

# Suche in multiparallelen Korpora

Multilingwis: <https://pub.cl.uzh.ch/purl/multilingwis>

Zum Vergleich:

Linguee: <https://www.linguee.com/>

Glosbe: <https://glosbe.com/>

Tradoit: <https://www.tradoit.com/>

## Einzelwörter (in FEP6 oder FEP9)

- 1) Welche Übersetzungsvarianten sind für Deutsch <sorgen> zu erwarten (in Englisch, Italienisch, etc.)?
- 2) Welche Übersetzungsvarianten kommen tatsächlich vor?

## Mehrwortausdrücke (Beispiele)

- tenere duro
- lasciare cadere
- fare tardi
- mit Kosten verbunden

Ausdrücke in eckigen Klammern:

- [Art und Weise]
- [Leib und Leben]
- [null und nichtig]
- [pour l' amour du ciel]

mit Platzhaltern:

- [keep \* head above water]
- [fare \* passo]

Ausdrücke kombiniert mit freien Wörtern:

- [unter den Tisch] fallen
- [in Frage] stellen
- [in Frage] kommen

- 3) Finde weitere Beispiele in verschiedenen Sprachen. In welchen Fällen liefert die Alignment brauchbare Resultate und in welchen nicht?

# Sprachenlernen mittels Exploration authentischer Übersetzungen

Suche in parallelen Sätzen mit PaCLE: <https://demo.spraakbanken.gu.se/johannes/PaCLE/>

1) Was bedeuten die folgenden schwedischen Ausdrücke? Wie lassen sie sich in andere Sprachen (Deutsch, Englisch, Italienisch, ...) übersetzen?

1	med tiden	
2	kort och gott	
3	över förväntan	
4	öga mot öga	
5	huvud på ett fat	
6	med flit	
7	sakta men säkert	

2) Welcher der Ausdrücke ist am einfachsten, welcher am schwierigsten zu verstehen? Welche Rolle spielt dabei die jeweils andere Sprache?

3) Wie hilfreich sind die jeweils angezeigten Beispiele? Welche Eigenschaften machen gute Beispiele aus?

# Kollokationen in parallelen Korpora berechnen

Programmcode unter [https://pub.cl.uzh.ch/users/graen/CAS\\_TT+AI/](https://pub.cl.uzh.ch/users/graen/CAS_TT+AI/)

## Vorgehen

1. Herunterladen des Parallelkorpus «Rumantsch Grischun» aus der «Parallel Corpus Collection» (PaCoCo) (<https://pub.cl.uzh.ch/purl/PaCoCo>)
2. Extraktion der Satzpaare Deutsch/Rumantsch (de/rm) [4-5]
3. Zählen der Vorkommen einzelner Wörter und der gemeinsamen Vorkommen von Wortpaaren de/rm [9]
4. Berechnung der «pointwise mutual information» (MI) für jedes Wortpaar de/rm [14]
5. Sortierung nach MI-Wert (je höher desto stärker die Assoziation)

$$\text{pmi}(x; y) \equiv \log \frac{p(x, y)}{p(x)p(y)} = \log \frac{p(x|y)}{p(x)} = \log \frac{p(y|x)}{p(y)}.$$

([https://en.wikipedia.org/wiki/Pointwise\\_mutual\\_information](https://en.wikipedia.org/wiki/Pointwise_mutual_information))