

# **Embeddings als Methode zur Analyse vom politischen Diskurs des Schweizer Parlaments**

IDPA in den Fächern Mathematik und Geschichte / Politik

Lukas Tobler

Wasserwerksgasse 14, 3011 Bern

Ausrichtung TAL

BMTAL-18M-S3-BE-Mo-0223

Abgabe: 03.02.2024, Bern

Betreut durch: Stefan Brenken

# Grobkonzept

## Problemstellung

Die Gesellschaft und Politik sind durchgehend in Bewegung, woraus sich Gruppierungen bilden, welche sich teils nur leicht, teils jedoch auch stark voneinander unterscheiden.

Die Sprache der verschiedenen Entwicklungen und Bewegungen zu analysieren, kann daher interessante und wertvolle Erkenntnisse liefern.

Eine Methodik zur Analyse des Sprachgebrauches sind Word-Embeddings. Diese Methodik kann helfen Unterschiede in den Bedeutungen von Wörtern in verschiedenen Korpora zu erkennen, damit diese anschliessend interpretiert werden können.

Diese Arbeit widmet sich ebendieser Methodik und soll die Fragestellung beantworten, „Wie können Word Embeddings genutzt werden, um parlamentarische Texte zu analysieren?“.

## Inhaltliche Abgrenzungen

Da Word-Embeddings nicht nur in der Sprachanalyse benutzt werden, sondern über enorm vielseitige Anwendungszwecke verfügen, ist eine Abgrenzung besonders wichtig.

Heutzutage werden Word-Embeddings meist mithilfe von künstlichen neuronalen Netzwerken generiert. Da diese Art Word-Embeddings zu generieren, zu komplex wäre und den Rahmen dieser Arbeit sprengen würde, wird der Fokus auf die simple und etwas ältere Methodik Word2Vec gelegt.

Um ein besseres Verständnis aufzubauen, soll eine Analyse der parlamentarischen Debatten und Geschäfte durchgeführt werden. Dabei werden jedoch nur die beiden grössten Parteien SP und SVP berücksichtigt.

## Nebenfach

Im Zuge des Grobkonzeptes wurde klar, dass die Frage nach den multilinguistischen Herausforderungen von Word-Embeddings zu komplex gewesen wäre. Zusätzlich ist die verfügbare Studienlage dürftig.

Daher wurde das Nebenfach nachträglich zu Geschichte / Politik geändert.

Somit sind mehr Studien und andere Quellen verfügbar. Ausserdem erlaubt dies eine Analyse der parlamentarischen Texte im Rahmen eines Anwendungsbeispiel.

## Inhaltliche Schwerpunkte

### Mathematische Funktionsweise von Embeddings

In diesem Unterkapitel soll das grundlegende mathematische Prinzip vom Erstellen und anschließende Auswerten von Embeddings erarbeitet und nachvollziehbar dargelegt werden. In diesem Kapitel soll es nicht um eine komplexe Anwendung dieser Technik gehen, sondern lediglich ein Verständnis aufgebaut werden, damit diese Technik im folgenden Teil angewandt werden kann.

### Anwendungsbeispiel

Dieses Unterkapitel soll an einem Anwendungsbeispiel aufzeigen, wie Word-Embeddings generiert und ausgewertet werden können. Ausserdem sollen Grenzen des Word2Vec-Algorithmus aufgezeigt werden.

Beim Anwendungsbeispiel soll es sich um eine Textanalyse von parlamentarischen Debatten und Geschäften handeln. Hierzu werden für die beiden Pol-Parteien SP und SVP jeweils zwei Thesen ausgestellt, welche mithilfe der Word-Embeddings be- oder widerlegt werden sollen.

### Vorgehensweise und methodischer Ansatz

Datum	Aufgabe
23.12.2023	Grobkonzept Rohfassung fertig
12.2023-01.2024	Literaturanalyse
12.2023-01.2024	Einarbeitung in mathematisches Konzept
24.01.2024	Erstes Gespräch mit Expertin (Sophia Schlosser)
31.01.2024	Finalisierung Grobkonzept
03.02.2024	Abgabe Grobkonzept
11.02.2024	Einarbeitung von Rückmeldungen zum Grobkonzept
11.02.2024	Datenaufbereitung für Textanalyse Fertig
18.02.2024	Embeddings für Textanalyse erstellt
29.02.2024	Analyse der erstellten Embeddings
03.2024	Schreiben
07.04.2024	Abgabe zur Kontrolle
20.04.2024	Abgabe IDPA

## Methodik

Gerade bei der Textanalyse wird eine gute Methodik ausschlaggebend sein. Das Vorgehen soll ungefähr wie folgt aussehen:

1. Definieren der zu prüfenden Thesen (Selbstbeschreibung) pro Partei mithilfe einem Topic-Model der Parteiprogramme
2. Definieren von Schlüsselwörtern pro These
3. Analyse der parlamentarischen Debatten und Geschäften mithilfe von Topic-Modelling und Distanzmassen (zb. Cosinus-Distanz)
4. Vergleich der beiden Korpora (Parteiprogramm / Parlamentarische Texte)

## Literaturverzeichnis

**Bubenhof, Noah, Calleri, Selena und Dreesen, Philipp. 2019.** Politisierung in rechtspopulistischen Medien : Wortschatzanalyse und Word Embeddings. *Osnabrücker Beiträge zur Sprachtheorie*. [Online] 2019.  
<https://digitalcollection.zhaw.ch/handle/11475/18403>.

**Das Schweizer Parlament.** Amtliches Bulletin. [Online]  
<https://www.parlament.ch/de/ratsbetrieb/suche-amtliches-bulletin>.

—. Sessionen. [Online] <https://www.parlament.ch/de/ratsbetrieb/sessionen/fruehere-sessionen>.

**Mikolov, Tomas, et al. 2013.** Efficient Estimation of Word Representations in Vector Space. [Online] 16. Januar 2013. <https://arxiv.org/pdf/1301.3781.pdf>.

**Moody, Christopher Erick. 2016.** Ida2vec: Tools for interpreting natural language. [Online] 2016. <https://github.com/cemoody/Ida2vec?tab=MIT-1-ov-file>.