
University of Zurich[UZH]

# Termextraktion & Fremdsprachenlernen

## CAS in Translation Technology and AI

Martin Volk & Johannes Graën

5. November 2021

# Termextraktion mittels Wortalignierung

# Termextraktion – Trados

## MultiTerm Extract

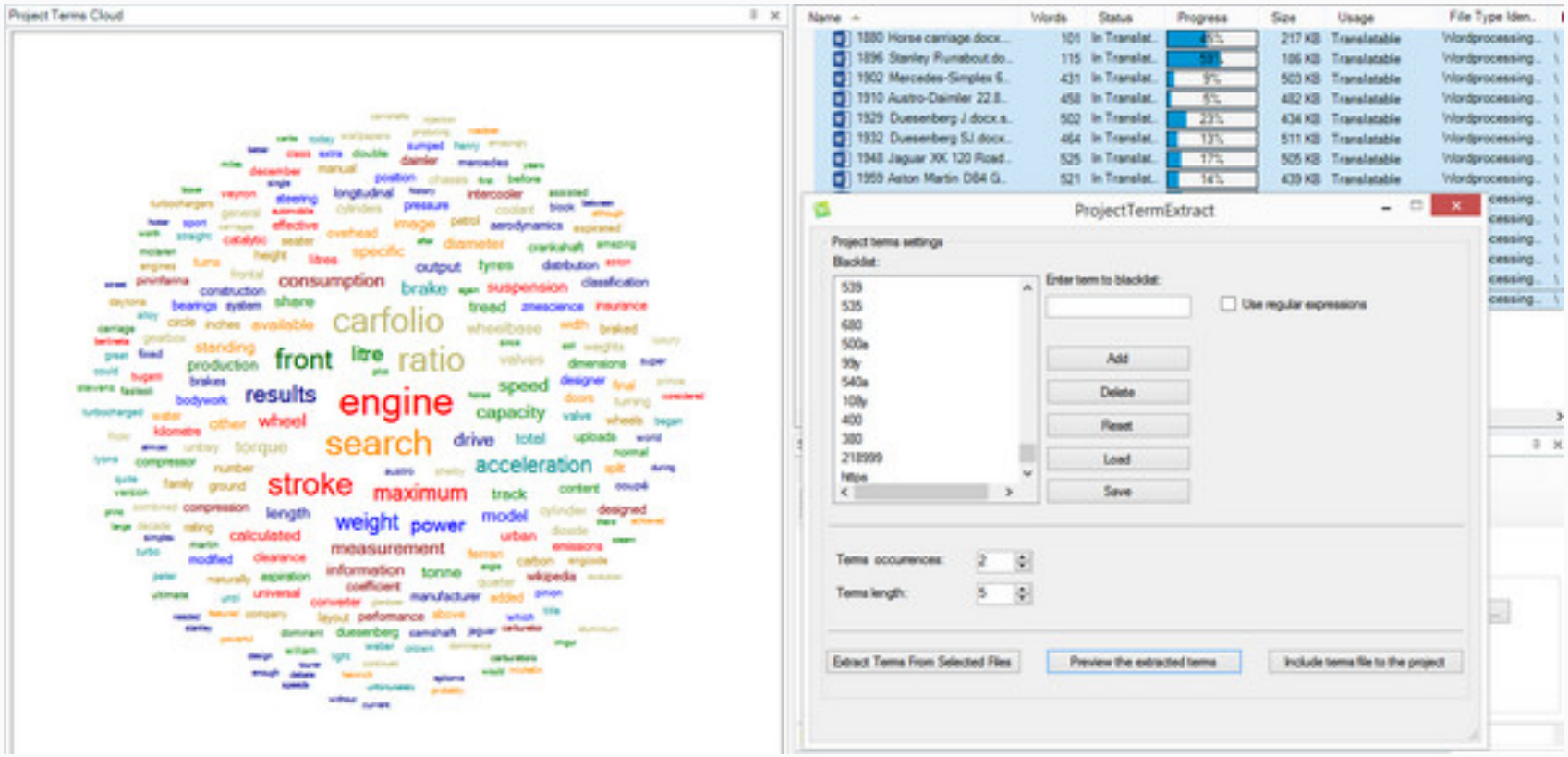
*„MultiTerm Extract checks the frequency of terms at a sub-segment level and enables you to build project glossaries without having to manually search for the terms. You can automatically locate and extract potential monolingual or bilingual terms from existing documentation or TMs to build termbases and glossaries quickly.“*

*„SDL MultiTerm Extract uses a statistical extraction method to determine the frequency of the appearance of candidate terms. It extracts term candidates and, for multilingual termbases, their probable translations found in sentences, incomplete sentence fragments and strings of code. The extracted terms are presented as a term candidate word or phrase.“*

## projectTermExtract

*„This projectTermExtract plugin adds a very neat feature to Studio allowing you to extract term candidates from your Project, or specific files within a project.“*

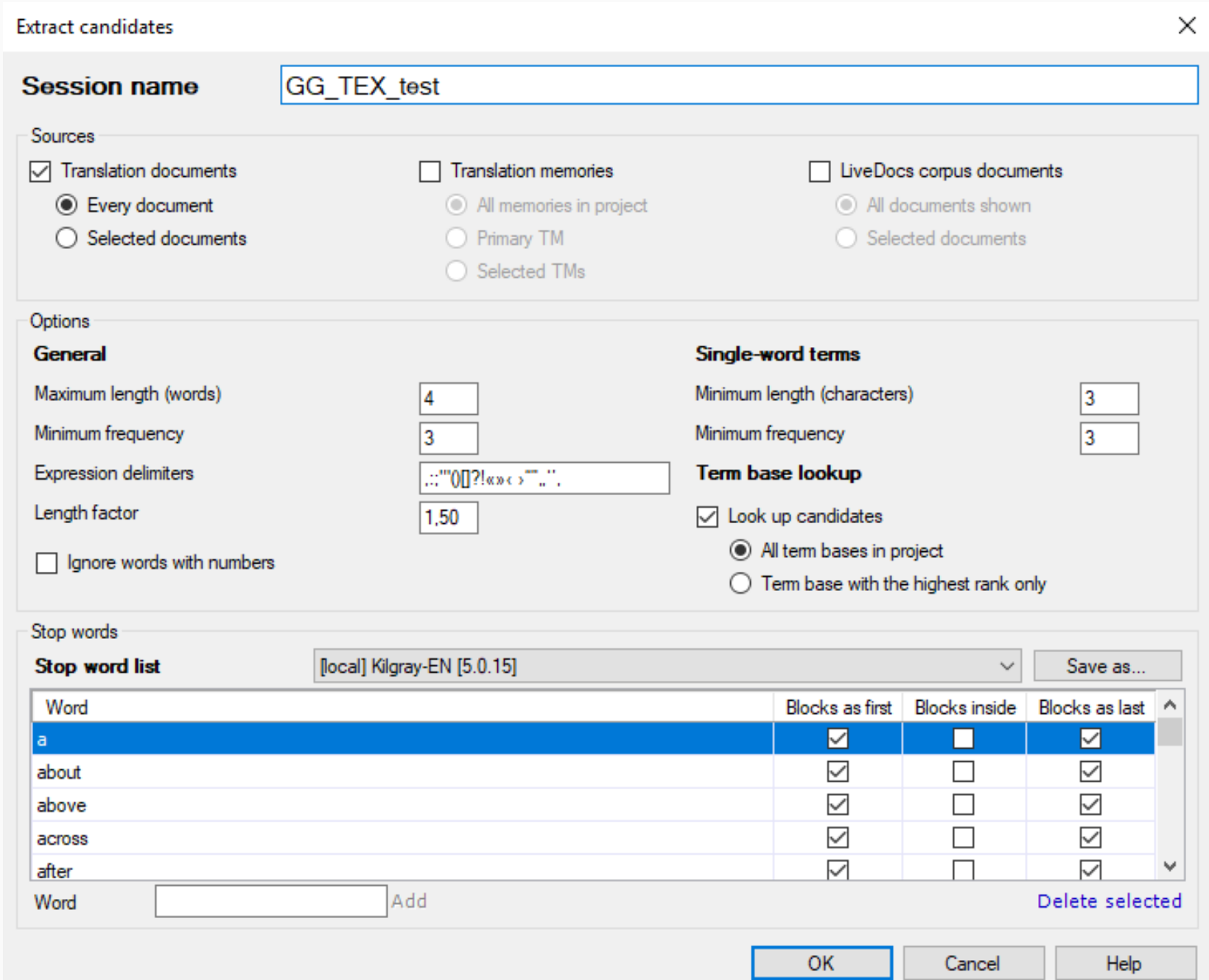
# Termextraktion – Trados (2)



# Termextraktion – memoQ

*„memoQ processes the text and gives you a list of candidates - possible terms. There may be a lot of garbage in the list: You may need to clean it up, filter, and edit it - and confirm “true„ terms before you can add those to a term base. After the extraction runs, memoQ opens the candidate list editor where you can do all this.“*

# Termextraktion – memoQ (2)



# Termextraktion – TermStar (Star)

*„Gleichzeitig unterstützen MindReader und Transit die Terminologiearbeit, z.B. mit der komfortablen Terminologieextraktion aus dem Übersetzungsprojekt oder dynamisch generierten Verwendungsbeispielen aus dem Translation Memory.“*

*„Mit dem Translation-Memory-System Transit NXT kann der Nutzer auch Terminologieextraktion durchführen. Möglich ist die einsprachige Terminologieextraktion aus einem Ausgangstext.“*

# Termextraktion – Across

*„Consistent terminology is a key precondition for high-quality translations. For this purpose, apart from the terminology management crossTerm, Across offers the possibility of extracting potential terms (term candidates) and subsequently determining and translating the actual terms before starting to translate the source document.“*

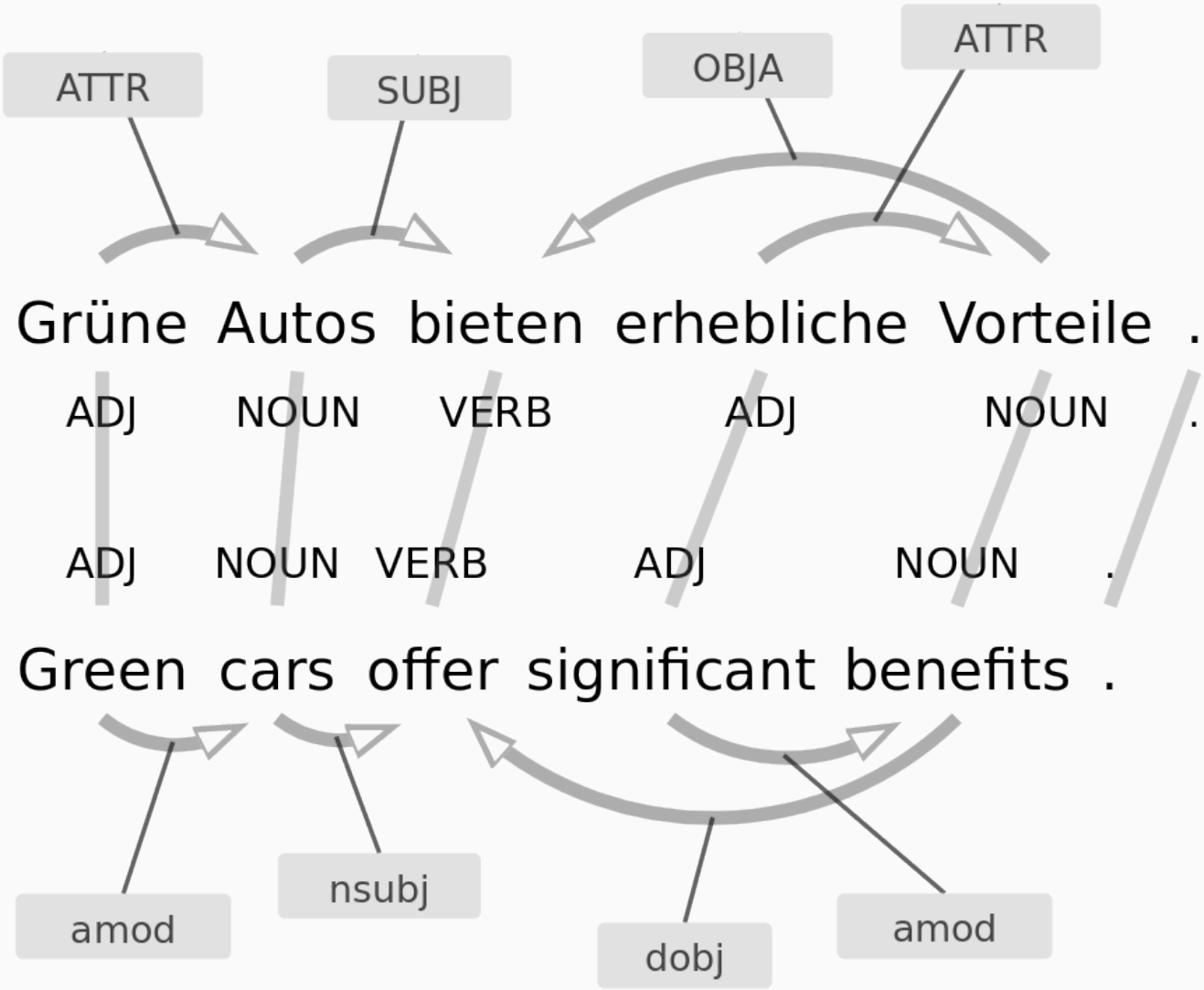
# Zusammenfassung

- Terme werden meist einsprachig ermittelt
- Die Methoden werden nicht offengelegt («a statistical extraction method»)
- Die Benutzer müssen Termkandidaten manuell einer Termdatenbank hinzufügen
- Alignierung scheint bei keinem der Produkte in Verwendung zu sein

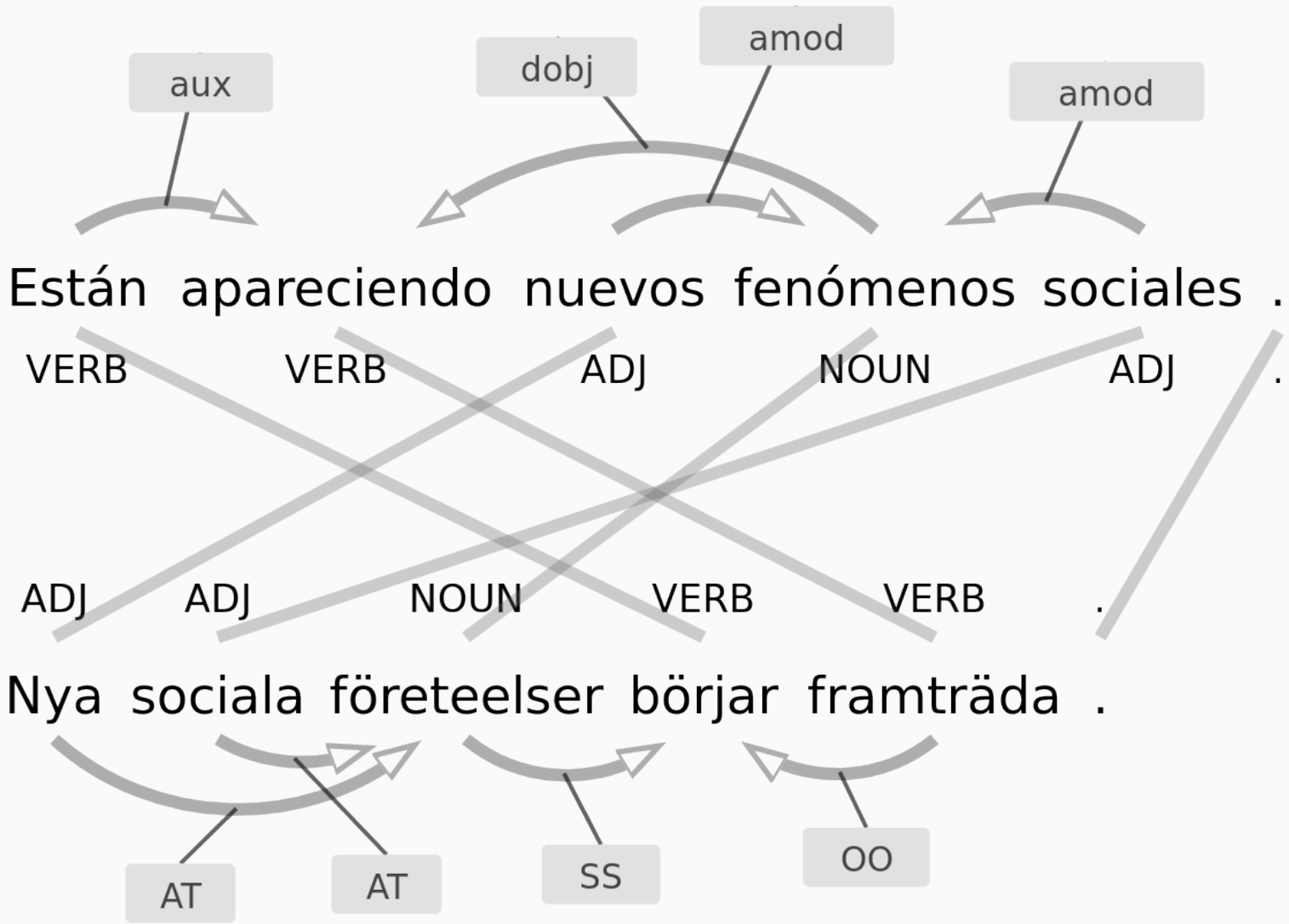
# Einsatz von Alignierung für die Termextraktion

- 1:n-Alignierungen als Indikator («nach wie vor»)
- Hinzuziehen syntaktischer Strukturen (ParZu →)
- Abschätzen der Erwartungshaltung vor oder nach einer Folge von Wörtern («formulaic speech»)
  - wie kann «Bausch und» fortgesetzt werden?
  - was kann vor «oder lebendig» stehen?
  - Beispiel: Binomiale Adverbien →

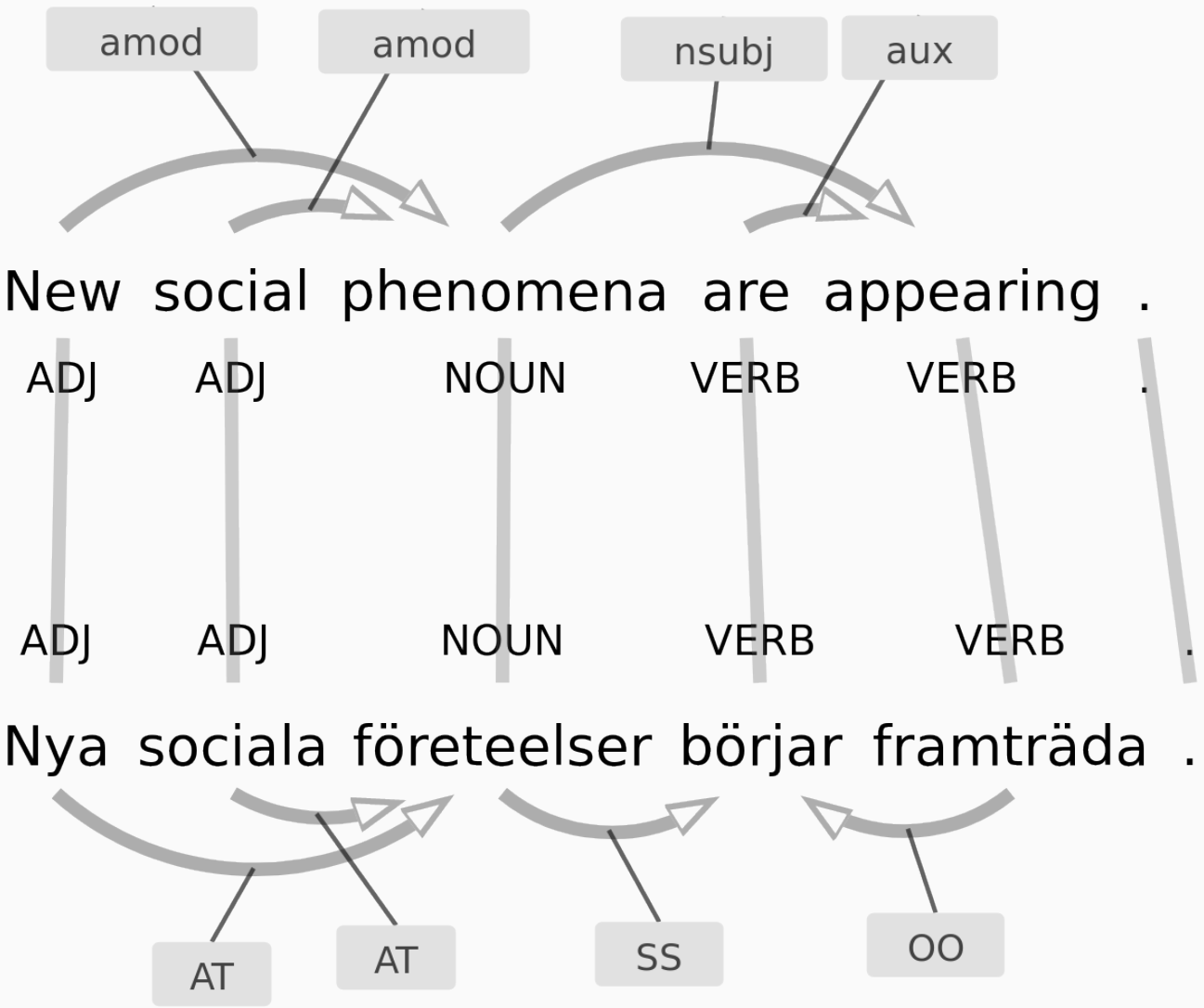
# Alignierung & Syntax (1)



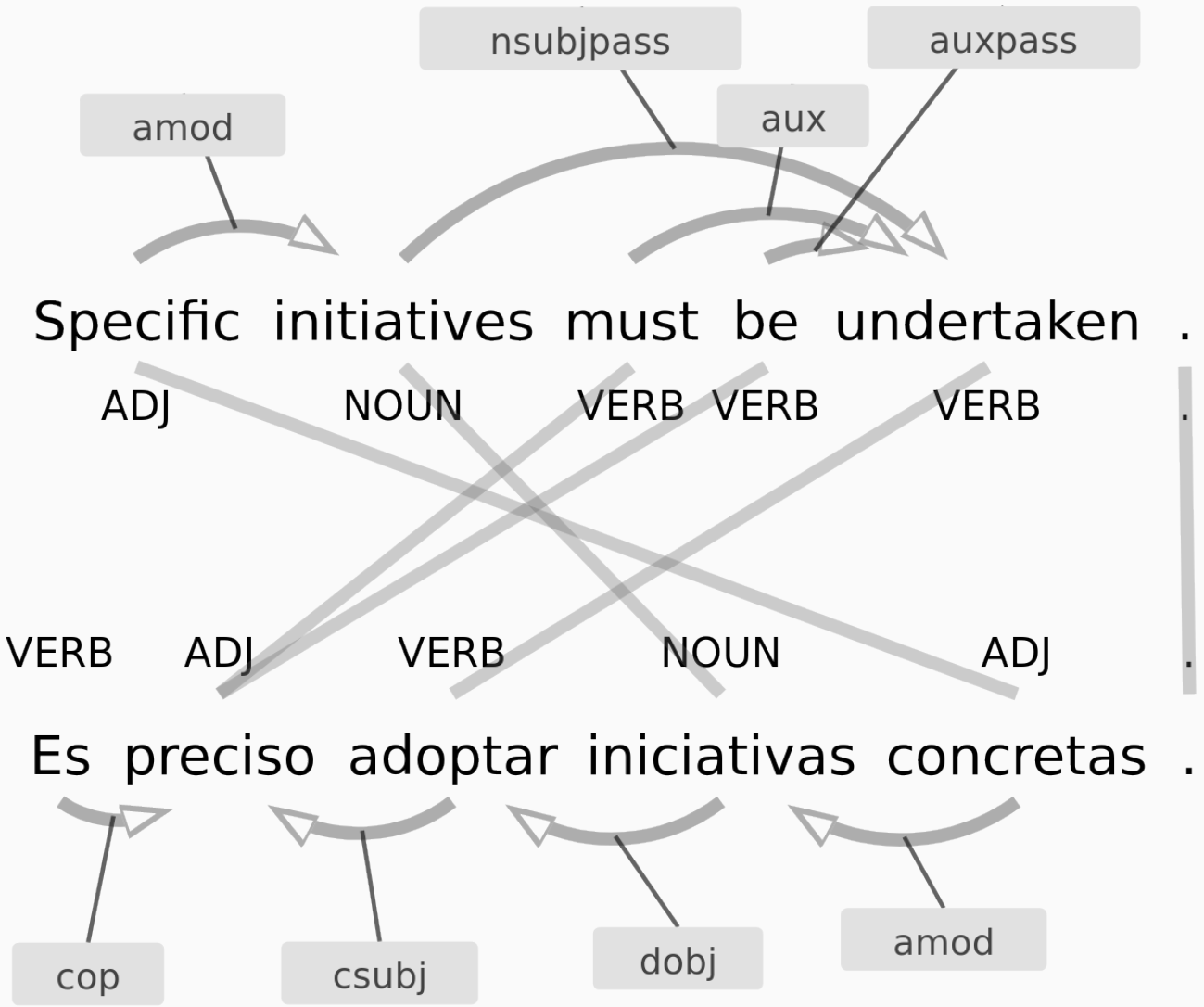
# Alignierung & Syntax (2)



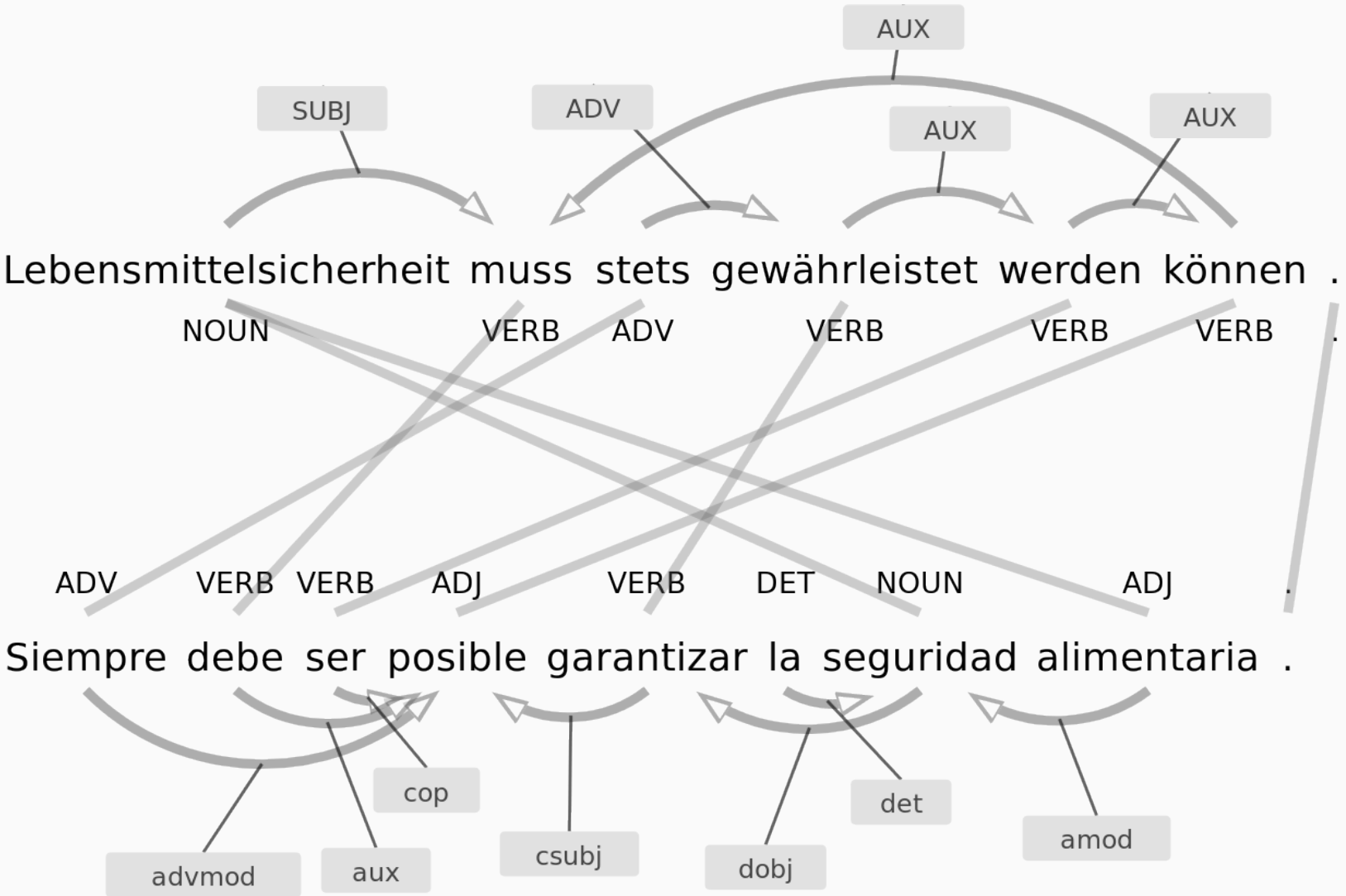
# Alignierung & Syntax (3)



# Alignierung & Syntax (4)



# Alignierung & Syntax (5)



# Sprachenlernen mithilfe von Übersetzungen

# Parallele Korpora zum Sprachenlernen

- authentische Texte
- (oft) professionelle Übersetzungen
- für unterschiedliche Domänen, Texttypen, Register verfügbar
- auch transkribierte Redebeiträge (Parlamentsdebatten, Untertitel) möglich
- Korpusabfragetools für Linguisten wenig geeignet für Lerner

# Date-driven learning

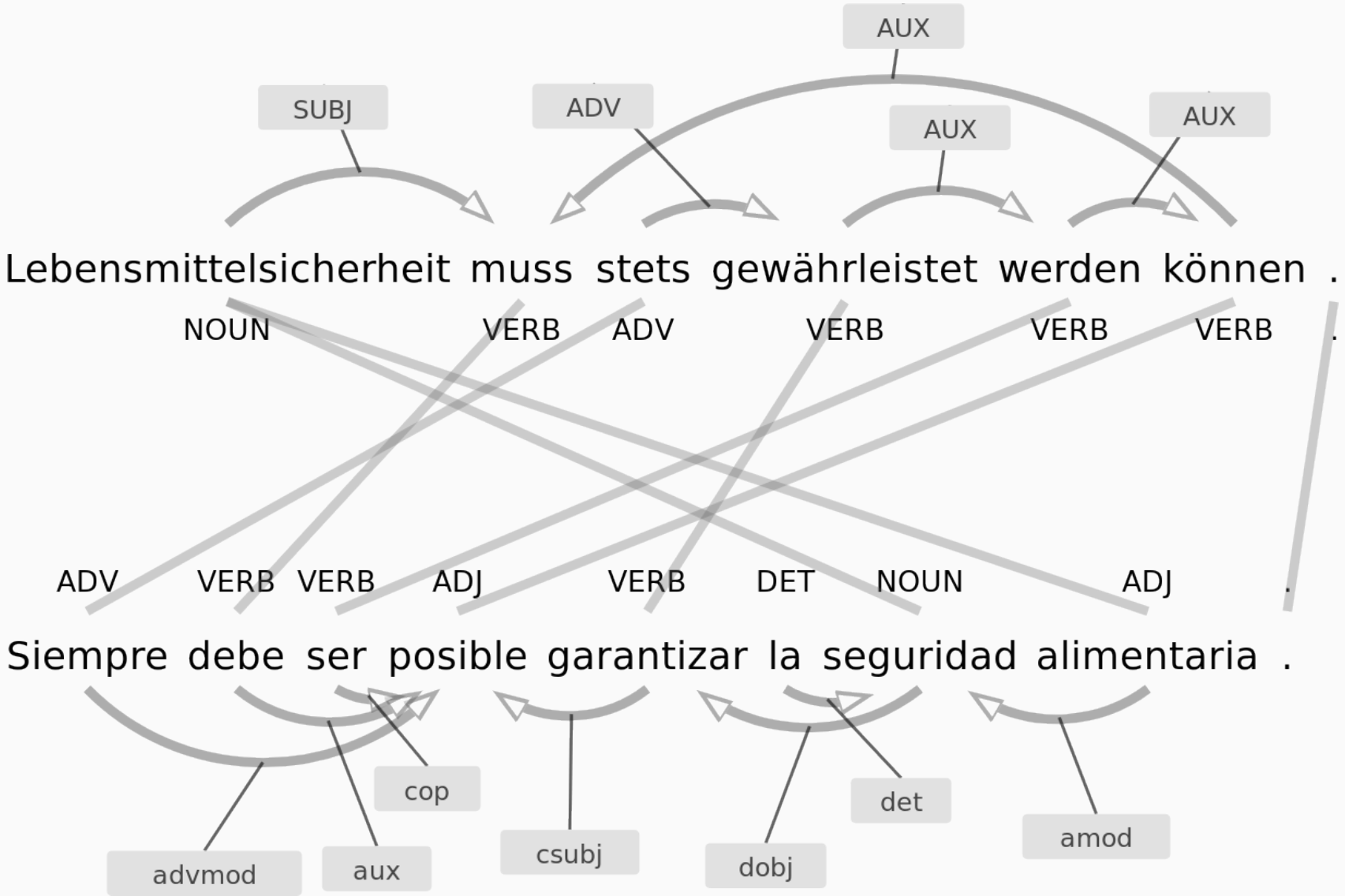
- Lerner explorieren die Zielsprache anhand realer Beispiele (Sätze)
- verschiedene Typen von Parallelität für verschiedene Lernerlevel
- ‹Good Learner Examples› (in Anlehnung an ‹Good Dictionary Examples› (GDEX))

# Parallelität (1)

I have also sent a message to the Spanish Government .

También he comunicado a las autoridades españolas .

# Parallelität (2)



# PaCLE

(Parallel Corpora for Language Learning Exercises)

Korpussuchwerkzeug, zum Auffinden paralleler Sätze

einziges Korpus zur Zeit: OpenSubtitles

für Sprachlehrer und -lerner

Nutzer können Satzpaare bewerten und ändern (Crowdsourcing)

einfache Suche mittels regulären Ausdrücken

komplexe Suche über PoS-Tags, syntaktische Beziehungen und Alignierungen von Token und Phrasen (future)

visuelle Übersicht über die gefundenen Varianten (future)

automatische Generierung von Übungen (future)

# Übung

siehe Handout

med tiden

kort och gott

över förväntan

öga mot öga

huvud på ett fat

med flit

sakta men säkert