

TypeShift: A User Interface for Visualizing the Typing Production Process

Adam Goodkind

Northwestern University

a.goodkind@u.northwestern.edu

Abstract

TypeShift is a tool for visualizing linguistic patterns in the timing of typing production. Language production is a complex process which draws on linguistic, cognitive and motor skills. By visualizing holistic trends in the typing process, TypeShift aims to elucidate the often noisy information signals that are used to represent typing patterns, both at the word-level and character-level. It accomplishes this by enabling a researcher to compare and contrast specific linguistic phenomena, and compare an individual typing session to multiple group averages. Finally, although TypeShift was originally designed for typing data, it can easily be adapted to accommodate speech data, as well.¹

1 Introduction

The task of “visualizing language production” is both broad and difficult to implement conclusively. Common visualizations relating to language production include dendrograms, word clouds and frequency counts. These summary visualizations, however, only provide a static snapshot of language. They do not capture the dynamics elements that go into the process of language production, replete with its hesitations, inconsistencies and repairs.

Capturing the dynamics of language production is important because information is transmitted in language not only via the words chosen (e.g. [Chenoweth and Hayes, 2001](#)), but also by tone of voice, pauses and slips of the tongue (e.g. [Bosker et al., 2013](#); [Goldrick et al., 2016](#)). Therefore, in order to accurately study language production, a researcher not only needs a record of the words that were produced, but also a record of *how* they were produced.

¹A web demo is available at <https://angoodkind.shinyapps.io/TypeShift/>. The source code can be accessed at <https://github.com/angoodkind/TypeShift>.

The TypeShift tool aims to visualize the dynamics of language produced via a computer keyboard. Compared to speech production data, keyboard typing production data is relatively easy to capture ([Cohen Priva, 2010](#)). A simple keylogger can record when a key is depressed and then released, making it relatively easy to measure when a word begins and when it ends. Similar timing metrics are much more laborious and ambiguous with speech production data, as it is more difficult to determine the boundaries of phonemes in a speech stream. Nonetheless, the TypeShift tool can be readily adapted to speech data, if sufficient data is collected, requiring only minor alterations to the tool.

1.1 Overall Motivation

The importance of a system such as TypeShift lies in answering the multifaceted question, “How can displaying temporal and dynamic information about language production elucidate linguistic processing?” This question is broad and difficult to answer concisely. However, certain aspects of a typing session can shed light on the question, and provide partial answers. Below is a list of examples, with elements of a typing session that TypeShift can help to shed light on:

- Is this particular typing instance fast or slow, both as compared to other typing instances from the same user, as well as compared to different subsets of the general population?
- Does a typist produce certain linguistic structures, e.g. noun phrases or function words, in a fashion distinct from other linguistic structures, e.g. verb phrases or content words? Does a certain structure result in slower production or more revisions?
- Does this typist produce language at a consistent rate, or do certain linguistic elements result in heavy revisions, or bursts of quick

typing interspersed with slow typing?

Answers to these questions can be useful for investigating psycholinguistic questions, as well as for tasks in related areas such as machine learning. For example, typing patterns have been shown to be a proxy for shallow syntactic parsing (Plank, 2016). However, in creating a training set, it is important to know if the typing data being used is typical or atypical, both for that typist and compared to others.

Finally, by being able to compare different trends, as well as investigate only certain linguistic elements and see the details of just one token, the TypeShift tool permits for targeted comparisons that are revealed through details-on-demand, an important principle of data visualization (Shneiderman, 1996). Using this capability, a cognitive scientist could use the tool to visualize how different linguistic structures affect processing, vis-a-vis pauses and production rates.

We begin by explaining the value of capturing keystroke information, and then outline how TypeShift builds upon and improves over previous methods of visualizing the typing process. Next, we briefly describe the large dataset around which the tool was originally built (although it is now easily extensible). We then describe the system design and its research motivations, using elucidating examples.

2 Related Work

2.1 Why keystrokes?

Keystroke dynamics is the science of capturing the detailed timing information of a typist’s keystrokes (Moskovitch et al., 2009). This timing information can vary from a parameter as coarse as the overall mean typing rate to a metric as fine-grained as the pause time between two particular keys. The details of a typing session can help predict many personal properties, from overall identity authentication (Ali et al., 2015, *inter alia*), to emotion and cognition prediction (Epp et al., 2011; Brizan et al., 2015). Further, it can convey social information and influence impression-formation in dyads and groups (e.g. Kalman et al., 2013).

Pausing within language production can also be especially informative. A pause can signal everything from increased cognitive effort, due to performing lexical retrieval (Erman, 2007), to a more intense focusing on the task at hand (Schilperoord, 1996).

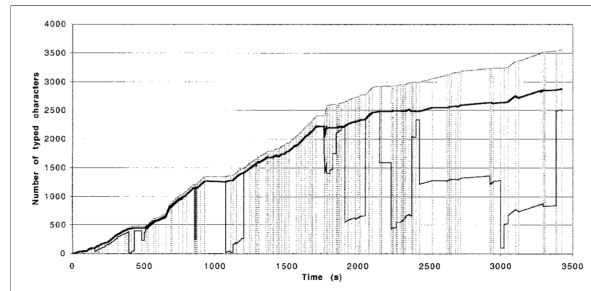


Figure 1: An example of an LS Graph, a linear representation of the typing production process. The lighter top line represents the total number of keystrokes produced. The black line in the middle represents the number of characters appearing on the page, i.e. the top line minus revisions. The lowest line represents where the user is working at that time. Finally, the vertical lines represent where revisions have occurred.

2.2 Previous visualization work

Given that the dynamics of typing can provide such a large amount of information, many researchers have devised methods to visualize a typing session. For a recent overview, see Bécotte et al. (2017). Many of the visualization techniques proposed for the typing process are limited to the “what” of typing, i.e. which keys were pressed and in what order. Some visualizations do focus on the rate of typing, which is reflected as a line chart where a steeper slope reflects a more rapid typing rate. This is illustrated in Fig. 1, which is an example of an LS Graph (Lindgren and Sullivan, 2002). However, this visualization only provides overall continuous lines, and lacks a discrete representation of separate words.

One visualization technique that *does* represent word tokens can be seen in Fig. 2 (Caporossi and Leblay, 2011). While the discrete nodes show word tokens as well as the relationship between the order in which tokens were produced, this visualization fails to capture the temporal dynamics or varying rates of the LS Graph. In fact, the x,y coordinates of this graph are completely arbitrary, and are arranged solely to present a compact visualization.

The visualization techniques described in this paper attempt to synthesize the best aspects of the visualization techniques mentioned above. It captures both the continuous, dynamics rate of language production, and highlights the discrete nature of word tokens.

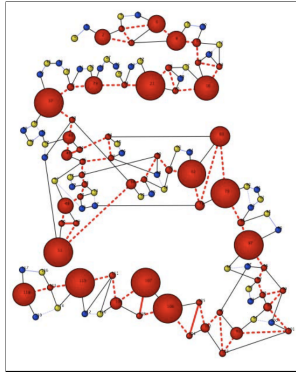


Figure 2: Graph representation of the online writing process

3 A New Keystroke Dataset

This system was originally designed for use with a large, open-source dataset of typing data.² Because of the scale and complexity of the data collected, the dataset can be utilized to answer a number of research questions, on its own. Nonetheless, though, the flexibility of the TypeShift system allows researchers to import their own keystroke or speech data, so long as certain parameters are included.

The typing data was collected from 1,013 university students, seated in front of a desktop computer. Each participant typed responses to 10–12 question prompts, and each answer had to be at least 300 characters. This resulted in approximately 1 million words consisting of 10 million keystrokes. Participants included all genders, were both left- and right-handed, were both native and non-native English speakers, and reported a wide variety of school majors and typing/computer experience. Further details are provided in Locklear et al. (2014) and Brizan et al. (2015).

Each prompt was created in order to elicit one of six cognitive tasks: *Remember*, *Understand*, *Apply*, *Analyze*, *Evaluate* or *Create*. This task label was determined by the experimenters, who assessed the cognitive demands of each question as they related to Bloom’s Taxonomy of Learning (Krathwohl and Anderson, 2009). Examples prompts include:

- (Remember) *List the recent movies you’ve seen or books you’ve read.*
- (Evaluate) *Do you think it’s a good idea to raise tuition for students in order to have money to make improvements to the University? Why or why not?*

²The data is available at <http://www2.latech.edu/~mike/TypingForTenData>

4 Design Motivations

4.1 Data Preprocessing

The collection procedures above resulted in an extensive amount of data. In order to understand the data with which we were dealing, the TypeShift tool was designed with the intention of creating meaningful data partitions, to make keystroke streams understandable and usable. To better illustrate the design decisions, we present a number of motivating examples below.

In order to first process the linguistic data, the stream of typing was clustered into word tokens using the Stanford CoreNLP parser (Manning et al., 2014). Each word token was annotated with its part-of-speech and semantic role. The semantic role is of interest because some psycholinguists posit that function words, e.g. *you*, *the*, *it* or *them*, are particularly psychologically informative (Chung and Pennebaker, 2007). This is in contrast to the traditional view that content words, e.g. *red*, *fast*, *smack* or *mug*, contain the most amount of cognitive information. In addition, the start time and end time of each token was recorded, along with the keystrokes that comprise each token.

Typing sessions were not of uniform length, which could make comparisons difficult. To allow for typing sessions of different time lengths to be compared, the elapsed time of each typing session was normalized and converted to the proportion of a complete session, i.e. 0.0–1.0. Similarly, the typing rate for each token (# of keystrokes/token time span) was normalized and converted to a z -score, to allow for comparison across typists. By normalizing the typing rate, a visualization can shed light on not only whether a typist is typing rapidly *overall*, but also rapidly *for himself or herself*.

4.2 TypeShift User Interface

The TypeShift tool was designed using the Shiny Web Application Interface for R (Chang et al., 2016). The tool allows a user to select exactly which typing session to visualize, as well as which aspects of a typing session to include in the visualization.

A screenshot of the interface is seen in Fig. 3. The menu on the side provides a number of options:

- Selected User – A list of available typists is presented along with the the typist’s age, gender and native language (L1).
- Typing Session – Within the selected typist,

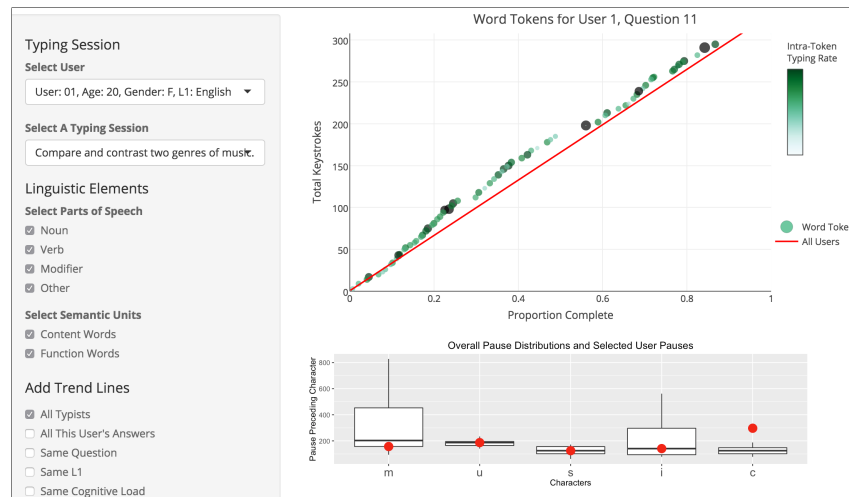


Figure 3: Screenshot of the TypeShift User Interface

a list of the question prompts he or she responded to is presented.

- Parts-of-Speech – The user can choose to display only certain syntactic categories, such as only nouns or only verbs.
- Semantic Units – The user can choose to display only function or content words.
- Trend Lines – The user can compare this typing session to a variety of overall trends:
 - All Typists – The entire dataset
 - All This User's Answers – Every answer from the selected typist
 - Same Question – All responses to the same typing prompt
 - Same L1 – All typing sessions from users with the same native language
 - Same Cognitive Load – All responses to typing prompts that require a similar level of cognitive effort

The main plot charts the total number of keystrokes produced versus the proportion of the typing session that was completed. In this way, the slope of the resulting line (series of points) reflects the typing rate: a steeper slope represents a more rapid typing rate. The color and size of a point represent the typing rate of a particular token, in contrast to the overall typing rate. For example, if a user types a single word faster than the other words, then the point representing that word will be larger and darker than the other points. Finally, a point's border thickness represents the number of times a token was revised. In other words, a token with a thick border represents a word produced with a high number of backspaces or deletes.

If a user wishes to investigate the character-by-

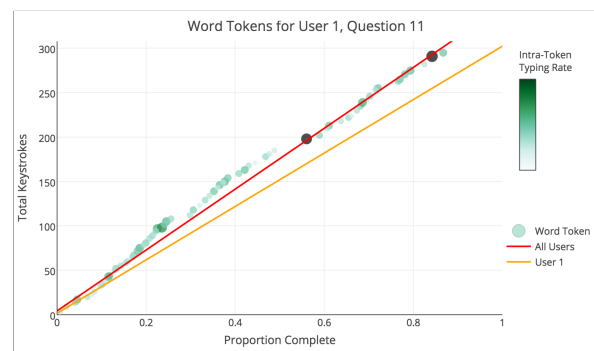


Figure 4: Average typing session for a slow typist

character specifics of a particular token, she can click on any individual word, to see a detailed view of that token's timing (lower-right portion of Fig. 3). The pause time before each character is displayed as a red point for that particular instance. In addition, a box plot for the overall population distribution of pauses before that character is also displayed. In this way, a user can ascertain how a typist produces individual characters of a specific word, relative to other typist's producing the same word.

4.3 Design Choice Motivations

The design motivations of the TypeShift system are best understood through examples. The examples below highlight how TypeShift can bring to light key attributes of a typist or typing session.

Fig. 4 illustrates the importance of comparing a typing session not only to the overall, across-subjects mean, but to the within-subject mean, as well. Compared to the overall population (red line), this typing session would be considered average. However, for this typist, this particular session is actually very rapid, as the timing is above the user

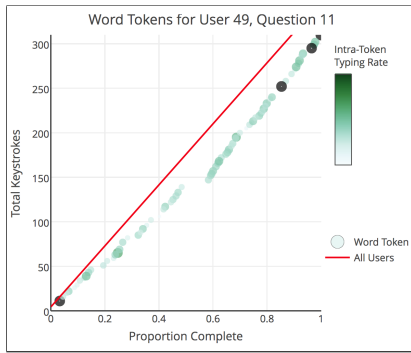


Figure 5: Slow but steadily-paced typist, with typing bursts

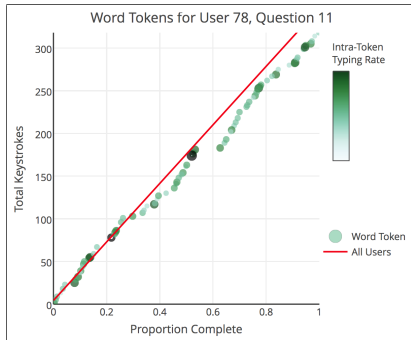


Figure 6: Inconsistent typing rate

mean (orange line). If a researcher were trying to understand this particular typist, this visualization helps show that this is not a typical typing session for this typist.

Fig. 5 illustrates a typist who is slow, but types in bursts at a relatively constant rate. The overall constant rate is illustrated by the uniform slope of the plot, along with similarly sized and colored points. In contrast, Fig. 6 shows a more inconsistent typing rate, with varying slopes and different colored points. The burstiness of the typing session is illustrated by the large gaps in the plot, when typing pauses for a significant period of time. By being able to select only certain linguistic structures, or zoom in on a single word, a researcher could investigate exactly what is *causing* these typing bursts.

Fig. 7 is a non-native English speaker, and looks very different than the previous plots. This typist makes a number of large revisions, as seen by the thick point borders. Likely this is due to less familiarity with English, and hence more spelling or word-choice errors. After the revision, there is often a pause or a change in typing rate. It seems that for this non-native speaker, these errors create greater confusion, and require time to readjust afterwards.

Fig. 8 is a faster-than-average typist, as illustrated by the steeper slope of the user trend line

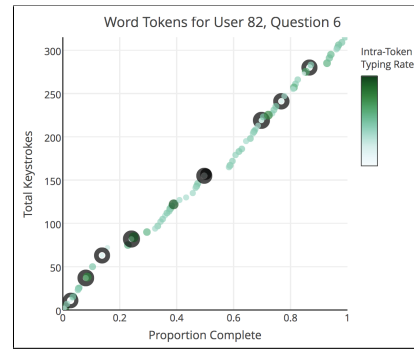


Figure 7: Heavy revising causes changes in typing dynamics

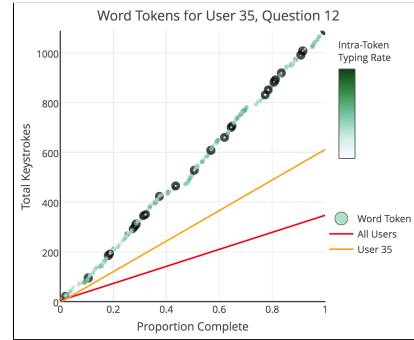


Figure 8: Faster than average typists that stays consistent

(orange), as compared to the overall trend line (red). Further, this particular session is fast-paced, even for this typist. Unlike Fig. 7, where revisions are a catalyst for a change in typing rate, this more proficient typist is not as affected by revising. Rather, the typist maintains a steady rate, despite frequent revisions. This further sheds light on the importance of comparison trend lines, as they allow us to infer more about the typist than could be inferred from typing session data in isolation.

In addition to comparing typing rates, being able to filter for only certain types of word tokens may also shed light on interesting trends and differences.

As can be seen in Figs. 9(a) and 9(b), this user types verbs at a faster rate than she types nouns. This could point to the importance of taking into consideration latent linguistic structure when analyzing a typist's performance. By being able to filter, a user of TypeShift can perform a more direct comparison of these linguistic variables.

Similarly, Figs. 10(a) and 10(b) demonstrate how differently a user types function words versus content words. A comparison of these visualizations could be an indicator of distinct cognitive processes driving distinct production rates of different types of words.

Finally, as seen in Fig. 11, a user of TypeShift

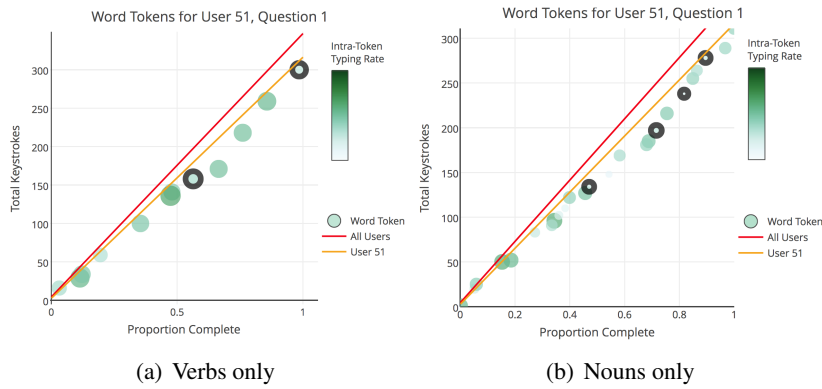


Figure 9: Comparing typing rates for different parts-of-speech

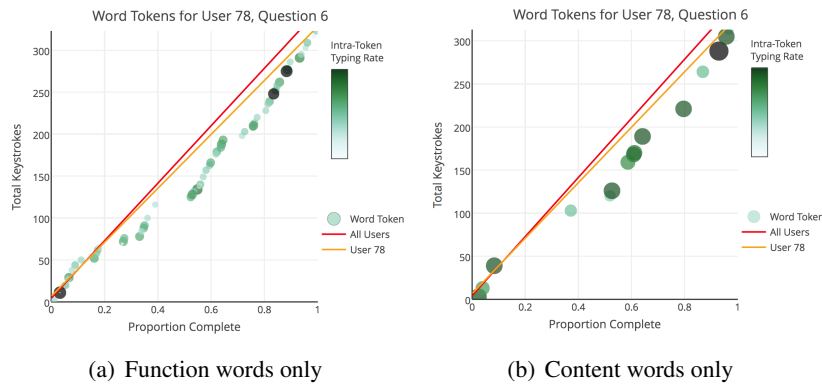


Figure 10: Comparing typing rates for different semantic categories

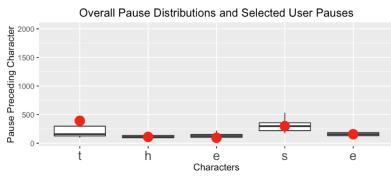


Figure 11: Pre-character pauses for each keystroke in the token “these”

can zoom in to a particular word token. The red points represents the selected typist’s pause durations, while the boxplots represent the overall population distribution of pause durations. This is motivated by the hypothesis that typing production is a hierarchical process: the initial pause is dictated by word retrieval difficulty; once the word is retrieved, subsequent pauses are determined by the difficulty of motor execution (Yamaguchi et al., 2013). We see evidence for this in Fig. 11, where the subject is a native Russian speaker. As can be seen, the pause before the initial character is of a longer duration, possibly because of more difficulty retrieving a word in a non-native language. However, the subsequent pauses are very near the population medians, since these pauses are determined

by motor skills rather than language familiarity.

5 Conclusion

The TypeShift user interface aims to provide a dynamic tool to visualize both the continuous and discrete nature of the language production process. Language production is both a stream of flowing words, as well as a series of separate word tokens. By allowing a user to capture both the holistic process as a single linear progression, as well as highlighting individual characteristics of each particular token, the tool can help a user understand both aspects. Further, by providing comparisons between typing sessions, TypeShift can help illustrate where a typing sessions ranks, both within- and across-subjects.

Ultimately, the goal of understanding language production is to understand how the mind categorizes and processes information. By allowing a user to better visualize and compare typing sessions, more rapid progress can be made towards conceptualizing human information transmission.

References

- Md Liakat Ali, Charles C Tappert, Meikang Qiu, and John V Monaco. 2015. Authentication and identification methods used in keystroke biometric systems. In *2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems*, pages 1424–1429. IEEE.
- Hélène-Sarah Bécotte, Gilles Caporossi, Christophe Leblay, and Alain Hertz. 2017. Writing and rewriting: Keystroke loggings colored numerical visualization. *Observing Writing: Logging Handwriting and Computer Keystrokes*. Leiden, Brill Academic Publishers.
- Hans Rutger Bosker, Anne-France Pinget, Hugo Quené, Ted Sanders, and Nivja H De Jong. 2013. What makes speech sound fluent? the contributions of pauses, speed and repairs. *Language Testing*, 30(2):159–175.
- David Guy Brizan, Adam Goodkind, Patrick Koch, Kiran Balagani, Vir V Phoha, and Andrew Rosenberg. 2015. Utilizing linguistically enhanced keystroke dynamics to predict typist cognition and demographics. *International Journal of Human-Computer Studies*, 82:57–68.
- Gilles Caporossi and Christophe Leblay. 2011. Online writing data representation: a graph theory approach. In *International Symposium on Intelligent Data Analysis*, pages 80–89. Springer.
- Winston Chang, Joe Cheng, JJ Allaire, Yihui Xie, and Jonathan McPherson. 2016. shiny: Web application framework for r. r package version 0.13. 2. URL: <http://CRAN.R-project.org/package=shiny>.
- N Ann Chenoweth and John R Hayes. 2001. Fluency in writing: Generating text in l1 and l2. *Written communication*, 18(1):80–98.
- Cindy Chung and James W Pennebaker. 2007. The psychological functions of function words. *Social communication*, 1:343–359.
- Uriel Cohen Priva. 2010. Constructing typing-time corpora: A new way to answer old questions. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 32.
- Clayton Epp, Michael Lippold, and Regan L Mandryk. 2011. Identifying emotional states using keystroke dynamics. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 715–724.
- Britt Erman. 2007. Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics*, 12(1):25–53.
- Matthew Goldrick, Joseph Keshet, Erin Gustafson, Jordana Heller, and Jeremy Needle. 2016. Automatic analysis of slips of the tongue: Insights into the cognitive architecture of speech production. *Cognition*, 149:31–39.
- Yoram M Kalman, Lauren E Scissors, Alastair J Gill, and Darren Gergle. 2013. Online chronemics convey social information. *Computers in Human Behavior*, 29(3):1260–1269.
- David R Krathwohl and Lorin W Anderson. 2009. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Eva Lindgren and Kirk PH Sullivan. 2002. The Is graph: A methodology for visualizing writing revision. *Language Learning*, 52(3):565–595.
- Hilbert Locklear, Sathya Govindarajan, Zdeňka Sitová, Adam Goodkind, David Guy Brizan, Andrew Rosenberg, Vir V Phoha, Paolo Gasti, and Kiran S Balagani. 2014. Continuous authentication with cognition-centric text production and revision features. In *IEEE International Joint Conference on Biometrics*, pages 1–8. IEEE.
- Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60.
- Robert Moskovitch, Clint Feher, Arik Messerman, Niklas Kirschnick, Tarik Mustafic, Ahmet Camtepe, Bernhard Lohlein, Ulrich Heister, Sebastian Moller, Lior Rokach, et al. 2009. Identity theft, computers and behavioral biometrics. In *2009 IEEE International Conference on Intelligence and Security Informatics*, pages 155–160. IEEE.
- Barbara Plank. 2016. Keystroke dynamics as signal for shallow syntactic parsing. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 609–619.
- Joost Schilperoord. 1996. *It's about time: Temporal aspects of cognitive processes in text production*, volume 6. Rodopi.
- Ben Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE symposium on visual languages*, pages 336–343. IEEE.
- Motonori Yamaguchi, M. J. C. Crump, and Gordon D. Logan. 2013. Speed-accuracy trade-off in skilled typewriting: Decomposing the contributions of hierarchical control loops. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3):678–699.