilot Experiment

- To minimize bias due to MT, show original language content to raters: French content to French rater, etc.
- To maximize information for raters: extract English translation of content that is provided from sources
- Assessed 50 docs/h (for comparison: assessment of pictures/short documents in CH: 400 docs/h)
- It is an extremely hard task due to topics, domain vocabulary and the nature of abstract (very compressed information)
- Show keywords

Evaluation

Example: Adapting Topic 'Addiction'

Highly relevant are documents investigating how addiction influences personality and the influence of personality on the probability of drug addiction. Partially relevant are documents on the general psychological effects of drug addiction.

Highly relevant are documents investigating how addiction influences personality and the influence of personality on the probability of drug or non-substance addiction. Partially relevant are documents on the general psychological effects of addiction.

Evaluation

Contents

1 Introduction

2 Queries in PubPsych

- Queries & Languages
- Keyword & Query Translation
- MeSH-based Quad-lexicon
- Application to Query Translation

3 Content Translation in PubPsych

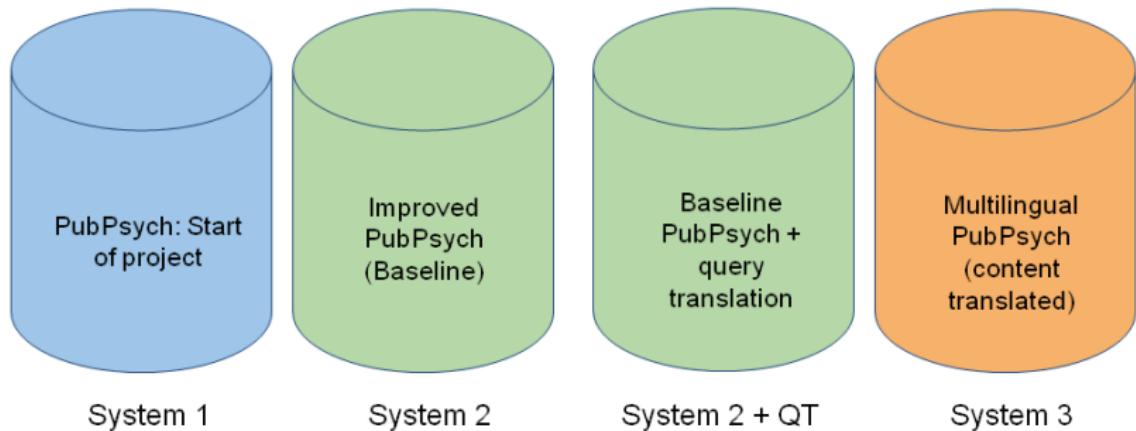
- CLuBS Settings
- Neural Machine Translation

4 Evaluation

- Experiment Design
- Experiment in Progress ...

Evaluation

Testing Systems



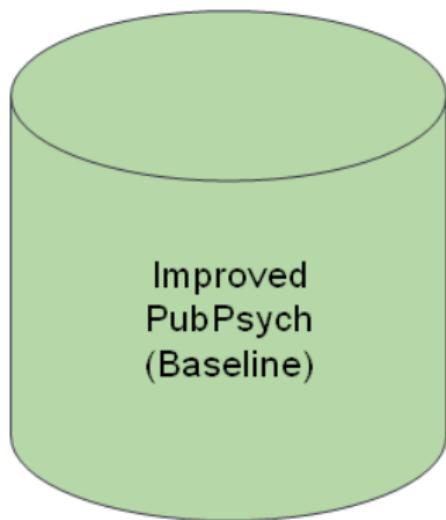
Evaluation

Effects of Improved Ranking



System 1

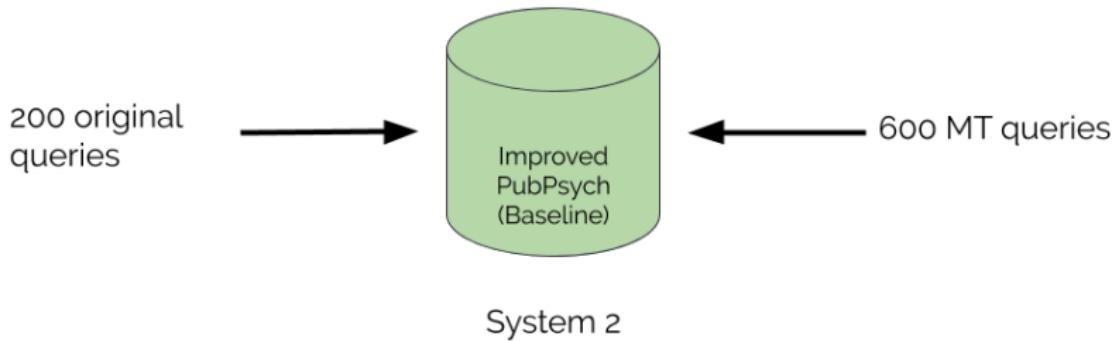
VS.



System 2

Evaluation

A Closer Look on Queries

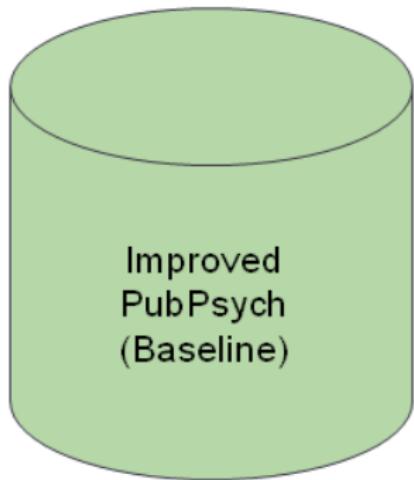


Which query produces the better results list?

- Original German query?
- German query translated from DE, EN or FR?

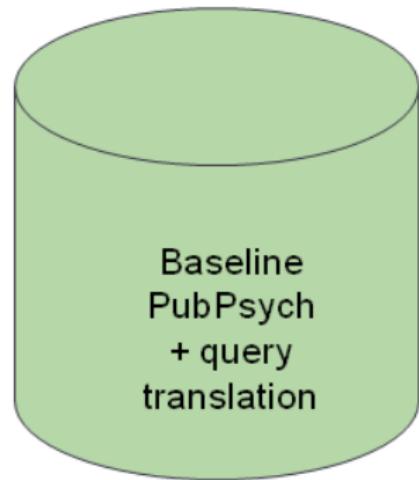
Evaluation

Effects of Query Translation



System 2

VS.



System 2 +
QT

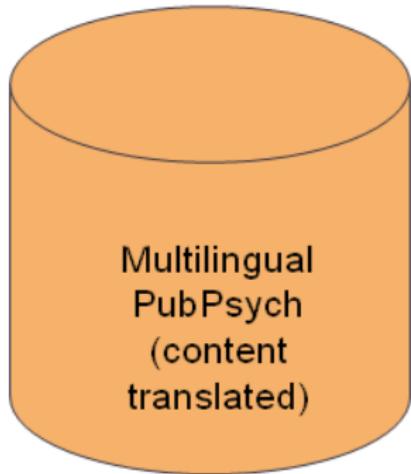
Evaluation

Effects of Content Translation



System 2

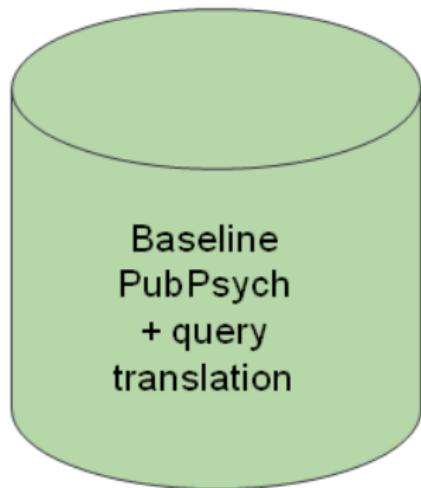
VS.



System 3

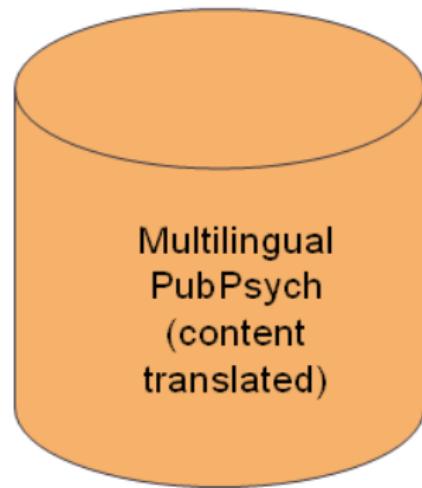
Evaluation

Translated Content vs. Translated Queries



System 2 + QT

VS.



System 3

Evaluation

Different Systems on a Glance

	System 1	System 2	System 2_QT	System 3
Queries	200 orig.	200 orig. + 600 trans.	200 orig.	200 orig.
Content	not altered	not altered	not altered	all translated

Evaluation

Result Set Sizes

Sending 200 original queries to each instance to retrieve first 10 results

Evaluation

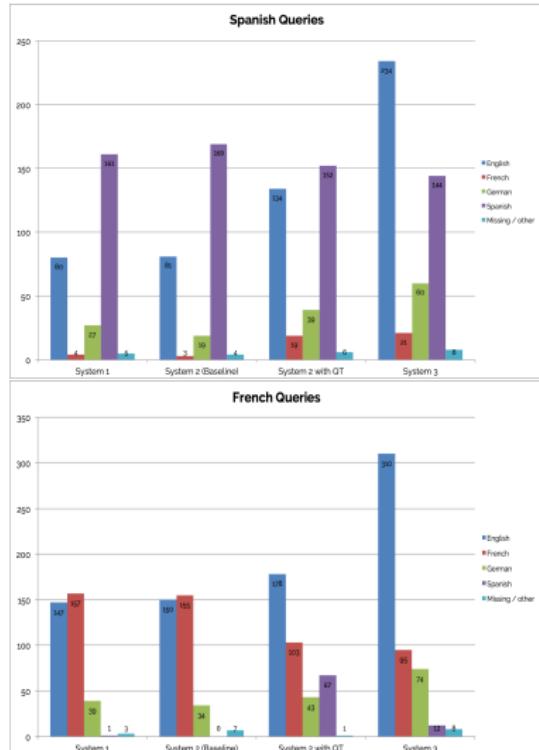
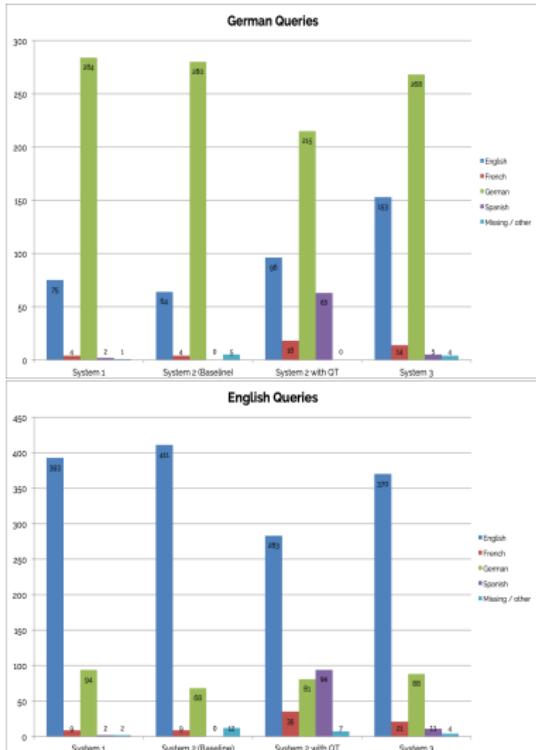
Result Set Sizes

Sending 200 original queries to each instance to retrieve first 10 results

	System 1	System 2	System 2_QT	System 3
0-result queries	53	24	15	5
Sum of found docs	123,182	118,987	964,975	377,074
# of docs top10	1,323	1,477	1,635	1,910

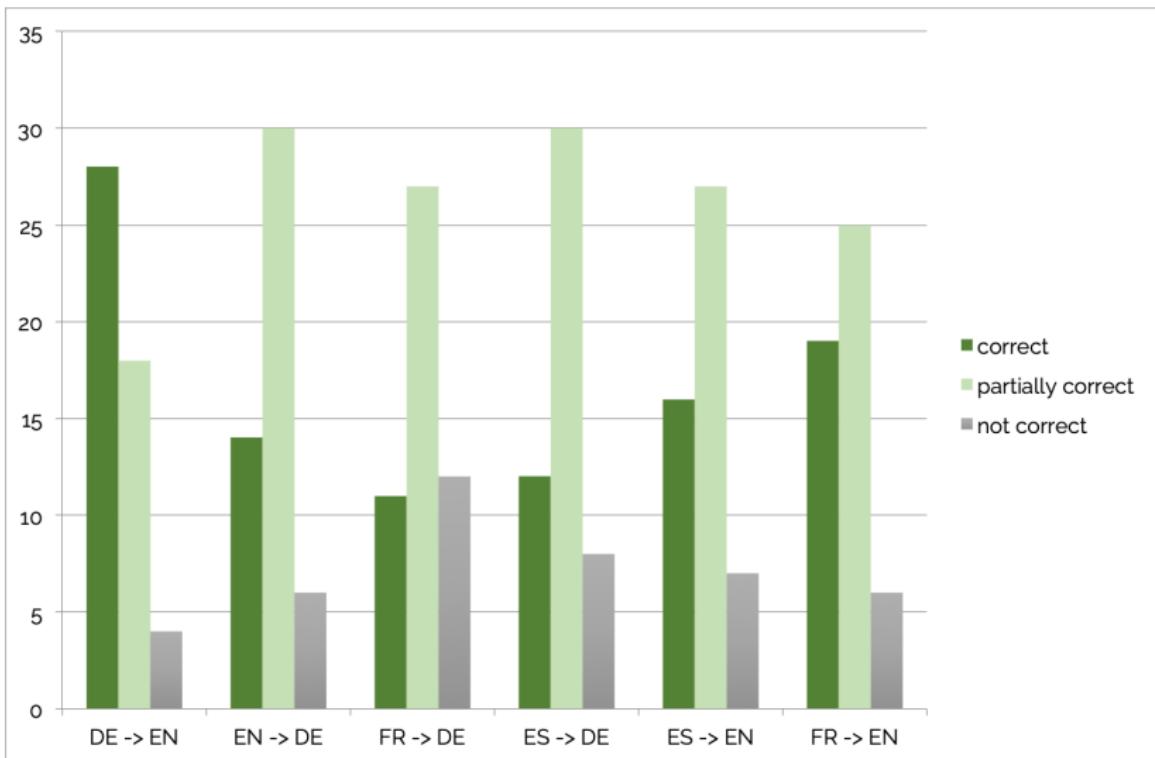
Evaluation

Language of Retrieved Documents



Evaluation

A Closer Look on Queries: Translations \leftrightarrow German



Evaluation

A Closer Look on Queries: German → English

Source query: *Frauen Mathematik*

Source English query: *women math*

Machine translation: *women mathematics*

Evaluation

A Closer Look on Queries: German → English

Source query: *Frauen Mathematik*

Source English query: *women math*

Machine translation: *women mathematics*

What is correct?

Evaluation

Results Lists

PubPsych women math Advanced Search

Results
1 to 10 of 190 for **women math**

by relevance

1. Girls and the leaky math pipeline. Implicit math-gender stereotypes and math achievement in female adolescents and women THESIS/DISSERTATION

Greven, Sarah S.; Ito, Tiffany A.; Park, Briandette; ...
2008
the number of **women** pursuing a postsecondary education and the number of **women** studying **math-intensive** majors. 5 studies were performed using either Implicit Association Tests (IATs) or go ...

2. The effects of gender composition on women's experience in math work groups JOURNAL ARTICLE

Groves, Sarah S.; Ito, Tiffany A.; Park, Briandette; ...
2017 - Source: Journal of personality and social psychology, 2017, 112 (8), 877-890
The present studies tested a model outlining the effects of group gender composition on self- and others' perceptions of **women's** **math** ability in a truly interactive setting with groups composed ...

3. Domain identification moderates the effect of positive stereotypes on Chinese American women's math performance JOURNAL ARTICLE

Saito, Shigehiro; Benet-Martinez, Vanessa; Zane, Nolan; ...
2015 - Source: Cultural diversity & ethnic minority psychology, 2015, 21 (1), 162-7
would improve performance for those more **math** identified (compared to a control condition), but would hinder performance for those less **math** identified. We examined 116 Chinese American **women** (mean age ...

190 Results

PubPsych women mathematics Advanced Search

Results
1 to 10 of 518 for **women mathematics**

by relevance

1. Why do women opt out? Sense of belonging and women's representation in mathematics JOURNAL ARTICLE

Good, Catherine; Rattan, Anesta; Dweck, Carol S.; ...
2012 - Source: Journal of personality and social psychology, 2012, 103 (4), 700-17
belonging to a particular academic domain. [Does] student perceptions of 2 factors in their math environment-the message that math ability is a fixed trait and the stereotype that **women** have less of this ...

2. Unstable identity compatibility: how gender rejection sensitivity undermines the success of women in science, technology, engineering, and mathematics fields JOURNAL ARTICLE

Arnett, Ghanshyam; London, Bonita; Rosenthal, Lisa; ...
2013 - Source: Psychological science, 2013, 24 (9), 1644-52
Although high perceived compatibility between one's gender and science, technology, engineering, and **mathematics** (STEM) fields (gender-STEM congruity) has been linked to **women's** success in ...

3. Motivation of men and women in mathematics and language JOURNAL ARTICLE

Vrijen, Anneke; Dierck, Paap J.; Wiersdorp, Uteke; ...
2009 - Source: International journal of psychology: Journal international de psychologie, 2009, 44 (5), 351-9
men and women, and the extent to which gender-appropriate subject-mathematics for **women** and Dutch for **men**-would result in stronger relationships between performance goals, social comparison, affect, and ...

519 Results

Evaluation

A Closer Look on Queries: English → German

Source query: *women math*

Source German query: *Frauen Mathematik*

Machine translation: *Frauen math*

Evaluation

A Closer Look on Queries: English → German

Source query: *women math*

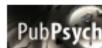
Source German query: *Frauen Mathematik*

Machine translation: *Frauen math*

What is correct?

Evaluation

Results lists

PubPsych Frauen Mathematik   search Advanced Search

Results

1 to 10 of 139 for **Frauen Mathematik**

1. Traumjob Mathematik. Berufswege von Frauen und Männern in der Mathematik

Abele, Andrea E.; Neurauter, Helmut; Többs, Renate.
2004
Diss. von der Volkswirtschaft seit 1998 geförderte interdisziplinäre Projekt "Frauen in der Mathematik. Determinanten von Berufswünschen in der Mathematik unter geschlechtsvergleichender ..."

 Full Record  Check full text availability

by relevance

by date



2. Doppelt benachteiligt? Sprachbehinderte Mädchen in Mathematik und Naturwissenschaften

Sporer, Barbara.
1993 - Diss. von der Sprachwissenschaft, 1993, 38 (2), 93-98, 6 Literaturang.
Anhand von eigenen Beobachtungen, Befragungen von Schülerninnen, Lehrern und Lehrerinnen einer Schule für Sprachbehinderte wird aufgezeigt, dass die traditionell zu bedachtende Berechtigung von Hid ... 

 Full Record  Check full text availability

JOURNAL ARTICLE



3. Girls and mathematics - A "females" issue? A control-value approach to gender differences in emotions towards mathematics

Prenzel, Anne C.; Petrun, Reinhild; Götz, Thomas.
2007 - Source: European Journal of Psychology of Education, 2007, 22 (4), 497-514, 99 Literaturang.
Analyzed gender differences in achievement emotions in the domain of mathematics. Based on R. Pekrun's (2000, 2006) control-value theory of achievement emotions, the authors hypothesized that there ar ... 

 Full Record  Check full text availability

JOURNAL ARTICLE



139 Results

PubPsych Frauen Math   search Advanced Search

Results

1 to 10 of 38 for **Frauen Math**

1. Girls and the leaky math pipeline. Implicit math-gender stereotypes and math withdrawal in female adolescents and women

Jetten, Petra.
2008 -
Investigated the effect of gender stereotypes on math performance and math interest in females. This study looks into the effect of math-gender stereotyping as the reason for the discrepancy between ...

 Full Record  Full Text

THESES/DISSENTIA...



2. When women can't do math: The interplay of self-construal, group identification, and stereotypic performance standards

Kötter, Johannes; Mühe, Udo.
2008 - Source: Journal of Experimental Social Psychology, 2008, 44 (2), 437-444, 46 Literaturang.
Studied the interplay of stereotypically induced self-construal, gender stereotyping, and gender-related performance standards on women's math performance. 114 females (mean age 19 years) were ...

 Full Record  Full Text

JOURNAL ARTICLE



3. Separating implicit gender stereotypes regarding math and language: Implicit ability stereotypes are self-serving for boys and men, but not for girls and women

Jetten, Petra; Götz, Thomas; Petrun, Reinhild; Götz, Thomas.
2011 - Source: Sex Roles, 2011, 64 (5-6), 324-335, 56 Literaturang.
Investigated implicit gender stereotypes related to math and language separately, using go/no-go association tasks. Samples consisted of 100 children (N = 187) and university students (N = 189) in ...

 Full Record  Full Text

JOURNAL ARTICLE



38 Results

Evaluation

Discussion

- Assessments in such a specialized domain are time consuming
- Domain specificity calls for the involvement of experts → crucial for our project
- Multilinguality makes every problem worse
- Real data is messy, real systems are messy
- Multilinguality is far from being solved, e.g. user interactions, preferred languages, UI design, etc.

Evaluation

Discussion II

- CLuBS addresses multilinguality through translation
- The nature of the engine depends on the element to be translated (*keyword vs. abstract*)
- All engines have to be adapted to the domain and that implies a need for in-domain data
- A real scenario such as Pubpsych generates continuously in-domain data (*1.0k docs. in 2017 vs. 1.8k docs. today*)
- So, the MT tomorrow will be better than the MT today.
But... CLIR will be better because MT is better?

-  C. Behnert.
Evaluation Methods within the LibRank Project.
Working Paper, LibRank, 2016.
-  A. Broder.
A taxonomy of web search.
In *ACM Sigir forum*, volume 36, pages 3–10. ACM, 2002.
-  J. Henrich, S. J. Heine, and A. Norenzayan.
Most people are not WEIRD.
Nature, 466(7302):29–29, jul 2010.
-  M. Khabsa, Z. Wu, and C. L. Giles.
Towards better understanding of academic search.
In *Joint Conference on Digital Library (JCDL)*, pages 111–114. ACM, 2016.
-  J. R. Landis and G. G. Koch.
The measurement of observer agreement for categorical data.
Biometrics, 33(1):159–174, 1977.
-  X. Li, B. J. Schijvenaars, and M. de Rijke.
Investigating queries and search failures in academic search.
Information Processing and Management, 53(3):666 – 683, 2017.

Thanks! But...

wait!



questions?

CLuBS

Cross-Lingual Bibliographic Search

Juliane Stiller & Cristina España-Bonet

Final Workshop
7th June 2019

Extra Slides

Comments on Query Translation

PubPsych



 search

Extra Slides

Comments on Query Translation

PubPsych

Risikobereitschaft PRM

search

Parsed query: +(+risikobereitschaft) +(prm))

String components of query: risikobereitschaft, prm

Extra Slides

Comments on Query Translation



Parsed query: +(+risikobereitschaft) +(prm)

String components of query: **risikobereitschaft**, prm

Translations:

en:::risk taking es:::asuncion de riesgos fr:::prise de risque

Extra Slides

Comments on Query Translation



Parsed query: +(+risikobereitschaft) +(prm))

String components of query: risikobereitschaft, prm

Translations:

en::risk taking es::asuncion de riesgos fr::prise de risque

en::photoplethysmography es::fotopletismografia

de::photoplethysmographie

Extra Slides

Comments on Query Translation



Parsed query: +(+risikobereitschaft) +(prm))

String components of query: risikobereitschaft, prm

Translations:

en::risk taking es::asuncion de riesgos fr::prise de risque

en::photoplethysmography es::fotopletismografia

de::photoplethysmographie

Projektorientiertes Risikomanagement

Extra Slides

Comments on Query Translation



Parsed query: +(alltagsmanagement)

String components of query: alltagsmanagement

Extra Slides

Comments on Query Translation



Parsed query: +(alltagsmanagement)

String components of query: **alltagsmanagement**

Translations:

cp::alltagsmanagement

Extra Slides

From Query Rule-based Translation to Embeddings



Parsed query: +(alltagsmanagement)

String components of query: alltags@@, management

Extra Slides

From Query Rule-based Translation to Embeddings



Parsed query: +(alltagsmanagement)

String components of query: **alltags@@, management**

Translations:

en::everyday es::cotidiano fr::ordinaires

en::management es::gestion fr::gestion

Extra Slides

Flashback: Context Translation Engine

- ML-NMT $\{en, es, fr, de\} \Rightarrow \{en, es, fr, de\}$
- 150K vocabulary (+BPE)

Extra Slides

Flashback: Context Translation Engine

- ML-NMT $\{en, es, fr, de\} \Rightarrow \{en, es, fr, de\}$
- 150K vocabulary (+BPE)

Example:

< 2en > bis zur Gerichtsverhandlung befand sie sich vor Ort in Polizeigewahrsam .

< 2en > bis zur Gerichts@@ verhandlung befand sie sich vor Ort in Polizei@@ gew@@ ahr@@ sam .

Extra Slides

Flashback: Context Translation Engine

- ML-NMT $\{en, es, fr, de\} \Rightarrow \{en, es, fr, de\}$
- 150K vocabulary (+BPE)

Example:

<2en> bis zur **Gerichtsverhandlung** befand sie sich vor
Ort in **Polizeigewahrsam** .

<2en> bis zur **Gerichts@@ verhandlung** befand sie sich
vor Ort in **Polizei@@ gew@@ ahr@@ sam** .

Extra Slides

Multilingual Word Embeddings in NMT

- ML-NMT systems create a semantic representation in a **multilingual common space**
- But (seq2seq) NMT uses much more than multilingual word embeddings for translating (e.g. sentence embeddings, attention)
- More important... NMT uses a *language model*, the **decoder**

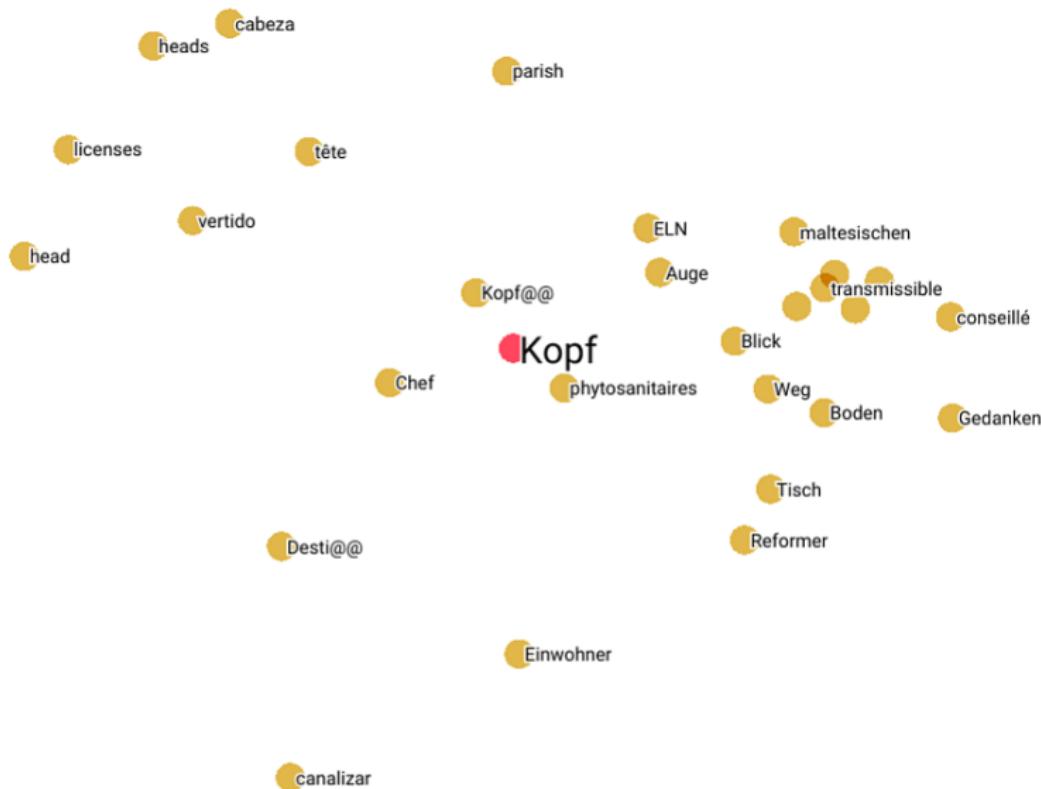
Extra Slides

Multilingual Word Embeddings for Query Translation

- Purpose: **improve coverage** with MLWE (avoid verbatim translations)
- **In-domain terms** still translated by MeSH quad-lexicon
- **Out-of-domain terms** translated by this by-product of the general domain NMT system

Extra Slides

MLWE, 2D representation I



Extra Slides

MLWE, 2D representation II



Extra Slides

MLWE, 2D representation III



Extra Slides

MLWE, 2D representation IV



Extra Slides

MLWE, 2D representation IV



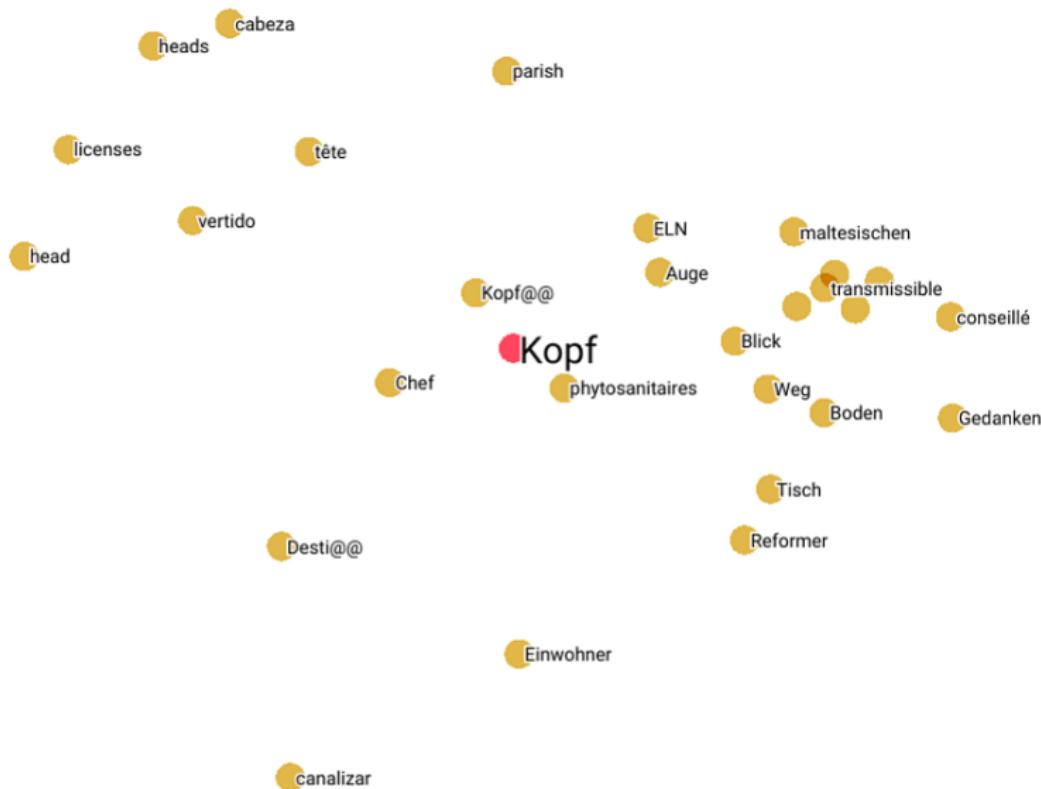
Extra Slides

Comments

- Word embeddings do not deal with multiword expressions or compounds
- Embeddings describe **semantics** not language, a synonym or a translation or a synonym of a translation have the same right to be close to a given word
 - **Query expansion** or query translation?
- Even within a language, translations do not need to be the **closest word**

Extra Slides

Flashback: MLWE, 2D representation I



Extra Slides

MLWE, the Need for Reranking

kopf vs. think

Nearest points in the original space:

transmissible		0.606
cabeza	★	0.618
head	★	0.619
heads		0.621
Kopf@@		0.631
conseillé		0.635
Desti@@		0.638
tête	★	0.657
maltesischen		0.664
capita		0.666

Extra Slides

MLWE, the Need for Reranking

kopf vs. think

Nearest points in the original space:

transmissible		0.606
cabeza	★	0.618
head	★	0.619
heads		0.621
Kopf@@		0.631
conseillé		0.635
Desti@@		0.638
tête	★	0.657
maltesischen		0.664
capita		0.666

Nearest points in the original space:

believe		0.193
consider		0.416
thinks		0.432
thought		0.432
feel		0.443
believes		0.492
thinking		0.522
know		0.531
pense	★	0.539
believed		0.549

Extra Slides

MLWE, the Need for Reranking

kopf vs. think

Nearest points in the original space:

transmissible	0.606
cabeza	★
head	★
heads	0.621
Kopf@@	0.631
conseillé	0.635
Desti@@	0.638
tête	★
maltesischen	0.664
capita	0.666

Nearest points in the original space:

believe	0.193
consider	0.416
thinks	0.432
thought	0.432
feel	0.443
believes	0.492
thinking	0.522
know	0.531
pense	★
believed	0.549

pensar	★	24 th	0.606
denken	★	56 th	0.670

Extra Slides

Reranker

Input

Pairs source word w_1 and n-best list of translations w_2

Dataset

BPEd Quad-lexicon

Features

22 features: basic, lexical, semantic and LM-like

Algorithms

Mainly SVM, XGB and MLP

Basic Features

- 2 words** (w_1 –source–, w_2 –target–)
- 2 languages** (L_1 , L_2)
- 2 subunits**: source is a BPE subunit instead of word
(`srcSubUnit`), or src and tgt have a BPE mark
(`bothBPEmark`)

Lexical Features

- 3 length** in characters without BPE mark @@ ($|l_1|, |l_2|, |l_1|/|l_2|$)
- 1 Levenshtein distance** between tokens w_1 and w_2 (lev)
- 3 character n-gram** similarity
(cosSimN2,cosSimN3,cosSimN4)
- 1 Levenshtein distance** between Metaphone 2 **phonetic keys** (levM2)

Semantic Features

1 Similarity between **words**

$$sim(w_1, w_2)$$

1 **Rank** of the translation, rankW2

4 Distances in similarities **to certain ranks**

$$sim(w_1, top1) - sim(w_1, w_2)$$

(simRankt1,simRankWnext,simRankt10,simRankt100)

Language Model-like Features

- 1 Similarity between **previous words**

$$sim(w_{1-1}, w_{2-1})$$

- 1 Similarity between **bigrams**

$$sim((w_{1-1}+w_1)/2, (w_{2-1} + w_2)/2)$$

Extra Slides

Dataset: BPED Quad-lexicon

Need for positive and negative examples:

- Positive: the translation w2
- Negative: a word w3 in the n-best list **close** to w2

Extra Slides

Dataset: BPED Quad-lexicon

Need for positive and negative examples:

- Positive: the translation w2
- Negative: a word w3 in the n-best list **close** to w2

Current approach takes as w3 the +1/-1 word in the cosine similarity ranking

- For top1 translations always +1;
- higher probability to -1's to get a balanced corpus (5/6!)

Extra Slides

Dataset: BPED Quad-lexicon

Need for positive and negative examples:

- Positive: the translation w2
- Negative: a word w3 in the n-best list **close** to w2

Current approach takes as w3 the +1/-1 word in the cosine similarity ranking

- For top1 translations always +1;
- higher probability to -1's to get a balanced corpus (5/6!)

Probably good for margin-based algorithms, but for others, should it be **completely random**?

Extra Slides

Dataset: Comments

- We can arbitrarily create a data set that allows $> 90\%$ accuracy
- No reason for a concrete design: need for extrinsic evaluation in MT
- Manual translation of the 500 queries test set (now, literally, ~ 300)
- Even then, we will not know the effect on retrieval

Extra Slides

Conclusions

No-one would use MLWE for word translation...

No-one would use MLWE for word translation... **But:**

- We do not aim at perfect translation, but at good retrieval
- It has been shown that retrieval can be improved by query expansion
- MLWE can perform at worst multilingual query expansion
- The specific in-domain terms are still translated by MeSH