

## **Textkorpora als verlässliche Ressource für Übersetzerinnen und Übersetzer**

### **Was ist ein Korpus? (vgl. Bennet 2010, Thomas 2016)**

- Sammlung authentischer Texte – Textauswahl auf Grundlage vorab definierter Kriterien, abhängig von Korpusstyp/Forschungsfrage/Zweck
- Sprache im Gebrauch
- Sprachliche Einheiten im Kontext
- Kann in elektronischer Form mithilfe bestimmter Software (Konkordanzprogramm) durchsucht werden.
- Z.B. Frequenzlisten (Einzelwörter und Mehrworteinheiten), Konkordanzen, verschiedene Suchfunktionen für gezielte Suche wie Kollokationssuche, Word Sketch, Thesaurus

### **Themen- und Anwendungsbereiche (vgl. Bennet 2010)**

- Welche (grammatischen/lexikalischen) Muster liegen der Sprache zugrunde?
- Wie unterscheiden sich diese Muster (Sprachvarietäten, Register)?
- Phraseologie
- Lexikogrammatik
- Übersetzen
- Fachsprachen
- Lexikographie
- Fremdsprachenunterricht

### **Korpusstypen**

- Allgemeine Korpora (Referenz- und Monitor Korpora)
- Spezialisierte Korpora (z.B. historische Korpora, Dialektkorpora, Lerner Korpora, fachspezifische Korpora...)

## Verfügbarkeit

### Webbasierte Korpora und Korpusanalysetools

- <https://corpus.byu.edu/> – freier Zugang zu ca. 15 englisch-, spanisch- und portugiesischsprachigen Korpora (limitierte Anzahl von Suchanfragen und Konkordanzzeilen, gebührenpflichtiges Upgrade erhältlich)
- <https://www.sketchengine.eu/> (Beispiel aus Workshop) – ca. 500 Korpora in ca. 90 verschiedenen Sprachen (gebührenpflichtig, ca. € 100 jährlich)

### Konkordanzprogramme zum Download (zum Durchsuchen eigener Korpora oder vorgefertigter Korpora, die ebenfalls als Download verfügbar sind)

- Liste verfügbarer Korpustools (Auswahl):  
[http://martinweisser.org/corpora\\_site/concordancers.html](http://martinweisser.org/corpora_site/concordancers.html)
- Z.B.: <http://www.laurenceanthony.net/software.html> (AntConc – Beispiel aus Workshop),  
<http://corpora.lancs.ac.uk/lancsbox/>, <http://www.textworld.com/scp/>,  
<http://neon.niederlandistik.fu-berlin.de/en/textstat/>
- Liste verfügbarer Textkorpora (Auswahl):  
[https://en.wikipedia.org/wiki/List\\_of\\_text\\_corpora](https://en.wikipedia.org/wiki/List_of_text_corpora)
- Z.B.: <http://opus.nlpl.eu/>, <http://www.statmt.org/europarl/>, <http://ota.ox.ac.uk/desc/2554>,  
<https://linguatools.org/tools/corpora/wikipedia-comparable-corpora/>,  
<http://rali.iro.umontreal.ca/rali/?q=en/node/71>
- Eigene Korpora können aus (eigenen) Dateien oder Internettexten erstellt werden.
- Programm zum Herunterladen von Internetseiten als Basis für eigene Korpora (in Sketch Engine bereits integriert): <https://bootcat.dipintra.it/>

## **Korpora als Ressource für Übersetzerinnen und Übersetzer (vgl. <https://www.sketchengine.eu/user-guide/translators-term-extraction/>)**

- Konkordanzen und (Bilingual) Word Sketch zum Finden und Überprüfen von typischen/authentischen allgemein- und fachsprachlichen Formulierungen
- Übersetzungskorpora (auch *Parallelkorpora*) – Ausgangstexte mit Übersetzungen (aligniert)
- Vergleichskorpora (auch *vergleichbare Korpora*) – vergleichbare Originaltexte in Ausgangs- und Zielsprache
- Terminologieextraktion – Vergleich fachsprachlicher Korpora mit allgemeinsprachlichen Korpora (Einzelwörter und Mehrworteinheiten)

## **Sketch Engine – Suchfunktionen (vgl. <https://www.sketchengine.eu/user-guide/user-manual/>)**

- **Simple:** alle Formen des lemmatisierten Suchbegriffs oder abgewandelte Form, Groß-/Kleinschreibung wird nicht berücksichtigt
- **Lemma:** alle Formen des Suchbegriffs (muss als Grundform eingegeben werden), Groß-/Kleinschreibung wird berücksichtigt, Wortart kann definiert werden
- **Phrase:** Mehrworteinheiten als Suchbegriff (exakte Zeichenfolge)
- **Word:** Einzelwörter als Suchbegriff (exakte Zeichenfolge)
- **Filter context:** Elemente links und rechts vom Suchbegriff können genauer bestimmt werden.
- **Collocation:** Liste von Wörtern, die vor und/oder nach dem Suchwort erscheinen (nach Erstellen der Konkordanz)
- **Word Sketch:** Zusammenfassung der grammatischen und kollokationellen Eigenschaften eines Suchwortes
- **Word Sketch Difference:** Vergleich von zwei Word Sketches (z.B. zum Vergleichen/Abgrenzen ähnlicher Wörter)

## **AntConc – Suchfunktionen (vgl. <http://www.laurenceanthony.net/software.html>)**

- Z.B.: Konkordanzen, Kollokationen, Terminologieextraktion (Funktion „Keyword List“), Wort- und Mehrwortlisten
- Zahlreiche zusätzliche Komponenten als Ergänzung (z.B. Textsammlung zur Erstellung unterschiedlicher fachspezifischer Korpora, Tool zur Analyse von alignierten Parallelkorpora etc.)

## **Quellen**

Anthony, Laurence (2018) - AntConc: <http://www.laurenceanthony.net/software.html>

Baroni, Marco/Kilgarriff, Adam/Pomikálek, Jan/Rychlý, Pavel (2006) “WebBootCaT: instant domain-specific corpora to support human translators”, in: *Proceedings of EAMT. 11th Annual Conference of the European Association for Machine Translation*. Norway, 247–252.

Bennett, Gena R. (2010) “An introduction to corpus linguistics”, in: Bennet (2010) *Using Corpora in the Language Learning Classroom. Corpus Linguistics for Teachers*. Michigan: University of Michigan Press, 1-18.

Sketch Engine user guide: <https://www.sketchengine.eu/user-guide/>,  
<https://www.sketchengine.eu/user-guide/user-manual/>

Sketch Engine for translators: <https://www.sketchengine.eu/user-guide/translators-term-extraction/>

Thomas, James (2016) *Discovering English with Sketch Engine*. VERSATILE.