

# Gedanken zur Modellierung von Zeitverläufen am Beispiel Kurzzeitwörter

Rainer Perkuhn und Cyril Belica, Leibniz-Institut für Deutsche Sprache Mannheim

## Abstract

Wenn wir ein intuitives Verständnis voraussetzen, was mit dem Begriff „Wort“ gemeint ist und was wir darunter verstehen wollen, dass „ein Wort zur Sprache gehört“, lassen sich vermutlich die meisten Wörter in wenige Gruppen einteilen, etwa salopp formuliert: (a) Wörter, die es schon immer gegeben hat, (b) Wörter, die es früher nicht gab und die irgendwann aufgekommen sind, (c) Wörter, die es früher gab, aber heute ausgestorben sind und eventuell (d) Mischformen dieser Typen. Kurzzeitwörter wären in diesem Sinne eine Kombination von (b) und (c), d.h. Wörter, die es zunächst nicht gab, die dann eine Zeit lang zur Sprache gehörten, mittlerweile aber wieder verschwunden sind.

Der äußere Rahmen dieser Charakteristika ist der Beobachtungszeitraum mit dem maximalen Endpunkt des heutigen Betrachtungszeitpunktes. Die Beurteilung, ob ein Wort zur Sprache gehört, sollte die Aspekte unterscheiden, ob das Wort bekannt ist und verstanden wird oder ob es tatsächlich auch aktiv verwendet wird, jeweils bezogen auf einen entsprechenden Querschnitt der Sprachgemeinschaft. Dass ein Wort über einen längeren Zeitraum nicht (mehr) aktiv verwendet wird, aber (immer noch) bekannt ist, ist für zurückliegende Zeitabschnitte im Nachhinein kaum überprüfbar. Bei unseren Studien konzentrieren wir uns auf die Verwendung; für die Betrachtung aus der Gegenwarts Perspektive kann der Nachhalleffekt bei der Interpretation mit berücksichtigt werden.

Als Maß der Verwendung arbeiten wir mit Gebrauchsfrequenzen aus dem Deutschen Referenzkorpus DeReKo, insbesondere mit entsprechend definierten vergleichbaren virtuellen Subkorpora. Der maximale Beobachtungszeitraum ist dadurch vorgegeben. Er wird eventuell durch vorgegebene Zeitabschnitte der Neologismenbeobachtung weiter eingeschränkt. In noch experimentell-explorativen Phasen testen wir verschiedene Modellierungen von Zeitverläufen. Während wir gleichzeitig versuchen, die Ansätze möglichst allgemein und übertragbar zu gestalten, sollen in der Pilotphase in Kooperation mit dem Projekt „Neuer Wortschatz“ insbesondere Kandidaten für Kurzzeitwörter ausgewählt und bewertet werden. Vorrangiges Ziel der Methodik ist es, die unterschiedlichen charakteristischen Zeitabschnitte und deren Übergänge zu erkennen. Um den Einfluss der Messtaktung zu dämpfen, sind gerade bei den Übergängen Toleranzen zu berücksichtigen. Ergänzend werden Durchschnittswerte für dynamische Formen der Taktung ermittelt. Als größte Herausforderung stellt sich aber heraus, geeignete Parameter für Werteintervalle für die verschiedenen Abschnitte zu definieren, die sowohl hinsichtlich *precision* als auch *recall* angemessene Kandidatenmengen ermitteln. Es deutet sich an, dass die Definition unterschiedlicher Kurzzeitwort-Typen eine Verbesserung über angepasste Parameterkonfigurationen mit sich bringen kann, die über eine Rückmeldung zur bisherigen Erkennung und einer daraus abgeleiteten verbesserten Trainingsmenge entwickelt werden sollen.