# Neue Ansätze zur Auswertung von Schreibprozessdaten: Textgeschichten und Satzgeschichten

Cerstin Mahlow
Malgorzata Anna Ulasik

Wir arbeiten derzeit an einem neuen Verfahren zur Analyse von Schreibprozessdaten auf linguistischer Ebene via syntaktisches Parsen. Dazu entwickeln wir ein plattformunabhängiges Open-Source-Werkzeug[1], um Keystroke-Logging-Rohdaten verarbeiten zu können.

Als Kriterium für die Extraktion von Versionen dient die Änderung im Produktionsmodus – d. h. der Wechsel zwischen Schreiben, Löschen von Text oder Einfügen von Text an anderer Stelle im bisher Geschriebenen (Mahlow 2015). Für unser aktuelles Interesse am sprachlich motivierten Editieren filtern wir (potenzielle) Korrekturen von Tipp- oder Rechtschreibfehlern und schließen diese als nicht-essentielles Editieren aus, da diese Korrekturen normalerweise nicht die syntaktische Struktur ändern. Auf diese Weise können wir relevante Versionen auf Textebene erzeugen, die aus Sätzen bestehen, die nicht von einer Revision betroffen sind (also nicht erneut geparst werden müssen), und Sätzen, die in Bearbeitung sind.

Relevante Versionen unterscheiden sich durch die *Transformationssequenz*. Wir verwenden *bursts* (Kaufer, Hayes und Flower 1986; Baaijen, Galbraith und Glopper 2012), um die Änderung im Produktionsmodus zu operationalisieren. Eine Editieroperation ist entweder das Entfernen oder Einfügen einer Sequenz ohne Unterbrechung. Die Transformationssequenz gehört so immer zu einer dieser beiden Editierkategorien. Dabei werden nur Versionen betrachtet, bei denen Sequenzen transformiert werden, die sprachliche Strukturen enthalten: Wörter, Phrasen, Teilsätze oder Sätze.

Die Sequenz der relevanten Versionen eines Satzes ist die *Satzgeschichte*. Ein Satz kann zu verschiedenen Zeiten bearbeitet werden, sodass diese

1. `https://github.com/mulasik/wta`

Abfolge von Versionen zwar linear, aber nicht unbedingt zusammenhängend ist. Siehe Abbildung 1 für eine Satzgeschichte aus einem Pilotexperiment.

*Abbildung 1: Satzgeschichte von* `Das Video als Video zu beschreiben und zu bewerten, fällt mir nicht leicht`*, die Ziffern sind die Versionsnummern, nicht-relevante Versionen sind unterdrückt.*

(43) *Das Video als Video fin*
(45) *Das Video als Video zu beschreiben, fällt mir*
(62) *Das Video als Video zu beschreiben und zu bewerten, fällt mir*
(75) *Das Video als Video zu beschreiben und zu bewerten, fällt mir nicht leicht.*

Abbildung 2 zeigt die Visualisierung der *Textgeschichte* einer Schreibsitzung, hier sind auch Änderungen im bereits geschriebenen Text sichtbar, nicht nur am Textende.

Für die weitere Entwicklung ist vorgesehen, das Werkzeug hinsichtlich dreier Parameter anpassbar zu machen: Änderung des Produktionsmodus, Pausen und Relevanz der Bearbeitung. Dies wird es ermöglichen, den Begriff der bursts und damit den Umfang der Transformationssequenzen anzupassen und so gezielt linguistische Aspekte des Schreibprozesses zu untersuchen.

Generell erlaubt diese neuartige Analysemethode die linguistische Modellierung des Schreibprozesses über kognitive Aspekte hinaus und ist unter anderem die Voraussetzung für die Entwicklung sprachbasierter Editierfunktionen (Mahlow und Piotrowski 2009).

## 1 Literatur


Baaijen, Veerle M., David Galbraith und Kees de Glopper (Juli 2012). „Keystroke Analysis“. In: *Written Communication* 29.3, S. 246–277. DOI: `10.1177/0741088312451108`.

Kaufer, David S., John R. Hayes und Linda Flower (1986). „Composing Written Sentences“. In: *Research in the Teaching of English* 20.2, S. 121–140. URL: `http://www.jstor.org/stable/40171073`.

Mahlow, Cerstin (2015). „A Definition of ”Version” for Text Production Data and Natural Language Document Drafts“. In: *Proceedings of the 3rd International Workshop on (Document) Changes: Modeling, Detection, Storage and Visualization*. DChanges 2015. Lausanne, Switzerland: ACM, S. 27–32. DOI: `10.1145/2881631.2881638`.

Mahlow, Cerstin und Michael Piotrowski (Okt. 2009). „LingURed: Language-Aware Editing Functions Based on NLP Resources“. In: *Proceedings of the International Multiconference on Computer Science and Information Technology*. Bd. 4. Polish Information Processing Society. Mragowo, Poland, S. 243–250. URL: `http://www.proceedings2009.imcsit.org/pliks/101.pdf`.

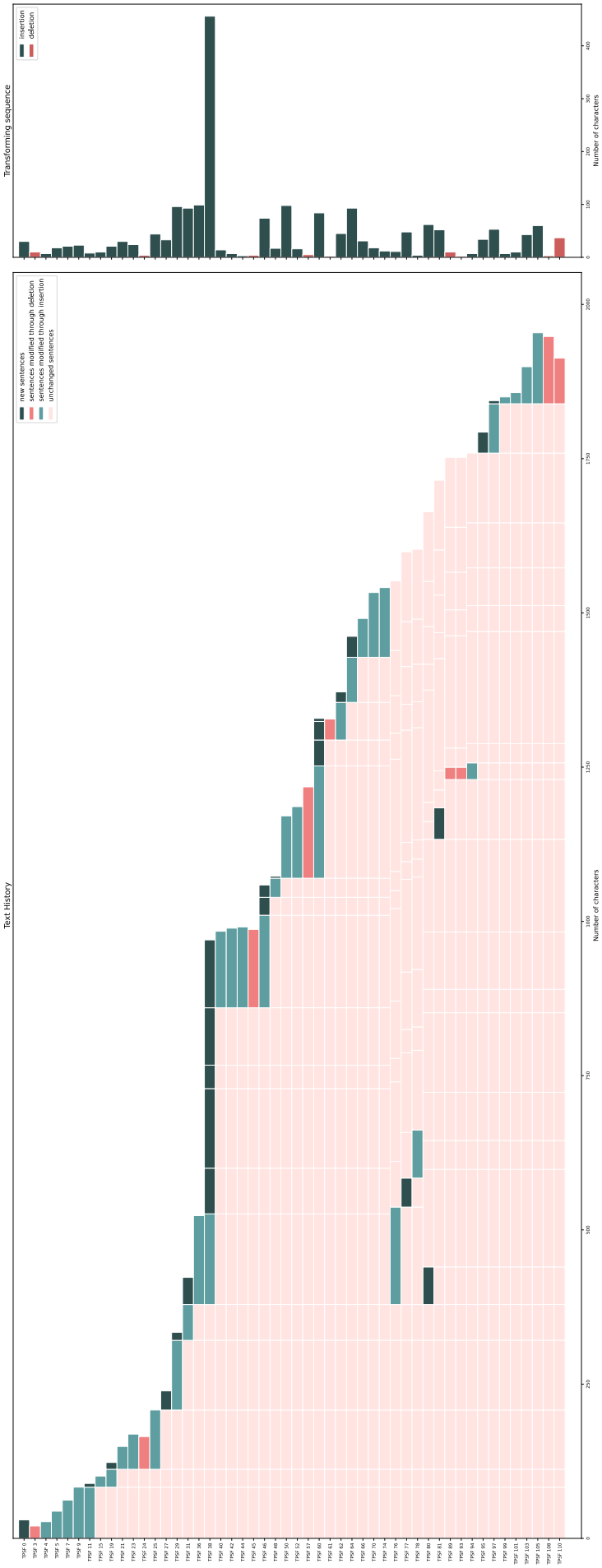


*Abbildung 2: Textgeschichte für kurze Schreibsitzung*