Fiona Draxler B.Sc.

Adaptive Writing Support – Suggesting Appropriate Tools Based on Cognitive Processes

MASTER'S THESIS

to achieve the university degree of

Master of Science

Master's degree programme: Computer Science

submitted to

Graz University of Technology

Supervisor

Eduardo E. Veas, Dr. techn. MSc

Institute for Interactive Systems and Data Science

Graz, August 2017

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master‘s thesis.

Date

Signature

Kurzfassung

Texte sind unabdinglich für die Informationsvermittlung und -speicherung. Für viele Leute bleibt das Verfassen von Texten jedoch weiterhin eine große Herausforderung, denn um den Inhalt gut zu vermitteln benötigt man Fähigkeiten beim Planen sowie sprachliche Gewandtheit und muss im Stande sein, die eigene Arbeit zu beurteilen.

In dieser Arbeit erörtern wir, wie Schreibende in allen Schreibphasen unterstützt werden können. Zusätzlich zu Literaturrecherche führen wir ein Experiment durch, in dem wir genauer analysieren, welche Eigenschaften Schreibprozesse aufweisen und welche Schwierigkeiten beim *Suchen* von Informationen, *Planen* und Strukturieren des Texts, *Übersetzen* von Ideen zu Text und beim *Überprüfen* bzw. *Korrigieren* auftreten. Wir formulieren Anforderungen für die Hilfe bei diesen Aktivitäten und schlagen Unterstützungsmöglichkeiten vor. Dabei betrachten wir vor allem digitale Lösungen.

Bereits vorhandene Tools unterstützen meist nur einen Aspekt und unterbrechen das Schreiben. Daher entwickeln wir einen Prototypen für ein umfassendes Schreibwerkzeug, das Schreibenden in jeder Schreibphase hilft. Das Tool ist als Google Docs Add-on implementiert, sodass es nahtlos in den Google Docs-Editor integriert werden kann. Abschließend führen wir eine Nutzerstudie durch, um die Features des Add-ons und den Arbeitsprozess unter dessen Verwendung zu untersuchen.

Schlüsselwörter: Mensch-Computer-Interaktion, Schreibprozesse, Schreibhilfen, Kognitive Psychologie, Google Docs, Ablenkung

Abstract

Texts are of crucial importance for communicating and managing information. However, text composition is still a challenge for many people: in order to effectively convey their message, writers need skills in planning and structuring, linguistic ability, and also the ability to evaluate their own work.

In this thesis, we look at how writers can be supported in all the tasks encompassed in the writing process. To this end, and in addition to literature research, we conducted an experiment to analyse the characteristics of the writing processes as well as difficulties writers typically encounter when they *search* for information, *plan* the structure of their text, *translate* their ideas to words, and *review* their writing. We formulate requirements for aiding these tasks and propose support possibilities, with a special focus on digital solutions.

Issues with existing tools are that they generally support only one aspect and interrupt the writing task. This was our motivation for developing a prototype of a comprehensive text composition tool which supports writers in all stages of their task. We chose to implement it as a Google Docs add-on, which means that it can be integrated seamlessly into the Google Docs text editor. The add-on offers a number of features specifically tailored to each writing process. Finally, we performed a user study to evaluate the features and the workflow while using the add-on.

Keywords: Human-Computer Interaction, Writing Processes, Writing Aid, Cognitive Psychology, Google Docs, Distraction

Acknowledgments

This thesis brings me yet another step closer to graduation, a goal I have worked a lot for in the past years. But this page is reserved for thanks to the people who have helped me with this:
First of all, to my thesis supervisor Dr. Eduardo Veas, who has been very encouraging, who provided many ideas and good feedback.
To Cecilia di Sciascio, for helping me with the implementation, for providing uRank and the keyword extraction algorithm, and for support in the evaluation.
To my family – the people who are always there for you.
And to my friends Doro, Filippo and Miguel for providing me with moral support and food for body and soul.

Graz, August 28, 2017 Fiona Draxler

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this thesis, we present the Writing Support add-on, a Google Docs add-on for text composition. It supports writing by integrating tools for all writing processes in an easy-to use editor environment.

Even though we write emails, scientific articles, reports, etc. on a daily basis, text composition still poses many difficulties, not just for novice writers. We have to pay attention to many factors such as the content, structure, language, and target audience. If we want to achieve good results, we need knowledge and skills in all of these areas.

Nowadays, a major part of the writing we do is digital and could be aided by digital support tools. In some areas, such as spell checking, this already works quite well. However, writing ist not just about spelling, and other processes that occur during writing can be harder to master. For support that caters to all writing processes, we need to know how people compose texts.

This remains very difficult, first and foremost because writing is a very complex activity from the cognitive point of view. It involves various mental processes which need to be coordinated well in order to be efficient. The widely used model by Flower and Hayes identifies the processes *planning*, *translating*, and *reviewing* [Flower and Hayes, 1981]. All these processes can occur on higher as well as lower levels. They have different requirements regarding long-term and working memory, context awareness, physical movements etc. and they vary significantly from one person to the next.

In order to support a specific writing process, we need to identify its characteristics and occurrence, and also develop strategies for the actual support actions. Research has focused on writing processes, but there is not yet a general model determining when which process takes precedence. There exist a number of writing tools, but they almost exclusively specialise on a small part of the entire writing procedure only. Many are explicitly triggered by the writer and thus interrupt the workflow.

Therefore, we can say that there is currently a lack of comprehensive integrated writing support environments and a lack of systems which take into account and adapt to the writers' cognitive needs in real time. We used this as a starting point and motivation for our own contribution.

In this thesis, we will present the development, implementation, and evaluation of the Writing Support Add-on.

Chapter 2 outlines related work on difficulties writers encounter as well as strategies to tackle them. We also list currently existing tools for factual texts and analyse which issues they address and how successful they are at doing so.

We performed a formative study to gather information on how to assess a writer's needs and identify current cognitive activity. The study is described in Chapter 4.

In Chapter 5 we introduce out writing support system. It combines of a number of process-based tools – some of which are already commonly used, but others also based on our own ideas. We concentrated on the interaction between our system, the writer, and the text editor: we wanted the writers to experience minimum disruption but still make them aware of critical points.

Finally, with the user study presented in Chapter 7, we evaluate the usability of the system. We look at the usage of the add-on for a specific writing task as well as quantitative and qualitative feedback on the add-on and its different modules.

Chapter 2

Related Work

2.1 The Writing Process – Literature Summary

This chapter will start with an overview of writing research and identification of needs at different stages before proceeding to the description of existing support systems.

Although writing is an everyday task, it has not yet been fully "decrypted" and remains an active area of research for psychologists, cognitive scientists, and language experts. This is mostly due to its complexity as well as the significant individual differences in the writing behaviour but also prior knowledge, which make it very difficult to predict what happens when. However, there are some general models that identify common actions and their relationship with one another.

In the following, the distinction between novice and expert, weaker and stronger writers will be a recurrent issue. It has to be noted that while all researchers use their own definition of these groups, they usually distinguish by writing experience and based on the text quality: skilled writers produce more fluent texts with clearer structure, better grammar and using more appropriate vocabulary.

2.1.1 Towards a Cognitive Process Model

Earlier models of the writing process mostly described writing as a linear sequence of actions: Rohman, for example, identified the stages *pre-writing*, *writing*, and *re-writing* [Rohman, 1965]; Britton *conception*, *incubation*, and *production* [Britton and others, 1975]. As these terms suggest, there is a clear distinction between the idea construction and the actual written execution, i.e. these models were still closely related to speech and rhetorics [Sommers, 1980]. But writing has a different temporal character: while in a speech, all words are pronounced sequentially from beginning to end, when writing a text you can also start with a later part; while words that have been pronounced cannot be taken back, written text can be changed – in a computer-written document this is particularly easy. You can usually not even see that the text has been modified. Therefore, characterisation of writing demands a model which

takes these differences into account. The most commonly known nowadays[1] is the one developed by Flower and Hayes [Flower and Hayes, 1981]. They used think-aloud protocols to track what writers would do while composing and, based on their findings, developed what they call the cognitive process model. The researchers identified three main processes that define the writing task: planning, translation, and review. These occur repeatedly and have a hierarchical order: one process can be a subprocess of another. Moreover, they can occur on a global, but also on local levels. For example, first, a writer would plan the general outline of an entire text and start writing. While deciding on the exact wording of a sentence, she plans its purpose on a sentence level, puts down some words, reads over the whole phrase again, finds some improvements and iterates once more. Thus, each process type occurs several times, and a high-level process contains instances of different lower level processes or subprocesses. A regulatory module named *monitor* controls which process is activated when. The characteristics of the writing processes are described below.

The Planning Process This is where "writers form an internal representation of the knowledge that will be used in writing", a representation which can still be very abstract. Important aspects here are *idea generation* and *structuring* and *organising ideas*. In this phase, writers gather information by accessing the long-term memory (LTM). External sources such as books are also considered part of the LTM.

Planning also includes goal-setting – on a content-related and procedural level, i.e. dealing with the questions of what to write and how to organise the writing. An important factor here is also the intended target audience.

The Translation Process This is "the process of putting ideas into visible language", i.e. the transformation of a mental representation of ideas into a "linear piece of writing".

The Reviewing Process This process is divided into two main aspects: *evaluating* and *revising*, i.e. rereading and identifying needs for improvement as well as implementing changes.

Overall, Flower and Hayes' paper remains an indispensable part of literature on writing, although their model has been refined by several other researchers since. In the following, we will present some articles which refine, expand or adjust parts of the model, starting with the planning process. We will also mention aspects such as the relationships that were shown to exist between behavioural patterns, text quality, the load of the working memory, etc.

[1] In the literature for this thesis alone, it is used in more than half of the articles, and not all of these were on writing.

2.1.2 Searching and Generating Ideas

Before starting to write, a writer needs to know what to write about. This is why gathering information and reading from external sources are integral parts of text compositions, even more so now that digital sources are readily available. When writing texts using a computer, it is particularly easy to integrate external sources both before and during the writing task. This development is also the reason why in our case, we will give the aspect of searching more prominence and divide the planning process into searching (including idea generation) and planning in the sense of structuring and organising ideas. The current section will give more details on searching as the first part of planning, the following one will concentrate on structuring ideas.

How reading is integrated into writing was, for example, analysed by Plakans [Plakans, 2009]. Using think-aloud protocols, she identified reading strategies and when they would typically occur. She found, for example, that writers were often "scanning the source texts for ideas to use in writing" or "skimming them to understand the main ideas", mostly before they started writing. At later stages, there was a more prominent focus on rereading and checking.

Kennedy also looked at individual differences [Kennedy, 1985]. She found that while in her study "all subjects referred to the reading sources as they composed", they did so at different points in the reading-writing process. In general, skilled readers' reading strategies were distinct from those of less able readers; skilled readers were planning more actively, for example.

Regarding the objective of searching and reading, according to Alamargot et al., writers consult sources to obtain new information, but also to check if the text written so far is in accordance with external documents [Alamargot et al., 2005]. Sometimes, it might be necessary to do an in-depth reading, while at other times the reading is more superficial or focused on one specific aspect. They also mention that the type of sources and the order in which they are consulted can have a significant influence on the final text – this has to be considered when recommending sources.

2.1.3 Planning – Structuring and Organising Ideas

A major component of planning is draft creation. There are large individual differences in how people manage it: Some people need a complete layout of their structure before they can successfully start writing, while others devise their concepts more generally first and only refine them when they get to the respective section. In addition, the planning depends on the nature of the text and the writer's knowledge of the topic. For example, a scientific article usually requires more planning than a narrative account of the writer's last holidays.

These differences were also apparent in a study by Piolat and Roussey, where they analysed essays written by first-year psychology students in an exam [Piolat

and Roussey, 1996]. Around two thirds of the students had created a written plan, out of which only very few were organised drafts. They looked at how drafting correlated with the grade the students obtained and found that better draft quality was correlated to higher grades. Essays with at least a bit of drafting were rated better than those without.

Similarly, Kellogg conducted an experiment to find out how different drafting strategies influence text quality of an informative essay regarding content, style, and fluency [Kellogg, 1990]. He asked students to either not do any explicit planning, to create a written outline, or to draw a cluster of ideas. The written outline proved to be the most successful strategy, while the clustering helped to generate ideas but was also very time-consuming and contributed less to the structuring.

These two studies focussed on informative texts – and obtained results that were confirmed by Hayes in his "New Directions in Writing Theory" [Hayes, 2006]. However, Hayes also raises the question for which kinds of writing a draft is helpful. He states that for some kinds of texts, *freewriting* – although not yet studied sufficiently – could be a good alternative because it might help generate even more ideas during the writing itself, which then, in turn, lead the text into a new direction. Freewriting could leave more room for such ideas than a rigid predefined frame allows.

In summary, we have seen that personal planning behaviour mostly varies in the time that is allocated and the detail to which it is pursued. For informative texts, a certain amount of drafting seems to have a positive influence on the text quality, while for other text families this may be different.

2.1.4 Planning – Goal-Setting

> "People start out writing without knowing exactly where they will end up; yet they agree that writing is a purposeful act."
> [Flower and Hayes, 1981]

Flower and Hayes separate goals into two categories: process goals and content goals. The first type refers to the "how" part of the writing, i.e. the actions a writer is taking. A process goal could, for example, be the creation of an outline, the decision to take some time to think about a specific aspect etc. Content goals, on the other hand, concern the the "what" part. Examples are the definition of a target group ("My text should be easy to understand for researchers in the area of cognitive psychology") or the main topic of a section ("In Chapter 2, my goal is to present the current state of the art").

Both types of goals can be *top-level goals* or *sub-goals*, they can be explicitly formulated or implicit, abstract, or very precise. The writer generally changes and adapts them over the course of the composition. A very abstract goal, for instance, may become more clear when the writer reflects more on it and starts planning at a lower level. Some goals may be discarded when a passage is finished, others may

be added when reviewing reveals need for revision or new knowledge is "discovered". Some goals may suddenly seem less important if a writer realises that her knowledge on the planned topic of a paragraph is a lot better than what she wanted to write about in the following one; she may be tempted to make it less prominent. This last point shows a competition between initial goals, knowledge, and also the current text and is an important point to be aware of.

Based on the think-aloud protocols Flower and Hayes collected, they were able to see a connection between writing skills and the type of goals: better writers defined more process and more concrete low-level goals than novice writers. This suggests that they were more conscious of what they were doing and that novice writers, on the other hand, could profit from clear prompts to define their content goals and from structural guidance in the writing task.

That predefined goals can help, at least for school children, was confirmed in a study by Schunk et al. [Schunk and Swartz, 1993]. They compared writing performance of fourth grade pupils who got a process goal of learning a strategy and a product goal of writing paragraphs. Both types of goals led to higher writing skill scores. The writing performance was even higher when the students were also provided with progress feedback.

2.1.5 Translation

Language skills play a very important role in the translation process. The higher the automatism in choosing words, the lesser the load on the working memory and the easier it is for writers to keep an overview of what they are doing (cf. Section 2.1.7).

Generating sentences from their ideas is especially difficult for second language (L2) writers, for whom translating is "more laborious, less fluent, and less productive" [Silva, 1993]. Nowadays, a major number of scientific texts is written in English but by far not all researchers are native English speakers. Foreign language writing therefore deserves additional support.

In digital environments, the typing skill is also an important factor: Alves found that slower typists paused more often [Alves et al., 2007]. He suggests they could not keep all their ideas in the working memory over the longer span of time and more often had to pause and recover their thoughts.

2.1.6 Characteristics of the Revision Process

Revision is actually not just one action, but a process from determining what should be improved to implementing changes. Scardamalia and Bereiter divided revision into three stages: *compare*, *diagnose*, and *operate* [Scardamalia and Bereiter, 1994]. In the first stage, a writer checks if the text written so far matches the goals and intentions. If not, she diagnoses why this is so. The last stage refers to the actual changes in the problematic passages.

How someone reviews and revises a text is again a factor which can influence the quality of the resulting text. Becker, for example, analysed a number of articles dealing with revision [Becker, 2006]. According to her observations, novice writers are more likely to perform only a superficial analysis of the text they have written and perceive rewriting as "punitive". Professional writers on the other hand "incorporate revision into every aspect of the writing process". Overall, she states that "diagnostic skill is often the most important factor in successfully revising texts" – which means that the development of the latter can have a huge impact on the text quality and is worth further investigation.

2.1.7 The Role of the Working Memory

Another important factor which has a large influence on the final text as well as the occurrence of processes and their duration is the working memory (WM) and its capacity.

Olive investigated the load of the working memory and the role of its components in the writing process [Olive, 2004]. His analysis is based on Baddeley's well-known model of the working memory [Baddeley, 1992]. This model defines three main components: the *central executive*, the *phonological loop*, and the *visuospatial sketchpad*. The second and third are responsible for dealing with speech-based information and visual images respectively, while the central executive coordinates them.

He analysed experiments which mostly used the dual-task technique, where he recorded the reaction times of participants who had to perform two different tasks (primary and secondary) at the same time. In this setting, the cognitive system has to distribute the available cognitive resources to the tasks. If they concern the same "slave" system of the working memory, they have to share resources and are more likely to interfere than if they concern two different systems (i.e. a visual and a phonetic task). Longer reaction times mean higher workload in the primary task. From these experiments, Olive concluded that the linguistic formulation mostly engages the phonological loop while planning is more image-based and thus attributed to the visual working memory.

One further example for the impact the capacity of the working memory has was a study by Kellogg and Olive. It showed that latency (and thus workload) was high when adults were writing and lower when they were pausing or only copying [Olive and Kellogg, 2002]. In the first case, planning, translation, and reviewing happen simultaneously, while in the latter two, only the high- or low-level processes are activated respectively – and the focus lies on one of the "slave systems".

In general, the writing task puts a high load on the central executive. And the more efficient the writing processes are, the lower their demand of cognitive resources. Then, in turn, there are more resources available for "activating other processes and for coordinating goals" [Olive and Kellogg, 2002].

The working memory capacity is also one of the main differences between a

novice and a skilled writer. These differences and their impact on the writing were a focus point in a study by Alamargot et al. [Alamargot et al., 2011]. They wanted to find out which effect the working memory capacity of adults had on the the ability to adapt their texts to the intended audience and also considered the "individual differences in the time course of the writing processes and the content of the final product". Their results showed that a greater capacity of the working memory indeed enabled writers to maintain a better awareness of the target group – and to further integrate the latter into problem representation and goal-setting.

Moreover, the capacity of the working memory for the text production depends on the level of routine in the the allocation of processes. Studies by van den Bergh and Rijaarsdam showed that there seems to be a conscious component in the decision which task should be executed when, and this decision requires cognitive resources [Rijlaarsdam and Van den Bergh, 2006]. At the same time, the resources available for other tasks decrease. Thus, the less conscious effort a writer requires for this decision, the more efficiently the working memory can be used.

On the whole, the analysis of the working memory's role and its impact on composition tasks is a very important point in the development of supporting tools and strategies. The more load can be taken off the writer's working memory, the more capacity is available for higher-level tasks and the more likely it is that a resulting text will be well-written and easy to understand.

2.1.8 The Role of the Long-Term Memory

As mentioned above, the original Flower and Hayes model does not differentiate between the long-term memory (LTM) and external sources but sees the LTM in a more abstract way and simply considers external sources as a part of it. The interaction between memory, sources, and the working environment, however, is in itself a matter worth investigating.

Dansac and Alamargot researched the questions of when and why knowledge is retrieved and if there are repeating patterns for doing so [Dansac and Alamargot, 1999]. Flower and Hayes had already identified two different kinds of access to referential information: during the planning and during the translation processes. In their experiments, Dansac and Alamargot were able to confirm that knowledge acquisition occurred during macrostructural and microstructural planning, that is before and within a block that deals with one aspect. They also analysed which variant occurred in which situation and concluded that this seems to depend on the cost of retrieval and knowledge organisation. Only if organising the content is relatively easy – and it is possible to hold the necessary information in the working memory – is it likely to happen before composition of a theme block, otherwise the writer will look up more information in the middle of translation actions.

Moreover, Alamargot et al. analysed the relationship between the usage of the

LTM and the capacity of the WM [Alamargot et al., 2005]. They concluded that both additional documents and the parts of the text that have already been written are integral to the information that is retrieved. The working memory then has the task of coordinating between the various sources. Its capacity largely influences the strategy chosen to plan and generate content: writers with a lower memory capacity are less likely to be able to deal with information from different sources at the same time, especially if those are contradictory. They also seemed to prefer referring to external knowledge rather than retrieving and reflecting on knowledge from the long-term memory.

One more aspect concerning memory was introduced by Kellogg in his article "Long-term Working Memory": in two studies, he showed that there were two kinds of knowledge or skills that independently influenced writing performance: domain knowledge and verbal ability. The best texts were written by participants who excelled in both areas [Kellogg, 2001].

Again, as for the working memory, the analysis of the long-term memory's role in text composition and the interplay of the working and the long-term memory is important for devising tools for supporting writers. It hints at when to present which kind of information based on the complexity of the topic and the knowledge as well as abilities of the writer.

2.1.9 Process Occurrence and Detection

An essential factor in the study of writing processes is detecting when they occur. This is not an easy task – both regarding the complexity of composition and the cognitive psychology behind it. In the following, I will present the methods that researchers commonly apply and some of the obtained findings.

Flower and Hayes used *think-aloud protocols* for this task [Flower and Hayes, 1981]. This means they asked study participants to voice their thoughts during composition. Everything they said was noted down. Afterwards, the statements were manually attributed to planning, translation and review.

Kellogg, on the other hand, used a retrospective method [Kellogg, 1987]. At set intervals, a beep would interrupt study participants. They were asked to react as quickly as possible and then to think back to what they had been doing just before the beep. After a bit of training, they were able to attribute their last actions to a writing process by pushing one of four buttons (assigned to planning, translation, review and unrelated processes).

Piolat et al. employed the so-called *Triple Task Technique* to record not only the currently active process but also the working memory load [Piolat et al., 2001]. They did so by measuring the time it takes a writer to react to an auditory probe played back as in Kellogg's experiment. Thus, the first task is the writing itself, the second task the detection of the audio signal, and the third the retrospection of what the writer was thinking and doing before the sound was played.

Experimental results included a typical pattern for when processes occurred over the course of a composition: as could be intuitively suggested, in the beginning, there was a stronger focus on planning and translating, while in later stages, evaluation, revision, and execution prevailed. The writing task was automatically given the most attention and thus became the primary task, even when writers were prompted to focus on a different task.

Alamargot et al. tracked eye movements and the position of the pen on a digitising tablet [Alamargot et al., 2011]. Thus, they were able to detect pauses and shifts in focus of attention, e.g. from the text to external sources. They had previously divided the study participants into a low and a high working memory capacity group and could find that participants with higher capacity made longer pauses – this could be a reason for their better audience awareness, as mentioned in Section 2.1.7.

Levy and Ransdell used video recordings and logged the time that study participants needed to react to an audio signal in order to measure the cognitive load [Levy and Ransdell, 1994]. The video and audio signals were played back after the sessions and raters attributed time frames to a writing process. They considered using a keystroke logging software but it proved too cumbersome to use for their purpose. Since logging software has greatly advanced since 1994, this should however not be seen as a reason to disregard logging tools.

Indeed, in the experiment described in their 2009 article, Wengelin et al. combined eye tracking with keystroke logging in order to find patterns that would be characteristic for a specific writing process [Wengelin et al., 2009]. They argue that the combination of the two gives a better insight into what happens during composition. One use case would be the more detailed analysis of pauses. A fixation–saccades sequence typical for reading, for example, suggests that the writer is going over the text written so far. Overall, their analysis so far remains very hypothetical but gives clues as to how these tools can be used to monitor writers – for more insights also in combination with retrospective methods or think-aloud protocols. One factor that limits the use of combining gaze tracking and keystrokes, however, is the fact that in a composing task, the characters are not guaranteed to stay on a fixed position on the screen but are moved whenever the page is scrolled, which means that the correspondence between direction of gaze and characters on the screen is hard to establish.

Some of the techniques used for process detection may be applicable in a real-life environment and on a day-to-day basis: logging software is unobtrusive, and eye tracking is already possible with relatively cheap and compact hardware. More precise methods based that require interaction with the writer (e.g. pushing a button), however, interrupt the flow of the composition. It was not possible, so far, to create a sufficiently reliable model which would predict process occurrence but not disrupt the writer.

2.1.10 Quality Measures

Supporting writers means enabling them to write faster, make the writing less cumbersome, and to improve the quality of their texts. But how do we measure how good a text is?

We can consider different aspects here, e.g. how easy a text is to read and understand, how well it is oriented towards the target audience and how sophisticated the content is.

A manual approach is, for example, the set of guidelines used for SAT essay scoring[2]. In order to obtain a high score, essays must cover requirements in different categories: they must show good command of the English language, cohesion, variety in sentence structure, an adequate tone and style, a logical order of ideas, and a precise statement.

Pitler and Nenkova present an approach based on "lexical, syntactic, and discourse features to produce a highly predictive model of human readers' judgments of text readability" [Pitler and Nenkova, 2008]. Discourse features give insights into the structure of a text, e.g. if it is written in a coherent way. From their features, they calculated a measure which yielded good predictions for readability; however, at the time of writing, there was no reliable way for automatic discourse annotation, i.e. feature extraction.

2.2 Requirements for Writing Support

The previous section gave an overview of cognitive processes during writing and difficulties regarding the skill of the writers, the capacity of their working memory, and the retrieval and organisation of knowledge were identified.

Let us now summarise in which specific areas writers could benefit from support tools. We will list the requirements such tools must fulfil in order to enable writers to compose texts of higher quality with respect to content, structure, and language, to write theses texts in less time and with less difficulties and mental effort. We will start with some general requirements that are process- or state-independent.

1. Standard user experience requirements such as easy usage, consistent design, good integration of the functionality,...

2. A balance in the trade-off between disruption and awareness: Writers must be able to concentrate on their task without being interrupted (e.g. by attention-seeking blinking interfaces). At the same time, writers should be made aware of changes like new information, of guiding prompts, or other cues for possible interaction.

[2] https://collegereadiness.collegeboard.org/sat/scores/understanding-scores/essay

3. Context-awareness and adaption: Not all information is useful at all times. A good writing tool “knows" the needs of its user and guides through the composition process by displaying the right type and amount of information at the right time.

4. Accessibility: the more important a feature is at a given moment in time, the quicker and easier it should be to reach and to use.

5. Ease of use and efficiency: as stated in the section on working memory (cf. Section 2.1.7), writing requires substantial cognitive effort. The more capacity is used up by overhead actions of a writing support tool, the less is available for the writing task. In an ideal state, on the other hand, it should act as an extension of the writer's memory and enable a writer to perform better in tasks that would otherwise be disregarded, e.g. the awareness of the audience (which seems to be a particularly difficult point for less skilled writers).

6. Support reflection: make writers better judges of their own actions and thus “teach” them how to improve.

In addition to this, we can formulate process-specific requirements:

- Searching
  - When accessing external sources, both an overview and a detail mode should be supported, so that scanning and skimming as well as detailed research are supported
  - If sources are suggested, then they should be relevant to the topic and of high quality. The selection should be unbiased (remember that the order in which items are presented also has an influence), and appropriate for the selected genre. Suggestion of sources could be realised with a search engine or a personalised recommender system.
  - Make access to and coordination of external sources as simple and immediate as possible. Support easy organisation of the content and consequently also less disrupted writing by reducing look-up times. Such a search tool could be considered an extension of the long-term memory.
- Planning
  - A planning tool should generally incite writers to engage in planning. They should be conscious of what they want to achieve and how they want to get there. A tool could, for example, give prompts to explicitly formulate goals.

  - For factual texts, it should be possible to easily create at least a rough draft: In Section 2.1.3, we saw that drafting – if it does not require too much time – went hand in hand with better text quality.

- Translation
  - Generally, a translation tool should help writers generate text from a structured collection of ideas. This text should be structured and well-connected so that the message is clear to readers.
  - Particular attention should be paid to real-time aid.
  - Not only low-level linguistic features, but also fluency, logic, and connectivity should be supported.

- Revision
  - A revision tool should help detect mismatches between the set goals and the composed text, e.g. whether the linguistic level and the content of a text are appropriate for the target audience.
  - It should raise awareness for style and formal structure issues, e.g. the use of unusual phrases, inconsistent section division etc.
  - On top of this, semantic checks are of particular value: coherence, fluency, and the logic of arguments are generally the most difficult parts to evaluate but at the same time provide a lot of potential for improvement, especially for novice writers.

2.3 Writing Support and Tools

This section provides a list of available (digital) writing tools, grouped by the writing process that is their primary focus. We will check which kind of support they provide and which requirements they fulfil. Screenshots of the presented tools can be found in Appendix A.

2.3.1 Searching

There exists a number of tools intended for the retrieval of related content. The simplest ones basically provide search-engine functionality with options to insert references and create a bibliography from scientific articles. Examples for this are:

- the Google Scholar extension for Chrome[3]. It presents a list of articles based on currently selected text on a website or keywords the user enters into the search box.

[3] https://chrome.google.com/webstore/detail/google-scholar-button/ldipcbpaocekfooobnbcddclnhejkcpn

- the Google Docs Explore button[4]. Google recently added an "Explore" feature to all Drive documents. For Docs, it replaces the Research tool and provides a sidebar with the following features:
  - an overview panel with sections for topics (used as search request triggers), images, and research links (not necessarily scientific resources). The content is suggested based on the document's content and updated periodically.
  - a search button for explicit querying of the web and the user's Drive documents. Results are grouped by web, images, and drive.
- Microsoft Word Researcher[5]. Similar to the Explore tool, Office 2016 was equipped with a sidebar which uses the Bing Knowledge Graph to search for related content from the web. The retrieved material can be explored and directly cited in the document.
- the EEXCESS[6] add-on and browser extension. The browser extension adds a footer to websites, where keywords which were automatically extracted from the currently selected paragraph are displayed. They can be dropped and filtered and the main topic can be changed. When hovering over a keyword, its source in the text is highlighted.

  The extension uses EEXCESS to look up related content for keywords and displays titles in a grid. In addition, there is visualisation for filtering the results based on several metadata characteristics like time, language, and license. A major drawback is that the retrieved items are not ranked by relevance.
- the Google Docs add-on Paperpile[7]. It provides a search box to look up articles based on authors, keywords or DOI. The search results each have a cite button to include them into the document. The citation in the document is linked to a page on the paperpile website – which also stores all citations for a document.
- the Zemanta[8] plugin for wordpress, which searches for related content and suggests images, mostly for blog posts. The displayed related articles are also mostly blog articles. The user can choose preferred blogs as sources for personal search. The plugin also extracts keywords from the text that can then be linked, e.g. to a Wikipedia article, with a single click. The search is initiated automatically and can also be triggered by the user.

[4] https://blog.google/products/docs/explore-docs-sheets-and-slides/
[5] https://blogs.office.com/2016/07/26/
[6] http://eexcess.eu
[7] https://paperpile.com
[8] https://wordpress.org/plugins/zemanta/

This list only presented a small selection of available search tools, many more exist for different domains and applications. They mostly differ in the kind of information they offer, be it scientific content, images, newspaper articles,... Some of them automatically suggest relevant resources while others need the user to enter keywords for triggering search queries. The tools designed for creating scientific content usually provide a way of citing the results when they are used in the composed text. There are variations for all kinds of working environments and text editors. Some are indeed very easy to use and well integrated into the respective environment, so they are not too distracting.

What these tools generally lack is context-awareness, that is they do not take into account what the writer is currently doing. In his doctoral thesis, Rhodes [Rhodes, 2000] describes a *Just-In-Time Information Retrieval Agent* (JITIR) which does not only recommend related content but also takes great care of when and how to present it. An important characteristic of such an agent is that it should reduce the "cognitive effort required to find, evaluate and access information". He also mentions that the higher the effort or cost to perform a task, the less likely it is that it will be performed. In the case of looking up relevant data, it is likely that writers often refrain from using an external search engine because it will take too long – a delay of two seconds already makes a task less likely to be executed. This means that a lot of potentially useful information will not even be referred to. An integrated system which proactively searches for related content and displays it in the same interface could greatly reduce this cost and thus make it easier for a user to include more sources.

One section of Rhodes' thesis is devoted to the question how potentially useful information should be presented: it should not distract the writer too much from the current task but it should be easily noticeable when necessary and easy to skim over in order to check the relevance.

Rhodes implemented and evaluated four different JITIRs, including the *Remembrance Agent*[9], a tool for emacs which displays suggestions – a list of related documents – at the bottom of the editor window. They are updated every few seconds while the user is active. Experiments showed that when using the Remembrance Agent, users retrieved a higher number of documents; external queries through the search engine were still used nearly as often as before. Therefore, the RA was not a replacement, but an extension and only partially a substitute for the search engine. Rhodes also notes that for the usability of the RA, the interface and recommendation quality were very important factors.

More recently, Melguizo et al. developed a proactive recommender system for writing named IntelliGent [Melguizo et al., 2010]. The tool "proactively submits queries based on a broadly defined user profile in combination with what the user is currently typing or reading". Recommendations are displayed in a dedicated

[9] https://github.com/zzkt/remembrance-agent

area on screen, on hovering over them, a corresponding URL is expanded and can be navigated to with a click. Melguizo et al. also looked at the disruption their tool caused: participants stated to have been most distracted while translating and reviewing and less during planning, where the suggested information is also intuitively most valuable. Moreover, they found that adaptation to user needs in specific writing stages is quite complex. IntelliGent did not predict active writing process, but participants were prompted to perform their actions sequentially – which is not a real-life setting.

One further limitation of proactive systems is that they need a starting point or context to use as a basis for queries, whereas in search engines, users themselves provide the necessary input. Thus, proactive systems are more easier to apply for idea-driven tasks than for data-driven tasks.

2.3.2 Planning

Planning tools support goal-setting, idea generation and structuring of ideas, but also draft or outline creation.

The simplest way to support goal-setting is probably with a to-do list manager such as trello[10], which allows users to create tasks and to add notes and a deadline. However, this is not specific to text composition, so writers would have to come up with their own items.

A common practice in teaching is to give cues cards. For instance, Feretti et al., designed prompts telling students to "think of two or more reasons to back up your opinion" and to "think of two or more reasons that other people might use to back up their opinion" [Ferretti et al., 2009]. With this support, students' consideration of alternative views increased, so the cards had a positive effect on the achievement of goals.

Nowadays, most text editing applications include paragraph style predefinitions which automatically generate an outline based on the heading styles used in the text. Prominent examples are Microsoft Word and Google Docs. With the latter, the outline can then be displayed as an extra panel on the left side of the browser window. These tools are a good first step for drafting: headlines can be used to define the structure of a story or an article. As was stated in Section 2.1.3, not every writer plans the same way and one person can and will choose different strategies depending on the context. If an outlining tool is directly integrated into the text editor, it can accommodate for these needs. For instance, when a writer wants to change her outline later on, it is sufficient to update the and the outline will be adjusted accordingly. However, even from personal experience when collaborating on documents we know that a number of people do not use the heading styles, but

[10] https://trello.com

format the text directly – in this case, the outlining cannot be used. Besides this, the provided drafting support is very basic.

More in the area of local structuring, Englert et al. [Englert et al., 2005] developed a web-based tool for planning paragraphs which was aimed at children with learning disorders. For each part of their text, the children were asked to fill text boxes for topic sentence, paragraphs and a concluding sentence. This gave them guidance in forming their arguments. Most of the children managed to achieve higher text quality than without the scaffold. In the end, the parts were combined into a single text.

Scardamalia and Bereiter did an extensive study on methods to improve reflection capabilities of 6th graders over a duration of 15 weeks [Scardamalia et al., 1984]. The children in the experimental class got some general strategic instruction for planning and cue cards which they were supposed to use in a planning phase whenever they got stuck. The phrases on the cue cards (e.g. "An important point I haven't considered is...") were then integrated into a think-aloud monologue. The children in the experimental group performed better in the essays they had to write than the control group without cue cards and their actions were generally more reflected. The cue cards had apparently provided guidance and made children think about aspects they had not considered before.

Idea generation and grouping can, for example, be aided with mind mapping or clipboard-style tools. One implementation is the so-called "Klemmbrett" (clipboard) of the German application Papyrus Autor[11]. It provides a surface for the organisation of boxes with notes and also the option to create a character book, especially useful for novels, where information about characters as well as timelines can be stored for consistency. The content of a note can be dragged to the text area and is then automatically inserted.

2.3.3 Translation

Translation is generally the least regarded writing process when it comes to the availability and diversity of tools. However, it has to be noted that spell checking, for example, is already implemented in most operating systems nowadays and thus automatically more often used than not.

In general, spell checkers work with a list of properly spelled words and determine whether a word in the text is contained in it or not. That is, there is a positive list of options [Mahlow and Dale, 2014]. Difficulties arise when a word is misspelled in the current context but is actually a correctly spelled word in a different context, e.g. *from* and *form* are easily mistyped but have a different meaning. In addition,

[11] https://www.papyrus.de

proper nouns and neologisms make it hard to say for sure if a word is correct or not. Depending on the language, there are also issues such as composite nouns in German. Spell checkers can suggest replacements or even auto-correct misspelled words. However, the auto-correction is prone to errors as well; the replacements may well be worse alternatives than the original word. Despite these difficulties, spell checkers can have a good overall performance and be a great help to writers in all circumstances.

In addition, real-time grammar checks immediately show a writer if a text meets the syntactic rules of the language it is written in. A typical approach would be that the checker considers "a grammatical error to be present when the parser's standard grammar is unable to find a parse; in such a circumstance, some additional processing is then invoked in order to determine the nature of the error" [Mahlow and Dale, 2014]. Commonly detected errors include a mismatch of subject and verb or wrong cases of pronouns and adjectives. As with spell checking, grammar checking requires a large knowledge base, which is by no means guaranteed to be complete.

Another tool that helps with the actual formulation of phrases is a synonym suggestion feature as implemented in Microsoft Word: on right clicking on a word, the user is provided with a list of possible replacements (see Figure 2.1). Using varied language can greatly improve the impression of fluency and make a text more interesting to read.

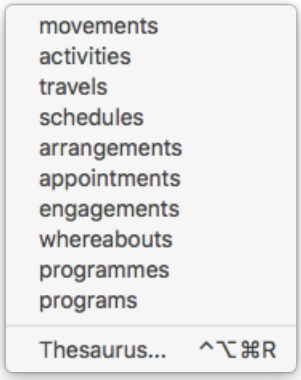


Figure 2.1: Microsoft Word synonym suggestions for "actions"

These tools all deal with low-level linguistic aspects, but there is only little support for semantic features which would tell a writer if her writing is fluent and comprehensible.

Two tools, Coh-Metrix [Graesser et al., 2004] and TAACO [Crossley et al., 2016], compute a number of measures for cohesion and coherence. Cohesion and coherence are concepts for linguistic and semantic connection and thus understandability of

texts. Both Coh-Metrix and TAACO provide an interface where users can paste text or select files to be analysed. Then they output values for a number of measures such as the ratio of particular word classes, the number of connectives, the type-token ratio etc. However, if you are not an expert in dealing with cohesion/coherence, it is unlikely that you will know how to interpret the results (Coh-Metrix at least provides help links). Since the systems are designed to work with full texts, they are more easy to use in a revision state than during translation – although generally, real-time cohesion and coherence checks could be very helpful there. This is also why they are mentioned here in the section on translation.

2.3.4 Revision

Automatic style analysis can be a starting point for successful revision. Mahlow and Dale mention syntactic measures such as "the length of sentences or words, the frequency of particular parts-of-speech, the length of noun phrases, the relative proportions of sentences in active and passive voice, and the complexity of sentences" [Mahlow and Dale, 2014]. Highlighted points give writers an idea which aspects of a text could be critical for a reader and how they could make it seem more natural and easy to read. On the other hand, there is the danger that users simply trust style recommendations and accept them as correct, which means that such tools can have a large influence on final outcomes – and that is not a good idea if it limits individuality and creativity.

Most of the available digital reviewing tools focus on linguistic issues – these are a lot easier to identify by an application than content issues would be, as no topic knowledge is required. Below, we will list and describe several applications.

- Grammarly[12] is a widely used proof-reading service. In the basic version it is actually rather a translating tool, since it supports grammar and spelling checks. The paid full version, however, also unlocks features for style and vocabulary improvement. Grammarly can be used as a standalone app, in the browser, or integrated into Microsoft Word.

- Papyrus Autor does not only help with planning (cf. Section 2.3.2), but also performs a thorough analysis of the writing style – German texts only, though. As can be seen in the screenshot in Figure 2.2, the application highlights repeated words, and crosses out adverbs which might be superfluous, points out spelling mistakes, usage of the passive form, overly long sentences etc.

- GradeProof[13] provides, amongst others, a Google Docs add-on. Its suggestions for style review are presented in a similar way as those of Papyrus Autor. A text is analysed regarding spelling, grammar, and style (rephrasing suggestions). Common recommendations are words or expressions which can be left

[12] https://www.grammarly.com
[13] https://gradeproof.com

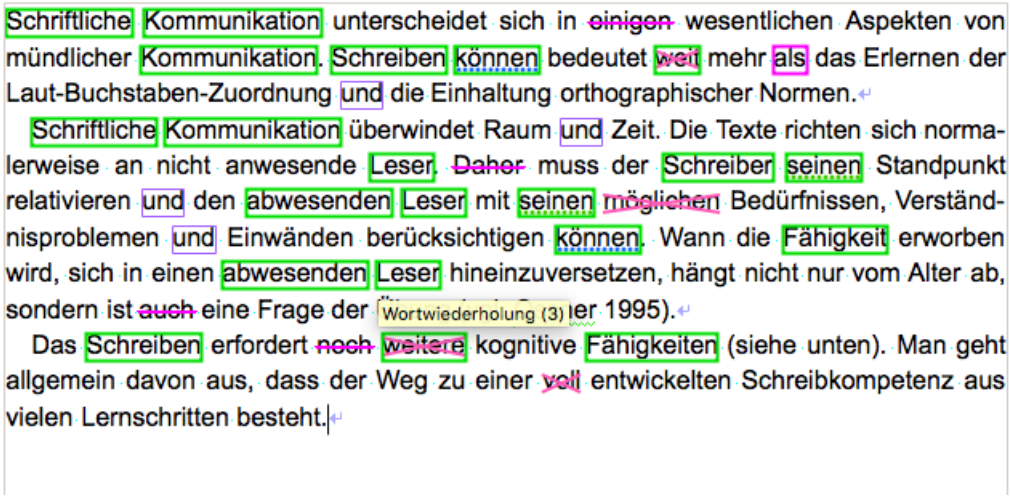


Figure 2.2: Papyrus Autor style analysis

out, e.g. "that" as an introductory word in a subordinate sentence or adverbs like "very". In addition, the panel displays statistics such as the word count and the Flesch-Kincaid grade level [Flesch, 1948]. The latter predicts readability as the number of years of education necessary to understand the text. The score uses the ratios between the total number of syllables as well as between the number of words and sentences in a text, assuming that words with fewer syllables and sentences with fewer words are easier to understand. These two ratios are scaled and combined in such a way that the result is supposed to represent the number of years of education necessary to understand the given text. This measure is of course only heuristic since it does not consider aspects such as the type of sentence construction, the register of words, etc., but is still useful for the detection of mismatches. It is important to note that as such, the measure only works for English texts, but there are variations for other languages.

- ProWriting Aid[14] also comes with a Google Docs add-on which provides several different reports on a document. The key figures are presented in comparison to the results of other ProWriting Aid users. The summary report gives an overview of:
  - word count, sentences, paragraphs, and characters
  - vocabulary, including a list of the most used, most unusual, overused words
  - sentence length, number of long sentences

[14] https://prowritingaid.com

- style suggestions: similar to Papyrus Autor, that is counting adverbs, checking for passive voice. Lists possible replacements for expressions
- grammar issues
- consistency

Most detailed reports are paid features of the full version of the add-on. Words or phrases which were identified in a report are highlighted in the text.

How quality feedback can be integrated into a (writing) activity was shown be Fan et al. [Fan, 2017]. Their mobile app CourseMIRROR was not developed for writing revision, but for reflections on university lectures. After each lecture, students are prompted to answer reflective questions with short statements. The app then checks the quality of the reflections using Natural Language Processing algorithms and gives real-time feedback via a coloured progress bar and improvement suggestions. This helped students to improve their reflections and to go into more detail on unclear points – an outcome which writing tools would also greatly benefit from.

2.4 Summary

In this chapter, we looked at components of the writing process and analysed why it is such a complex activity. We saw in which situations writers are likely to struggle and which strategies have been found so far to support them and teach them to improve.

We also presented a selection of implemented writing tools. In some areas, there are many available tools: we listed six search plugins or add-ons which retrieve documents and four revision tools which create reports on text statistics; and these are by far not the only ones available. On the other hand, it was difficult to find tools which do not only scratch the surface by extracting keywords or linguistic statistics, but offer further assistance for semantic aspects such as fluency and coherence or tools that analyse the logic of an argument. Semantic analysis is of course a lot more complex and challenging, but it is also an area where more sophisticated support has great potential.

Furthermore, most of the items we mentioned cover exactly one aspect. Only some are integrated into an editor (e.g. as a Google Docs add-on) and they have to be explicitly opened when they are needed. It is upon the writer to decide when this is the case, thus causing a disruption in the flow of writing. While their functionality may have great potential, the usefulness of the tools is therefore limited and their usage often cumbersome.

Chapter 3

Method

In the previous section, we described our extensive literature research on difficulties writers encounter as well as strategies to tackle them. We used this knowledge to compile a list of requirements for writing tools. Then we looked at currently existing tools and analysed which issues they address and how successful they are at doing so.

Our own contribution consists of a formative study on writing processes, the implementation of a writing support tool for factual texts, and its evaluation. Details on our approach are presented below.

3.1 Formative Study

As a first step, we perform a formative study to further explore writing processes. The study is meant to be explorative in the sense that we also want to find out how much information we can obtain from data collected in a real-life setting, i.e. without any additional hardware or external software. The goals of the experiment are as follows:

- Confirm findings from the literature regarding process occurrence and behavioural patterns, e.g. process occurrence and pausing behaviour.
- Discover writing patterns from the activity data and common actions depending on the writing phase, e.g. pausing behaviour and writing progress
- Identify needs of writers depending on the context
- Determine to which extent the gathered information is sufficient for identifying a writer's current needs

For creating a process-based writing support system, such information is crucial for the detection of processes and, based on this, the suggestion of appropriate tools.

3.2 Implementation of a Writing Support Tool

The next step is to combine the outcomes of the formative study, related work, and analysis of existing tools in order to design a concept for a writing support tool. This tool is meant to aid the entire process of writing, in a method that integrates seamlessly and keeps disruption to a minimum.

Regarding functionality, we need features for all the writing processes searching, planning, translating, and reviewing. For efficient usage, close attention has to be paid to the design of the interface and interaction. Moreover, we have to adhere to the further requirements for writing support tools we listed in the previous chapter.

3.3 Evaluation of the Writing Support Tool

The third part of our contribution is an evaluation of the tool we present. To this end, we perform a user study with the following goals:

- Assess the usability
- Identify usage patterns of the add-on
- Relate text quality to prior knowledge and usage patterns
- Identify areas for further development

Chapter 4

Formative Study

This section describes procedure and outcomes of the formative study presented in Section 3.1.

4.1 Experimental Setup

Two requirements can be derived from the goals of the experiment: the observation of process occurrence and characteristics of writing behaviour.

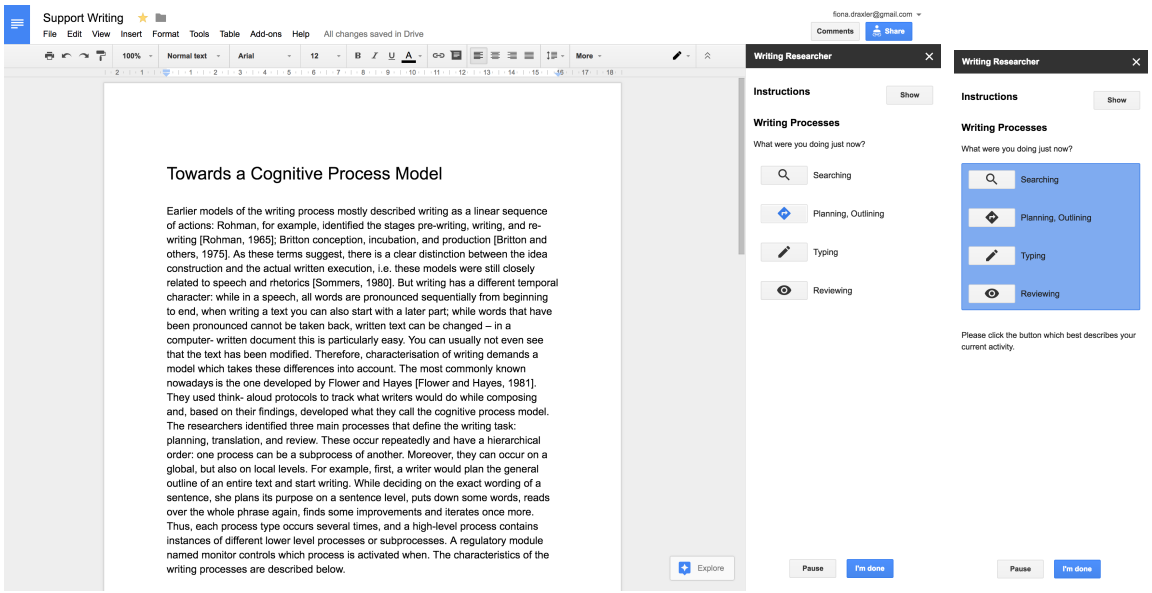
To this end, participants were asked to write a scientific text of 1 to 2 pages in length. During the composition task, they were also prompted at given intervals to indicate the process that best fit what they were currently doing. This setup resembles Kellogg's retrospective method as described in 2.1.9. We did not add keystroke logging, video or audio recordings, eye tracking, or screencasts and did not apply a think-aloud technique because we wanted the setting to be as close to a real-life situation as possible and wanted to evaluate how much information such a reduced setting can provide. The tracking is directly integrated into the writing environment and can thus not only be used in an experimental context but easily be adapted to a different setting.

The texts were written using Google Docs and an add-on whose main interface was a sidebar on the right side of the Google Docs editor window, as illustrated in Figure 4.1. The sidebar showed instructions and buttons for indicating active processes. In addition, there were two popups supporting the study progress. The workflow was devised as follows:

1. The research add-on is started via the *Add-ons* menu and then displays a dialogue welcoming the participant and redirecting to a pre-study questionnaire.

2. On returning from this questionnaire, the instructions for the text to be composed during the study are displayed.

3. The main interface is opened in a sidebar. It recaps the instructions, but most importantly contains four buttons, each associated with one of the writing



(a) Normal state **(b)** Prompting state

Figure 4.1: Sidebar for the formative study

processes *searching*, *structuring*, *typing*, and *reviewing*. The participant is asked to click them at will and when prompted to indicate what she is currently doing (by highlighting the button container).

4. As soon as the writing is done, another popup leads to the post-study questionnaire.

4.1.1 Collected Data

In the background, the add-on collected data on the writing process and sent it to a server where it was stored in a database. A number of measures were collected between one and five times per second:

- timestamp
- currently selected process
- full text of the document
- text length
- cursor position within the current paragraph
- selected text and selection offsets

- used paragraph styles (differentiates between heading types and normal text)
- state of the browser tab (active or not)

In addition, the revision history for all documents was extracted. The revision history is made up of network calls used to keep the local and remote state synchronised. It can be accessed for any document a user has write access to and included the following:

- timestamp
- inserted text and insertion position
- start and end indices of deleted text

More details on the implementation can be found in Sections 5.5 and 6.

4.1.2 Participants

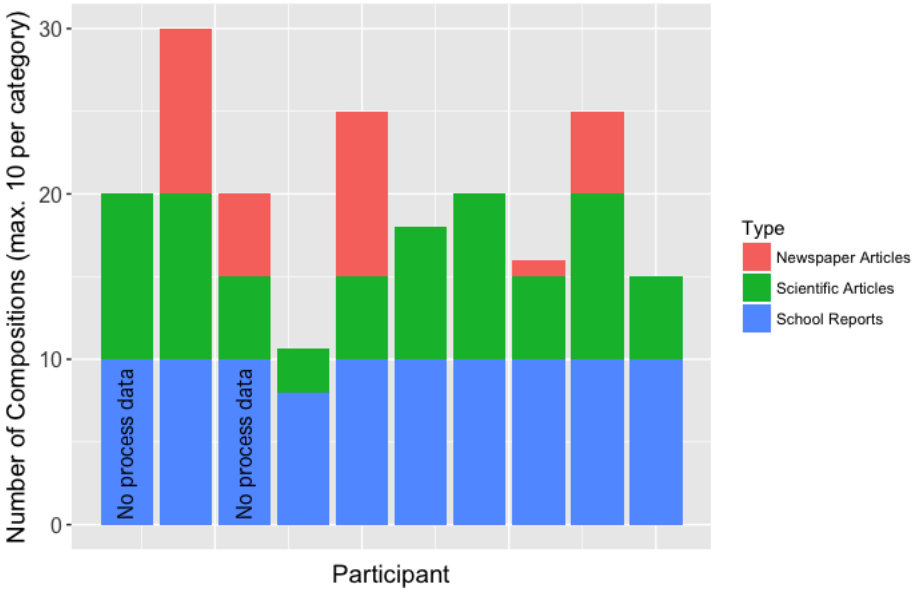


Figure 4.2: Number of articles written in English, in different categories, for each participant ("writing experience")

The participants of the formative study were mostly, although not exclusively, researchers and students working in the area of computer science. The age ranged from 20–29 to 50–59, with 20–29 being the largest group.

We did a basic assessment of their writing experience, based on the number of English articles they had written in school, for newspapers or blogs, and scientific

reports or articles, see Figure 4.2 for more details. The highest number they could choose was 10, representing 10 or more articles. Most of them had written many English texts in school, so this was not a distinguishing measure, but together with the stated English skills (which were mostly "advanced", one "native speaker", one "intermediate"), they indicate a good minimum level in English writing. The other categories were more varied.

Eight participants concluded all parts of the experiment, which means that they filled in the questionnaires and recorded their activity with the Writing Researcher add-on. In total, they wrote 12 usable texts; two participants wrote three texts each. One participant extended an already existing text. Closer analysis of his text shows that he started using the research tool once he had added headings and created an outline, that is, concentrated on planning. Because there was no data for more than half of his writing, we only used his information in graphic analyses, e.g. the table in Figure 4.3.

In addition to that, questionnaire and revision history data is available for further two – therefore in total ten – participants and further five – therefore in total 17 – documents.

4.2 Analysis

The data we gathered was used for two main purposes: firstly, relating the information about active processes collected through the button clicks by participants to context variables and secondly finding indicators for areas where writers are more likely to need support.

4.2.1 Process Occurrence

This section will describe what the clicks on the process buttons revealed about the occurrence of processes. First of all, it has to be noted though, that this data is noisy: some participants commented they would occasionally focus so much on the composition task that they did not notice the highlighting of the process buttons. Moreover, due to the hierarchical nature of processes (cf. Section 2.1.1), even for experts it is not always obvious which process an action is to be attributed to. In a reviewing process, for example, when a writer identifies a section to be changed she then switches to a lower-order typing phase. Some participants then said that this was again translating while others continued to attribute the action to the higher-level reviewing process.

Still, there are some overall trends which can be identified. First of all, there are obvious differences between the overall time allocated to a process. Table 4.1 shows that translating is the most dominant process, making up more than half of the time. There is not a big difference between searching, planning, and reviewing.

How does it look, though, if we do not consider all documents at the same

Process	Time Share
Searching	16.3%
Planning	14.3%
Translating	52.8%
Reviewing	16.5%

Table 4.1: Percentage of time that each process was active, as calculated from button clicks

time but look at them individually? Here, it quickly becomes apparent that there are large differences between them. Figure 4.3 gives an overview of the processes as selected by participants over the progress of their document, that is, over the relative time passed from start to completion. The documents are ordered according to their absolute length, decreasing from top to bottom, see the bar to the right representing the number of characters.

One clear difference lies in the granularity. During the composition of documents 6 and 7 for example, there were fewer changes in processes than in both longer and shorter documents. Moreover, document 5, contains many searching phases spread out over the entire time, whereas in documents 11 and 12, nearly no searching takes place.

However, even if they are not very distinct, there are some recurring patterns. Looking at the relative starting point of each process, averaged per document (see Figure 4.5a), we can find that planning activities most often start in the first half of the time passed whereas reviewing phases are more prominent in the end. A Wilcoxon rank sum test (the data was not normally distributed) shows that the average starting times of planning and reviewing are actually significantly different with $p \approx 0.001$.

On the other hand, the time at which a searching phase typically occurs in a document varies greatly, showing that there are indeed different strategies – some participants gathered almost all their information before actually getting to the typing, while others continued to retrieve new information as they went on. One participant, for example, summarised his actions as follows: "I did more planing in the beginning and then I was alternating between writing, searching and revising."

If we also consider the distribution of the averaged relative starting times (cf. Figure 4.5b for the standard deviation), we can see that it remains fairly constant for the translation process, but changes more for the three others. This shows that in some documents, writers used a process in the beginning and end while others kept them either closer to the beginning or end. Note, in particular, the outlier points, which indicate that there were occurrences at very different times or very close to one another.

Another indicator for an active process could be the pausing behaviour, that is,

Figure 4.3: Occurrence of processes while the text composition progresses. The bar on the right side indicates the document length. Recording pauses (not writing pauses) of more than 10 seconds were removed.

the time where no text is typed. Intuitively, we could say that a writer is searching, planning, or reviewing when she is not typing. However, pauses also occur during the translation process, as is visible in Figure 4.6. What is different, though, is the length of pauses. Reviewing and searching pauses tend to be a lot longer than planning or translating pauses.

4.2.2 Progression Analysis

Analysing writing behaviour does not only give us insights on the occurrence of processes. It is also important in order to identify weak points or deficits, and consequently present tools which help overcome them.

An example of this is the characterisation of the writing progress – a factor which varies a lot from one writer to the next. Some people might prefer writing a sentence, then immediately going over it again to review and adapt it while others note their text down in one go and keep changes for the end. We identified all insertions and deletions of characters in all 17 documents, aiming to identify dominant patterns for different writing strategies.

First of all, we calculated the ratio of inserted and deleted characters for each document, that is the percentage of inserted characters which were also deleted

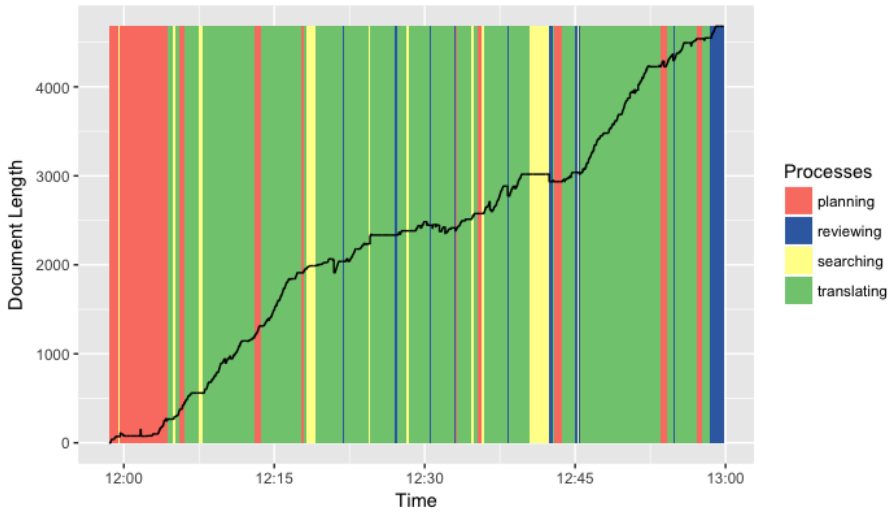


Figure 4.4: Occurrence of processes over time for a sample document. The line indicates the development of the document length.

again. This ranged between 18.5% to 77.2%, which supports the hypothesis that the workflow differs quite a lot from one writer or document to the next.

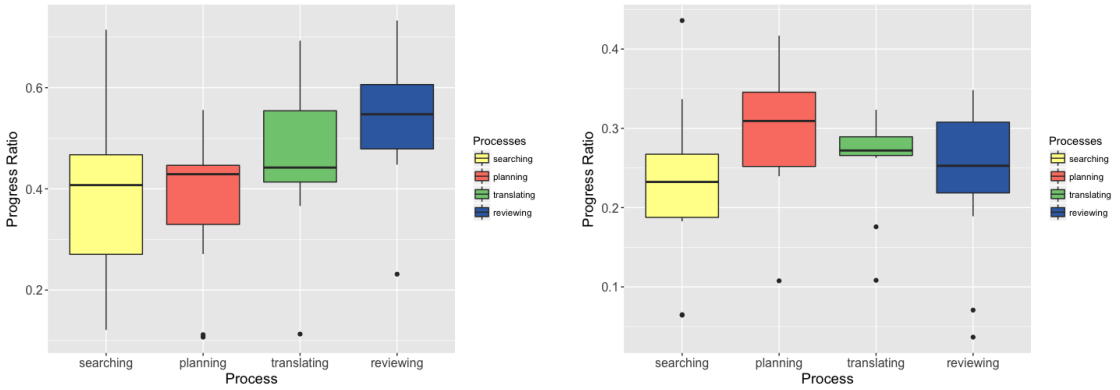
Next, we related the number of scientific articles the study participants had written to this ratio (for the participants that had written more than one document, we used the mean of this ratio) and discovered a Pearson correlation of 0.57 between the two variables. This means that overall, writers who had written more articles – and thus more experience in scientific writing – were more likely to delete a larger percentage of their text.

Moreover, more experienced writers also deleted, on average, more characters in one go: the Pearson correlation between the number of articles and the average number of characters per deletion was 0.47.

4.2.3 Methodology of the Study

The study was useful to collect helpful data, even though we kept the setting as close as possible to a real-life setting by limiting ourselves to the data we could gather via Google Docs alone.

Issues that were mentioned by participants were as follows: one participant said that he forgot to click the process buttons or realised too late that he had switched to a different process. Another one criticised the necessity to move the mouse to click buttons because this caused a disruption and suggested using keyboard shortcuts instead – an option we had considered, but in the gives setting ist was unfortunately not possible to capture keystrokes. Overall, participants felt moderately interrupted by the secondary task (average 3.17 with 1 = “not interrupted at all” and 5 = “interrupted completely”).



(a) Average relative starting time of a process, per document

(b) Standard deviation of relative starting time of a process, per document

Figure 4.5: Starting times of processes

On the other hand, all but one participant found it easy or very easy to associate their actions to the four processes (average 4.25 with 1 being "very difficult" and 5 being "very easy"). One also liked the fact that by clicking the buttons she was already made more aware of what she was doing during writing and felt that this could improve the habit of writing.

4.3 Discussion

4.3.1 Consequences for Process Detection

Regarding the process detection, some patterns that were suggested in literature, e.g. higher likelihood of reviewing processes to occur towards the end of a text composition, could be confirmed.

The extraction of patterns for process occurrence indicates which process is more likely at a given time, and a further potential factor was the length of writing pauses. But the differences between individuals and from one document to the next also show that we will not achieve precise predictions only based on this. The information is thus not sufficient to clearly determine the active process. Moreover, the question of granularity remains: is a subprocess of a top-level process to be considered separately or as a part of the top-level process? Therefore, for more precise predictions, it is probably necessary to create personalised models based on a number of texts written by one writer and to include more measures.

4.3.2 Consequences for Needs Assessment

The tendency that more experienced writers (with respect to the number of scientific articles written) deleted more text, and more in one go, could be explained by

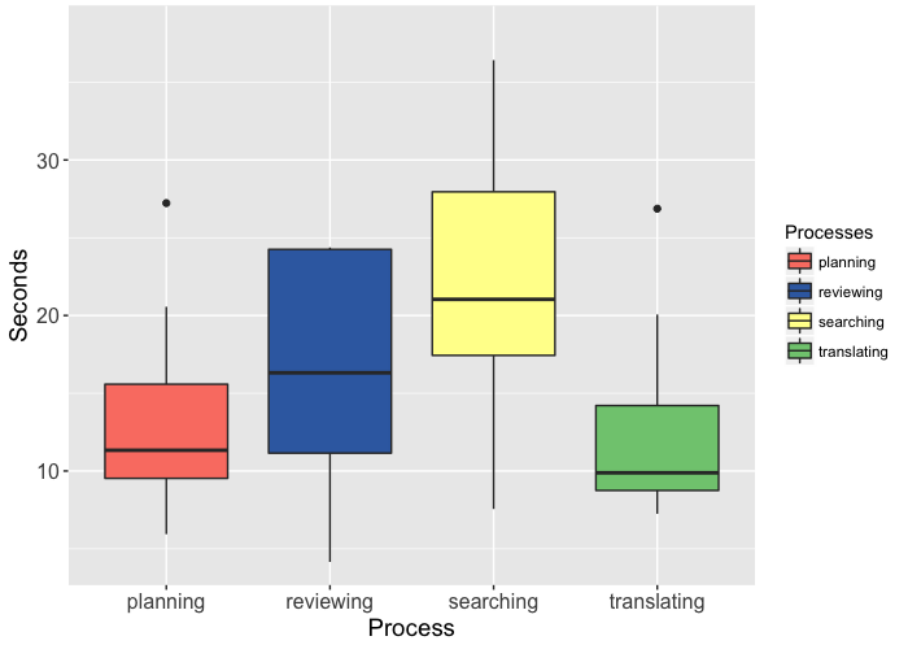


Figure 4.6: Mean duration of writing pauses during a process, per document

experienced writers being more likely to review their compositions and in more detail: larger deletions suggest more profound rather than superficial ("cosmetic") changes. In practice, we could then use average count of deleted characters, for instance, as a basis for tool selection: if it is small, this is an indicator that the writer needs more help with in-depth revision.

This is only an example for what can be derived from actions during the text composition. Especially personalised analysis is likely to reveal several other possible indicators for beneficial support.

Chapter 5

The Writing Support Add-On

In Section 2.3, we reviewed available writing tools. We found that they generally focus on one specific writing process, and even more often just a small fraction of possible support within this writing process. Our add-on, on the other hand, aims to support the entire process of writing, and integrated seamlessly into the working environment, so that the writer does not need to switch context when undertaking one process.

Based on the needs identified in the previous section and applying the knowledge about writing explained there, we developed a prototype for an integrated writing support system in the form of a Google Docs add-on. It consists of four parts, each dedicated to one of the writing processes *searching*, *planning*, *translating*, and *reviewing*. Each process requires different type of information, and having a tool for each helps to focus the tool in supporting that process.

Each tool set is displayed in a separate tab and only one tab is visible at a given point in time. Thus, it is possible to adapt the state of the add-on to current user needs. So far, it has not been possible to develop an accurate model for process detection and needs assessment yet and it would be far beyond the scope of this thesis to do so. At the current stage, it is therefore up to the user to decide which tool she would like to activate when and to select it by opening the respective tab. However, the design of the add-on makes it easy to integrate a process model in the future.

Our main focus areas were possible tools for each process, interaction with them and a logical composition in general – using the findings in cognitive psychology described in Section 2.1.1 – and embedding them in the Google Docs environment. These aspects were given preference over a more elaborate design of the graphical user interface and perfection of the various tools; these still have prototypical character. In this chapter, we will present all the tools we implemented, grouped by the process they support. We will relate our concepts to the requirements presented in Section 2.2. Chapter 6 will provide details on the implementation and explain decisions we took.

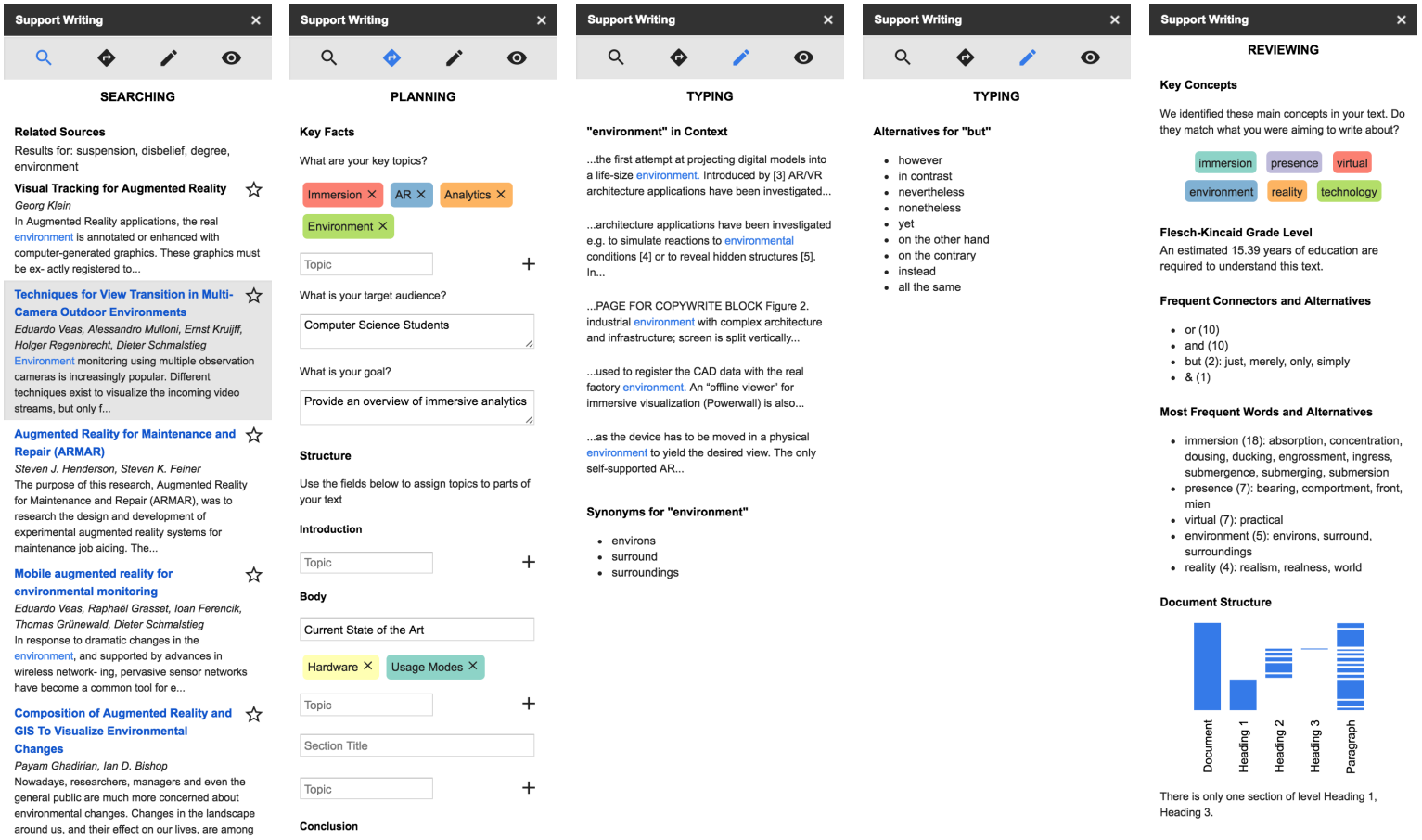


Figure 5.1: Screenshots of the tools for each writing process

5.1 Search Tool

The search tool gathers and displays recommended content. It extracts keywords from the current paragraph in the text written so far and matches them with documents in a database; the most relevant documents are displayed as a list, including title, authors and the first part of the abstract.

Recall that the specific requirements we formulated for searching were the following (cf. Section 2.2):

1. overview and detail modes for skimming and exact reading
2. easy and quick access to sources
3. relevance and balance of suggested sources

We address requirements 1 and 2 by two-level interaction: If the document is associated with a URL, clicking on its title opens the URL in a new tab. Thus, we make it possible to both scan the titles and short description to check if the suggested documents are likely to be relevant at all, and to access the full text with only one further click. This caters to information needs at different levels of detail and minimises the cognitive and physical effort required to check sources. This type of technique is also presented by Rhodes [Rhodes, 2000]: "One display technique is [...] a "ramping interface," where information is conveyed in stages. Each stage of a ramping interface provides a little more information, at the cost of requiring a little more attention to read and understand it."

We obtain up-to-date information by triggering the search proactively, more specifically whenever the text in the current paragraph changes or the cursor moves to a new paragraph – the paragraph is then also used as a basis for the keyword extraction. This is also the principle Melguizo et al. used in the IntelliGent system [Melguizo et al., 2009], cf. Section 2.3.1. To make sure that the searching tool is not too disruptive, we use a simple blue loading indicator which is visible whenever a new search has started. Its movement suggests that there will be new information available, but does not force the writer to look at it – or as Maglio et al. put it: "some motion actually provides effective feedback, helping users efficiently schedule glances at the peripheral display" [Maglio and Campbell, 2000]. The fact that the add-on is displayed as a sidebar in an otherwise unused area of the editor window also supports continued focus on the task.

Another feature we implemented are bookmarks: Items can be bookmarked by clicking a star icon on the right. This means that when the search is refreshed, they are maintained in a separate list at the bottom so the user can keep them for later usage. This also avoids situations where a possibly valuable source cannot be located anymore.

When there is no text yet, the topics entered in the planning tool (see below) are used as keywords. However, this feature was only added after the usability study

and could therefore not be evaluated.

The search tool can easily be connected to various document bases as long as they provide an API and provide information such as title, authors, URL, and abstracts (it is then up to who chose the data source to check that point 3 is fulfilled). In our full version, we used the uRank recommender system developed by di Sciascio et al. [di Sciascio et al., 2015] and before that also EEXCESS[1].

For our user studies, however, it was important to ensure a controllable environment and we decided to replace the connection with a search engine providing a fixed set of documents for each topic. Keywords were automatically extracted from these sets of documents and TF-IDF scores were used to rank the documents for a given query.

5.2 Planning Tool

A good planning tool should make writers engage in planning and should make them aware of what they want to achieve (see 2.2). In our add-on, we ask users to note key facts about the text to be written. There is the possibility to enter the key topics and answer questions about the intended target audience and the writer's goal. The idea is that clearly stating these facts will (a) make the planning more conscious, like the cue cards described in Section 2.3.2, (b) help the writer keep them in mind during the composition and (c) make it easier to check later on if the requirements were fulfilled.

The second part of the planning tool supports structuring the text by providing a frame with introduction, main part, and conclusion. For the main part, there is also the option to add sections for more detailed planning. For each part, the user can add a number of keywords. The resulting structure can easily be used as an outline and a point of reference during the development of the text – thus fulfilling our second requirement for planning tools.

The structuring could also be extended based on the type of text and suggestions could be given for the general outline. A conference paper, for example, would have a different outline than a thesis. For a part of our usability study, we already did something similar by adding predefined section headings which participants were supposed to use for guidance.

In the section on goal-setting (2.1.4), it was mentioned that goals are continuously changing over the course of the text production. One example was that a writer knows more about one topic than another one and gives it more room and importance than originally planned. This is very likely to be a subconscious process and the writer might not even realise what happened. In this case, noting down the main topics of each section also helps to become aware of such changes and to compensate for them.

[1] http://eexcess.eu

It has to be noted that due to data storage issues in Google Docs add-ons (cf. Section 6.1.3), the content of the planning tool is currently lost once the add-on is closed. For a single writing session, this could be circumvented by storing its content in the document cache. However, for long-term persistency, all user data has to be stored externally, e.g. on a server, and retrieved on starting the add-on.

5.3 Translation Tool

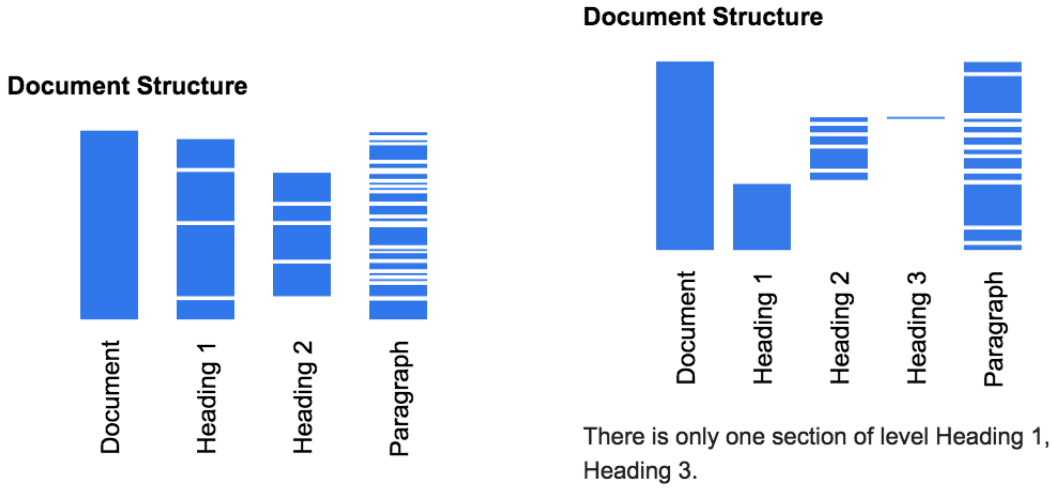
A translation[2] tool should, first and foremost, help writers to generate text from a structured collection of ideas, but also to connect in such a way that the structure and message become apparent for readers. Concrete requirements listed for translation tools were twofold: they should react quickly, and they should ideally provide support not only for spelling, grammatical or mechanic issues but also for higher level aspects such as coherence.

Our translation tool has two distinct operation modes depending on the current cursor position. If the cursor is positioned in or directly after a connector, it suggests other connectors with a similar meaning. Similarly, if a keyword is selected, the sidebar shows synonym suggestions retrieved from the WordNet corpus [Kilgarriff and Fellbaum, 2000]. On top of that, there is also a section showing how this specific word is used in context: a list of actual usage examples gathered from a number of documents (in our case the same ones as for the content recommendations in the search tool) is displayed.

We achieve instant feedback by updating the content when the writer's cursor is moved to a new word. The semantic support, though, is more difficult to put into practice, because most solutions demand complex analyses. Suggesting alternative connectors does not require any of this, since the choice and variability are limited. Nevertheless, this tool has great potential: it can help writers increase the variety of their text and make it more fluent. Furthermore, it is possible that a suggested solution is a better fit and makes a text more coherent than the connector which was originally used, especially when a writer was not entirely sure how to connect phrases. The same also holds for synonyms. It is particularly valuable for L2 writers who are likely to be less confident and knowledgeable when it comes to vocabulary and sentence connections. Shaw, for example, found that non-English-native PhD students resorted to thoroughly studying English source texts to achieve better command of the subject register [Shaw, 1991].

The context examples follow a similar approach in the sense that we do not need full understanding of a word's meaning. Instead, we rely on the knowledge other writers have applied before when they wrote sentences containing the word a writer is currently typing. Context examples show how a term is usually used in a sentence

[2]In the add-on, translating is described as "typing" as the term "translating" is generally understood as translation between two languages and not "thoughts-to-letters". We wanted to avoid confusion regarding this.



(a) A balanced document structure: never only one child and a consistent hierarchy of section levels.

(b) An unbalanced document: there is only one very short child of section level 3, the level 2 part has no level 1 parent.

Figure 5.2: Document structure visualisation

linguistically speaking. Furthermore, they give an impression of the topics a term tends to be associated with. On a content level, the usage examples can therefore work as a correcting instance (was the term applied correctly?) and as a source of inspiration (how did others use this term?).

To conclude, while it is hard to identify semantic issues computationally, we can increase the writers' awareness and give them first ideas for changes.

5.4 Revision Tool

Revision is important from both a linguistic and a content point of view. We stated that a revision tool should help the writer check that the vocabulary and style are adequate for the target audience, that it should make her aware of potential style and cohesion issues. In addition, problems with the text structure should be identified and the writer should generally be aided checking if goals were met.

We therefore have a number of independent aspects, for which we decided to design a collection of smaller tools to work on different aspects of the text. We cover linguistic, structural and content-related issues.

The first set focusses on checking if the content goals that were set for the composition were met. To this end, we extract the main keywords from the text which can then be compared to the topics added in the planning tool. When the mouse hovers over the key words, their occurrences are highlighted in the document text. The key concepts give an indication whether the predefined topics were actually discussed. Were the points that a writer wanted to focus on also more prominent in

the final text? Are there any topics which were not planned explicitly but show up as a key topic? If the two sets of concepts are not aligned well enough, this can be a cue for what to change when going back to the text.

In the language section, we compute Flesch-Kincaid Grade Level Score, which helps determine if the readability of the text is adequate for the intended target audience. For an explanation of this measure, please refer to Section 2.3.4.

In the same part, we also have word-based style review. Both the most frequent terms and the most frequent connectors are listed, together with a list of synonyms. If a text contains the same words too often, it easily sounds repetitive and consequently boring, therefore it makes sense to replace them with equivalent formulations – unless they are topic terms for which there is often no straightforward alternative available. In addition, imagine a situation where a user can only think of a word that comes close to what she wants to say. It may well be that amongst the suggested synonyms for this word there is a better one which expresses her idea more clearly. Thus, the suggestion of synonyms can help make a text more precise and therefore improve its quality. The frequent words are also highlighted in the text on hover.

Finally, structure analysis is done via an overview of the paragraph types as they are used in the document: Google Docs offers the option to mark a paragraph as hierarchical headings as well as normal text. The paragraph hierarchy is analysed by iterating over all paragraphs and identifying their parents. Then we can count how many children each paragraph has and check the difference between their levels. Looking at the consistency of the paragraph markup then gives insights into the structural planning: Are there any sections which only have one child section? Are the heading styles used at all, and if so also in the correct order, e.g. no type 1 heading is directly followed by a type 3 heading? Is one section a lot longer than all the others? Figure 5.2 shows examples for a balanced and an unbalanced section division.

5.5 Study Versions of the Add-on

For both the formative study and later the evaluation prototype, we used variations of the add-on described above. In the formative study, only a bare version of the add-on was used, and in both studies, a workflow including pre- and post-study questionnaires as well as usage data recording were added. For details, please refer to Sections 4.1 and 7.1 respectively.

Chapter 6

Implementation Details

The Writing Support add-on was implemented as a Google Docs add-on because the Google Docs environment provides a platform-independent, full-blown, yet easy-to-use graphical user interface for text editing. It requires no installation because it is run in a browser and offers many possibilities for developers. In addition to the Google Docs part, the add-on also relies on a Node.js server component and a MongoDB database and communicates with external services. An exemplary request which encompasses all add-on components is shown in Figure 6.1.

Key features of the Google Docs environment and add-on development are described below, more information and introductions to the the other frameworks and technologies we used can be found in Appendix B.

6.1 Google Docs and Google Apps Script Add-Ons

Google Docs[1] is an advanced writing environment fitted into Google Drive, cloud storage associated to Google accounts, and the G Suite, Google's solution for office applications. Its functionality is similar to that of Microsoft Word, but it is executed in the browser. Documents are stored online and they can easily be shared and worked on collaboratively.

6.1.1 Google Docs Add-Ons

Google Docs can be extended through add-ons, bundles of scripts that provide additional functionality. There is a wide range of available add-ons, popular examples are:

- Speech Recognition[2] makes it possible to dictate a text rather than typing it with a keyboard.

[1] https://www.google.com/docs/about

[2] https://chrome.google.com/webstore/detail/speech-recognition/idmniglhlcjfkhncgbiiecmianekpheh

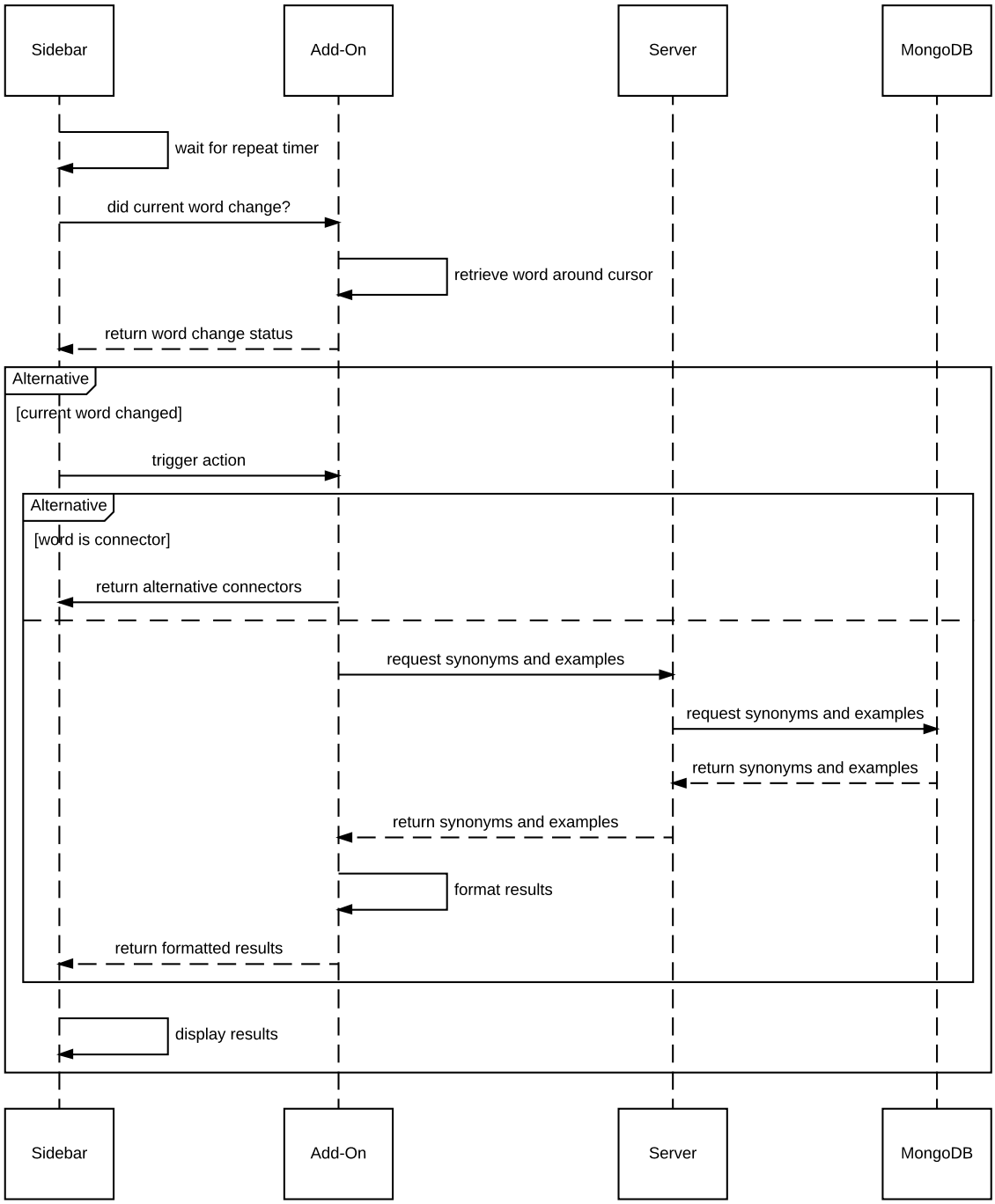


Figure 6.1: Sequence diagram for the translating tool

- EasyBib Bibliography Creator[3] lets you manage a list of citations and add them to your document in the desired format
- Easy Accents[4] improves language support by presenting a list of language-specific characters that are not easy to find or missing on your national keyboard.
- With Lucidchart Diagrams[5] you can create, for example, UML or mind map diagrams and add them to your document. The sequence diagram in Figure 6.1, for example, was created with Lucidchart,

You can see that these add-ons cover a wide range of functionality and improve productivity in very different ways. They are distributed via the Chrome webstore[6].

6.1.2 Features

The main part of an add-on is always the Google Apps Script (GAS)[7] back-end. GAS is a programming language, a subset of standard JavaScript extended with document-specific functionality and a couple of convenience methods, e.g. for HTTP requests.

In addition to the GAS part, the following interface components can be added: (1) menu items in the main window of the editor, (2) an HTML sidebar and (3) popup dialogues.

The GAS scripts are executed in a non-blocking way on Google servers. The content of sidebar and popups, on the other hand, are generated from HTML files which are executed on the client-side and displayed in an iFrame. Here, the browser's JavaScript engine can be used.

Add-ons have reading and writing access to the document. It is possible to insert content, format it (highlighting, bold print,...), and to add comments. Furthermore, there are methods available to retrieve information like the current cursor position, selections, and paragraphs including their style (e.g. "Heading 1", "Normal text"). If the user gives permission when opening the add-on, the add-on may access details such as the document ID, user ID and revision history. Document, user, and script cache allow storing a certain amount of data with different accessibility levels.

[3] https://chrome.google.com/webstore/detail/easybib-bibliography-crea/dnhomniofbmbomomggjpkakilbbgfkhc

[4] https://chrome.google.com/webstore/detail/easy-accents/njbfdnnpglelehliilonpfcbckcdhnid

[5] https://chrome.google.com/webstore/detail/lucidchart-diagrams/klhpiejgjijjcglpcaanggahphcgckpm

[6] https://chrome.google.com/webstore

[7] https://www.google.com/script/start/

6.1.3 Limitations

The google docs add-on environment is rather restrictive for developers. On the one hand, it is possible to assure security and ease-of-use, but on the other hand, this imposes some important limitations on developers, particularly in the areas interactivity, communication, library usage and data storage. Below, we will describe the issues that were relevant for the present work and for which we had to find solutions that unfortunately made the architecture of the add-on more complex.

Interaction between Add-On and Docs Editor

An add-on has only little influence on what happens in the main editor window because interface manipulation is limited to the predefined sidebar, menu items, and modal windows. Scenarios such as dragging-and-dropping of text blocks into the text or adding custom markers, e.g. to identify and edit citations in a document or to associate a paragraph with its corresponding item in a draft, are therefore not possible.

Moreover, an add-on cannot track everything that happens in the editor window. It is, amongst others, not possible to capture keystrokes in the editing area or to get the scroll position.

This makes reacting to user actions and integrating the add-on in the workflow a very challenging task and individual solutions have to be found for seemingly simple functionality. It also complicates the design of a generic experimentation framework. Most studies of computer-aided writing use keystroke logging as basis for analysis.

Communication between iFrames and Google Apps Script

HTML user interface components and the Google Apps Script part cannot communicate directly since they are executed on different sites. Data exchange is therefore solved through requests in the HTML file in combination with success and failure handlers. The exemplary code snippet below illustrates the interaction for retrieving the document length.

- Google Apps Script side

```
1 function getDocumentStatistics() {
2   var length = DocumentApp.getActiveDocument().getBody().
      getText().length;
3   return length;
4 }
```

- HTML

```
1 <!DOCTYPE html>
2 <html>
3 ...
4     <script>
```

```
5
6           function onSuccess(length) {
7               console.log(length);
8           }
9
10          google.script.run
11              .withSuccessHandler(onSuccess)
12              .getDocumentStatistics();
13      </script>
14
15  </html>
```

You can see that the interaction is started by the sidebar, on the client side. It is not possible to create custom communication which would be triggered by the GAS component, e.g. the latter cannot send a message telling the client that the user has added new text and the sidebar can update itself. It would only be possible to replace the entire HTML and that would mean losing the state of the sidebar.

Communication between iFrames and External Services

From within iFrames, data can only be requested using the secure HTTPS protocol, and not via HTTP. Not all APIs we used set up for secure communication – therefore requests had to be routed through the Google Apps Script component.

Google Apps Script Language

As mentioned before, GAS only allows for a subset of JavaScript. In particular, the function `setInterval` cannot be used. We relied on tracking status changes, though, and therefore used `setInterval` in the sidebar, and then triggered calls to the GAS part from there.

Similarly, external JavaScript libraries cannot simply be embedded through their URL. For keyword extraction, we relied on Natural Language Processing, but to use it directly in our backend component, the entire library we used would have to be ported to GAS. Then it could either be added to our project or referenced from another GAS project. Because of this, we used our Node.js server to execute all keyword queries.

Data Storage

A Google Docs add-on does not save its state when it is closed. For data persistence, it is necessary to either reload data from the document cache, which is, however, cleared after a certain time, or from an external server. We used the cache for data storage within the session and propose using an external database for future extensions.

Despite the limitations being quite restrictive, we still think they are outweighed by the advantages of using Google Docs. Especially the accessibility without installation, its ease of use and quick learnability from the developer's point of view make it a very suitable environment, not just for prototyping and evaluation.

6.2 Other Components

The Node.js server and MongoDB database were mostly used for extracting keywords from the document text, for retrieving synonyms, and for computing text statistics such as the Flesch-Kincaid grade level score. For our studies, we also used it for gathering usage statistics as well as preparing the document set for the search tool. Node.js and MongoDB are introduced in Appendix B.

6.3 Modularity and Extensibility

Modularity and extensibility were key aspects when designing the architecture of the add-on. For the HTML sidebar, this means that each tab is independent from the others and can be replaced, removed or added with very few code changes. In the Google Apps Script "back end", functionality is separated into different files: files for each writing process and some for common operations such as the extraction of keywords from a text or the detection of the cursor position. Finally, the Node.js server implements separate REST calls for each operation, e.g. the retrieval of synonyms for a term or the calculation of the Flesch-Kincaid grade level score.

Furthermore, it is also easy to integrate external services. We have successfully used information retrieval using the EEXCESS API (cf. Section 2.3.1) and uRank [di Sciascio et al., 2015] and we also experimented with external synonym and context providers. The only requirement is a REST interface; POST or GET requests are started from the GAS component.

Chapter 7

User Study

In order to evaluate the usability of the Writing Support add-on and also to continue our research regarding writing processes, we conducted a user study. Regarding the writing behaviour, we focussed on finding out when, how often and for how long writers would select each tool. Regarding the usability, we wanted to know which tools the participants found helpful, which ones they liked least, and what they felt was missing – thus gathering ideas for future developments.

7.1 Experimental Setup

In this second study, the participants were asked to write a 2-page summary on one of two topics: *Immersive Analytics*, or *Knowledge Technologies (KT)* with special focus on *Artificial Neural Networks*. We made sure that they would have the opportunity to gain the necessary knowledge before the writing task: the first group of participants consisted of students in a Visual Analytics class where the lecturer Dr. Veas gave a talk on Immersive Analytics, the second of a group of researchers who attended a very similar talk, and finally a group of students who attended the lecture "Introduction to Knowledge Technologies" ("Einführung in die Wissenstechnologien") offered in the summer term 2017 at TU Graz. The participants also got a short introduction into the Writing Support add-on, including a brief explanation of each tool.

Afterwards, they were asked to use one of the prepared documents to write a summary of the lecture or talk while using a modified version of the add-on. In essence, the differences to the original add-on were as follows:

- Just as in the formative study, upon opening the add-on, a popup leading to the pre-study questionnaire was opened.
- Once the survey was completed, this popup also displayed instructions for the writing task.

- The sidebar was the same as in the original add-on with the exception of "Pause", "I'm Done", and "Info" buttons being added at the very top. "Pause" temporarily stopped the recording so the participant could take a break. "I'm Done" was to be clicked on once the text was complete and "Info" opened a tab that summarised the instructions.

- After the writing, the text in the sidebar changed and prompted the participant to fill in the post-study questionnaire.

Since the number of participants was limited, we did not split the participants into an experimental and control group with and without add-on usage, but only evaluated the performance with the add-on. Comparing performance with and without the add-on would, however, be a recommended step for future work.

7.1.1 Participants

In total, 22 people participated in the study, although one did not answer the questions in the pre-study questionnaire and one missed the post-study questionnaire. 15 were attendees of one of the lectures mentioned above and mostly had only a high school degree; the remaining 7 were researchers and we will consider them experienced writers. Almost all participants stated computer science, software development, or closely related topics as their field of study or area of research. Ages ranged from below 20 years to 40–49, the majority was younger than 30.

We also asked them to self-assess their English proficiency, since the reports were to be written in English and we wanted to know how confident they felt using it. Only two subjects stated their level was basic, four said it was intermediate, 14 advanced, and one was a native speaker.

All researchers stated they were very familiar with the Google Docs environment. Among the students only three stated that they were not familiar with the environment or never used it, all others said they had a (fairly) advanced knowledge.

7.1.2 Collected Data

As long as the participants were in the writing stage, the following items were recorded at regular intervals of approximately 200 milliseconds and submitted with the addition of document id and a timestamp:

- full text of the document

- cursor position within the current paragraph

- selected text and selection offsets

In addition, the following events were registered:

- tool selection, that is, clicks on the tool buttons in the tab bar of the sidebar
- clicks on the pause or resume button
- clicks on links in the search tool
- hovering over a term, which triggered highlighting of that term in the text. This term could be either a concept (topic) or a frequent term, this *type* was also submitted.

The pre-study questionnaire was used to collect demographic information and to assess characteristics of the subjects' writing behaviour. We asked them, for example, at which points of a text composition they would usually plan, search, and review; which tools they typically use to search for information, how many scientific articles they had already written etc.

The post-study questionnaire contained questions about the usability of the add-on in general as well as specific questions about the tools. It also asked if the participants found it easy to write their text and if they were satisfied with the result.

Most questions in both pre-study and post-study questionnaire were presented as a Likert scale with seven options, so participants could also choose a neutral option. Only the two outer options were labelled, suggesting equal differences between the possible replies.

Finally, we used a rating form to manually assess the writing quality of all texts regarding their structure, content, language, vocabulary, and grammar. In addition to that, we extracted metrics such as the vocabulary size, the ratio of specific types of words, and readability scores with the open-source tool expresso[1].

7.2 Analysis

In addition to insights into the performance and usability of the add-on, the various types of data we gathered also enabled us to look at different aspects of the general writing behaviour. We will start by outlining the characteristics we found and then continue with a more thorough analysis of the writing tool itself. Further plots of the data we collected can be found in Appendix C.

7.2.1 Writing Behaviour

Participants spent on average one hour and 56 minutes using the add-on, the minimum was 9 minutes (but some additional writing took place without the add-on in this case) and the maximum a bit over six hours. They produced texts of approximately 250 to 1370 words. Most of the texts had around 800 words. On average,

[1] http://www.expresso-app.org

the attendees of the KT lecture took around three times as long as the researchers to finish their work.

To find out more about which activities participants typically perform, we used the pre-study questionnaire to ask specific questions about writing habits. Some key points were as follows:

- Almost all participants used general search engines to search for information, one third also scientific search engines. A much smaller number, 42.8%, usually refer to books.
- At 61.9%, the most commonly used drafting strategy is to create a basic or detailed outline, but some participants also stated they make a list of keywords or a mind map.
- A third of our participants stated that when planning, they do not take the audience to account at all or only pay little attention to the audience.
- For the writing part, 85.7% said they used spell checkers, 57.1% also grammar checkers, 47.6% synonym suggestions and 28.6% translators.
- For reviewing, it can be summarised that most participants state they normally check the spelling, style and consistency of the structure at least to some degree. Less than half use external tools such as grammarly[2] to review their texts. Some participants seem to check how well their text fits the draft, but some do not look at this aspect at all.

We also asked participants at what point during the text composition they would usually create a plan, search for information, and review their writing (see Figure 7.3). The responses were varied overall, but we found clearer patterns for planning and searching. The majority said they plan before they start writing, seven participants stated only that they define the content and structure as they write. Only four chose more than one option, suggesting that the planning activity is concentrated at a specific point in time. The tendency for searching was also towards the very beginning or before starting a new section, but especially the more experienced writers stated they additionally look up information in the middle of a section. Most writers review after local completions and when the entire text is written, some also add revision steps in between.

Furthermore, there were positive correlations between the number of articles our writers had already written and the number of "time slots" they selected for each process: 0.45 for searching, 0.51 for planning, and 0.61 for reviewing respectively.

7.2.2 Tool Usage

In our activity analysis, the first step was to look at tool usage. We did this in a similar manner as we had analysed process occurrence by considering the selection of

[2] https://www.grammarly.com

Figure 7.1: Normalised times where a tool was open, areas without fill mean no tool was selected. The shapes indicate points of interaction with the tool, e.g. clicking on a link in the search tool. The bar on the right side indicates the final document length. Recording pauses (not writing pauses) of more than 10 seconds were removed.

a tool instead of clicks on process buttons. Please refer to Figure 7.1 for an overview of the tool selections and the interactions we recorded. These two measures are of course not the same, since a tool can be open regardless of the current activity. Figure 7.2, for instance, shows an example where a major part of the typing happened while the reviewing tool was opened. Nevertheless, a comparison can give us useful insights into the behaviour with and without add-on and the needs of writers.

Figure 7.5a shows how often each tool was chosen, Table 7.1 and Figure 7.5b how long the tools were open. One participant did not open any of the tools (apparently, he did not look at our instructions or did not want to follow them – one of the issues with unsupervised studies), and one only three. The rest opened each tool at least once.

Just as in the formative study, participants had very different strategies and some frequently changed tools, while others used mostly one. In addition, some tools were generally opened earlier than others. Figure 7.4 shows the distribution of the selection times of each tool averaged per document. A Wilcoxon rank sum test revealed that, for example, the planning tool was selected earlier than both the search tool ($p < 0.01$) and the reviewing tool ($p < 0.01$).

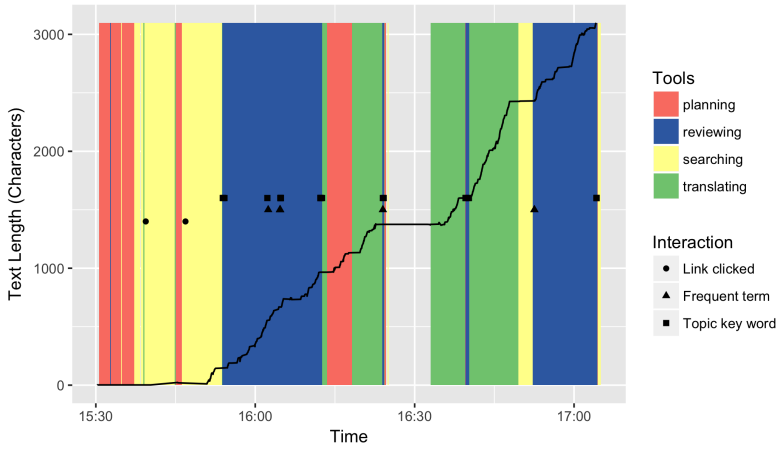


Figure 7.2: Development of text length, selection of tools, and interactions for one sample document

Tool	Time Share
None	19.8%
Searching	36.9%
Planning	15.5%
Translating	19.4%
Reviewing	9.1%

Table 7.1: Percentage of time that each tool was open

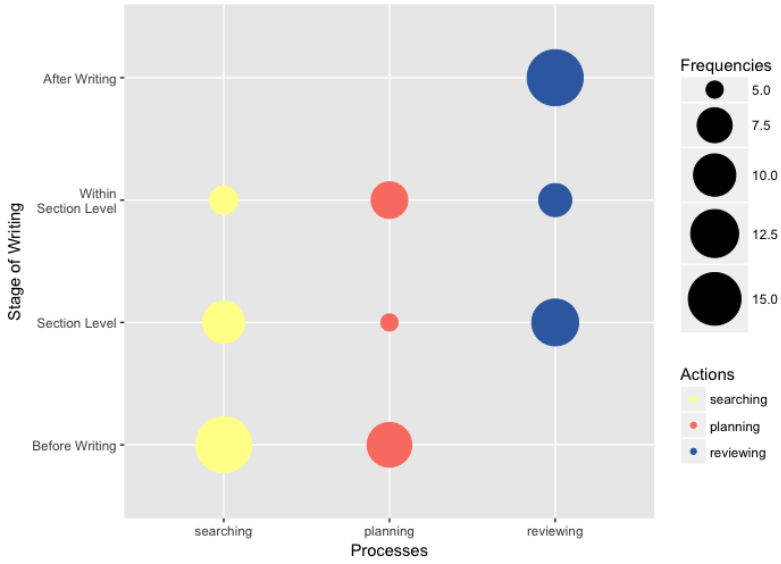
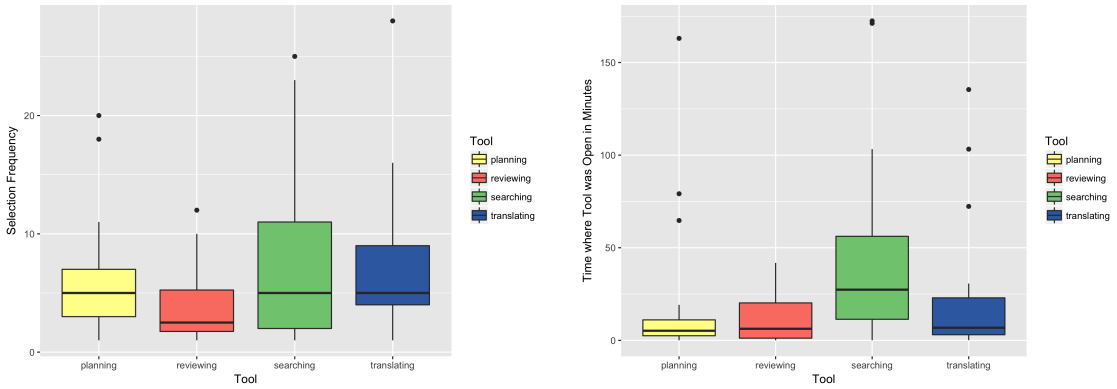


Figure 7.3: Writing stages where participants stated to typically perform actions. Translating was used as a reference and is therefore not included. Multiple answers were possible

Figure 7.4: Average relative opening time of a tool, per document



(a) Tool selection frequency, per document **(b)** Tool selection duration, per document

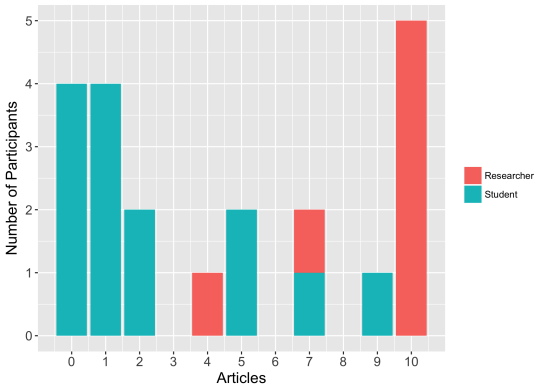
Figure 7.5: Tool selection

Another interesting point was that in the formative study, participants selected a process on average 59.8 times, but when using the add-on, they only clicked on a tool 24.4 times.

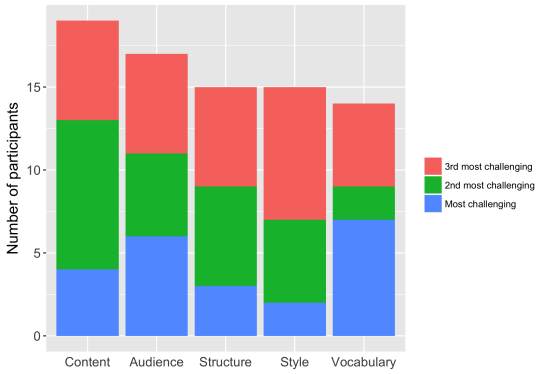
7.2.3 Writing Skills and Text Quality

Assessing writing skills was an important aspect for us before starting to analyse the quality of the texts our participants wrote. To get an impression of their experience in scientific writing, we asked them how many scientific articles or reports they had written prior to the study. Not surprisingly, this number was a lot higher for the researcher group. They had written 8.7 articles on average, and the students 2.3, please refer to Figure 7.6a for an overview.

In addition, we asked participants to pick the three most challenging of the



(a) Number of scientific texts written by study participants



(b) Challenging aspects in the writing process

aspects *content*, *adaption to audience*, *structure*, *style*, and *vocabulary*, the results are shown in Figure 7.6b. Content was mentioned by 19 of the 21 participants, and the audience by 17 – more support in these areas is therefore likely to be beneficial.

Moreover, there were some questions to self-assess several factors:

- the ability to structure a factual text
- the ability to build paragraphs in a factual text
- how likely they are to perform a number of actions, e.g. if they tend to verify the consistency of their structure when reviewing or if they used external revision tools

There were a number of points where "better" writers obtained higher scores. For example, participants who were more likely to review the style of their composition produced less long sentences and people who had written more articles used less weak verbs. The researchers were better at spelling.

7.2.4 Usability

The responses of the post-study questionnaire gave us useful feedback on the features and usability of the add-on.

For the question "I found the add-on was distracting", the option "I strongly disagree" was the one which was chosen most often, the mean was at 3.1 (1 = strongly disagree, 7 = strongly agree). At the same time, most participants said they were paying attention or even close attention to the add-on while writing. Furthermore, 45% of participants strongly agreed to the statement "I found it easy to understand how to use the add-on", 72% generally agreed. The mean was at 5.1 (1 = not easy at all and 7 = very easy). Similarly, we got a mean of 5.0 for good organisation of the add-on (1 = not well organised, 7 = very well organised).

The responsiveness obtained a mean rating of 4.3 (1 = not responsive at all, 7 = very responsive). Three subjects said it was not responsive at all, although two of them also mentioned technical issues. For one the bookmarks were apparently not working, one reported errors in the typing tool and slow updates.

Search Tool

As becomes apparent from Table 7.1, the search tool was the one participants used most. Its share was nearly 37% of the entire composition time. How they used it, though, differed individually. On average, they clicked on 2.7 links, but 10 clicked on none while one selected 13.

It seems that this also depended on how useful participants found the displayed sources: there was a correlation of 0.63 between how relevant they found the information for their writing to how much they said they had used it. Overall, the perceived relevance of items was slightly positive (mean of 4.1, 1 = very irrelevant, 7 very relevant). Some participants mentioned reasons why they did not find the selection so useful: an attendee of the Knowledge Technologies lecture said he would have preferred a wider range of sources, a researcher said that for experts the retrieved content was insufficient and not really new.

There were several suggestions for improvements. One of the researchers said, for instance, that he would like to be able to add references from the sources. From the technological point of view, there were requests for a more responsive implementation and adjustments of some details like the bookmarking system.

Planning Tool

The planning tool was open around 15% of the time. Five subjects stated they answered all the planning questions, and two none of them. For the KT lecture attendees, the given structure seems to have been helpful: one of them said the "rough structure of the text which should be written [...] was helpful to understand what is required". Helpfulness for the organisation of the writing task scored 4.52 out of 7.

Translation Tool

At 19% percent of the writing time, the translation tool was the second most used. Despite this, the average rating of 3.7 for the information in the tool was only moderately positive (1 = did not help to find information to use, 7 = helped a lot).

There were some very encouraging comments on the functionality, e.g. "The idea behind the tool is awesome", "The results are interesting as you can see the word in context", and "It helped me for not repeating some words".

Revision Tool

The revision tool was by far the busiest interface, with five different sections. Despite this, with a mean of 2.8, participants did not perceive it to be very cluttered or difficult to understand (1 = not cluttered at all, very easy to understand, 7 = very cluttered and difficult to understand).

All parts of the tool received positive or slightly positive ratings (all with 1 = not helpful at all, 7 = very helpful). The information about frequent words and synonyms obtained an overall rating of 4.4; this functionality was also used quite frequently by some writers with an average count of 5.1 highlights per document and a maximum of 31. Key concepts were highlighted even more often: 9.4 times on average, and a maximum of 51 in one document. At 4.0, it was also helpful for a majority of participants to align the topic of the text with the plan.

The Flesch-Kincaid grade level score as a support for audience awareness obtained 4.2. Two participants commented they had not seen the score before. Identification of weak points was rated with 3.9 out of 7 and overall support of the goal checking with 4.1 – a good rating for certainly non-trivial tasks. Finally, the overview of the document structure received a rating of 4.0.

In addition, writers who often highlighted frequent terms or key concepts were more likely to go back to planning and check they were going with the plan: we found Pearson correlations of 0.52 and 0.50.

7.3 Discussion

7.3.1 Writing Behaviour

From the writing behaviour, we draw the following conclusions:

- Most external information was accessed digitally. That makes it easy to directly retrieve them via an add-on and it also means that by using digital sources, it is possible to cover a wide range of the sources used.

- Creating an outline was the most used planning strategy – and our planning tool provides basic outlining support with the possibility to add keywords to sections.

- The fact that many participants do not seem to pay close attention to the target audience is a good motivation for adding a question about this, as we did in the planning tool. Raising awareness for goals is a very important point.

- The translation support participants were most likely to use, spell checking, is already a part of the Google Docs editor. Synonym suggestions are included in the add-on. Language translation would be a recommended extension, particularly for L2 writers.

- In the reviewing area, we found that in particular the checking of the draft does not happen as often as it could. This is a point that would deserve more attention, also in future versions of the add-on.
- We saw that more experienced writers (with respect to the number of scientific articles written) were more likely to plan, search, and review at more different stages of the writing. This suggests that more experienced writers are more flexible in their organisation of the writing process and can probably also react better to current requirements, e.g. if they see the need to get some more information in the middle of a section.

7.3.2 Tool Usage

In the formative study, searching had generally occurred earlier than the search tool was opened, but there was also no search functionality integrated into the interface. In addition, the share of the search time/tool now increased from 16.3% in the formative study (cf. Table 4.1) to 36.9% with the add-on. While a part of this change may be due to the different topic participants wrote about in the usability study, the difference is still large. This suggests that having external sources readily available within the working environment really does lower the threshold for writers to use them – see also Rhodes' experiments where users of his JITIR accessed almost three times as many documents as those using only a search engine [Rhodes, 2000].

In addition, participants probably did not select a tool whenever a new process became active. This indicates that writers might not desire very frequent changes of the interface – a point to consider when automatically suggesting tools based on the process and also when designing an interface.

7.3.3 Writing Skills and Text Quality

A very important question was of course how the writing experience skills related to the final text quality. A major problem here was the reported time difference between the researchers and lecture attendees. Furthermore, the students in the KT lecture probably also had more motivation to hand in polished texts: they were rewarded bonus points for an upcoming exam. So while the researchers were more experienced, they still delivered texts which were written in a rush and with too little revision.

This also limits the usability of writing experience and text quality as distinguishing features for discovering patterns in the writing behaviour and tool usage data collected with the add-on.

7.3.4 Usability

Overall, the add-on obtained good usability ratings and there were very encouraging comments. Some aspects were new for participants – this shows that there is indeed

potential for innovation. Below, a number of points are commented on in more detail.

Firstly, one of the general requirements from Section 2.2 was that a writing tool should not be too distracting. The combination of a low feeling of distraction while paying attention to the add-on suggests that the design of the add-on was a good compromise between necessary attention and too much distraction.

Secondly, while the overall rating of responsiveness was good, the problems participants mentioned should definitely be analysed more closely and fixed as soon as possible.

The response to the search tool was not very positive, but the participants still used it a lot. However, is is also known that the subjects had experience with search engines (cf. the question in the pre-study questionnaire) and probably knew the general concept; thus they did not comment so much on the principle but rather on the features they found to not work the way they expected.

The planning tool seems to have been helpful, especially when it provided structure guidance for the KT lecture students. The biggest issue, on the other hand, was the fact that the items in the planning tool were not stored and the participants who took a break could not use them again later on. For reliable persistence, the data would have to be stored on a server. Another mentioned improvement point was the interaction between document and planning tool, e.g. adding a structure to the document once sections have been defined in the tool and vice-versa.

The positive comments on the idea of the translation tool suggest that the idea was very well received. Problems were probably due to slow response times – we had some fast typists and because of the missing change trigger from the side of Google Docs, we had to resort to polling for text and cursor changes and then also to wait for a server response. Another problem is the fact that both the polling and the Google Docs editor are running in the client and we have make sure not to "eat up" too many resources with excessive polling.

In the last tool, the revision tool, participants particularly seemed to like the information about frequent words and suggested synonyms, but the other tools were also well received. This included the Flesch-Kincaid score as a measure which was probably new to several participants – thus there seems to be a curiosity and willingness to try new features.

A problem with the visualisation of the document structure was probably the fact that several participants did not mark their headings and paragraphs with the appropriate types but directly varied the font size and style instead. In this case, there was not enough information available to reconstruct the structure in a meaningful way. However, when writers become more used to applying paragraph styles or if they are explicitly prompted to use them, the benefit is very likely to increase.

Frequent usage of the search tool in general and the interaction with sources and the mentioned correlation between going back to the plan and the highlighting of frequent words and key concepts show that some of our writers engaged very actively with the add-on and also with the various tools in combinations. One of our main goals was to increase awareness and this interplay of reviewing the vocabulary and plans is a good example in this direction.

Finally, some deficits became apparent in the implementation. They were partially due to issues and limitations of Google Docs add-ons and the server we used, but also to the prototypical nature of the tool which made it less performant than it could have been and had a negative effect on the user experience. Improvement points in this area are mentioned in Section 8.1.

7.4 Summary

From the data collected in the user study we were able to learn more about the writing behaviour and usage of the writing support add-on and also related these aspects to text quality. We could confirm the variation of writing behaviour between individuals, but also the presence of general patterns, e.g. the planning tool being used earlier than the revision tool. We also found that tools are not exclusively used for the process they were designed for – which poses a further difficulty for automatic selection and presentation of tools.

Furthermore, evaluation of the questionnaires showed overall positive feedback of the add-on. We got encouraging comments on the ideas for tools, but also on possible improvements from the technological and functional points of view.

Chapter 8

Conclusion

8.1 Future Work

An experience that probably every researcher has made at some point is that the work is never done; there are always possible improvements and points which would deserve further investigation. In this subsection, we will list the most critical points from the technical point of view, but also look at a number of ideas for future extensions.

8.1.1 Evaluation

In order to evaluate the benefits of the writing support add-on more clearly, a further step in the analysis would be to carry out a much larger study where participants are divided into two groups, one group writing a text with the add-on and the second using Google Docs only, ideally in several sessions over the course of several months. Then the performance in both settings can be compared and the progress over time can be tracked.

8.1.2 Technical Improvements

First of all, we will look at some possible improvements in the implementation. For using the add-on in the future, and to supply it to a broader audience, it will be necessary to make the setup scalable. Currently, the external code is running on a slow server instance with limited capacity. It could of course just be migrated to a more powerful server, but ideally, some parts of the code could be converted to Google Apps Script and added to the add-on itself. Then the add-on would not rely on external connections for tasks like the keyword extraction.

Secondly, we would have to add at least a simple search engine providing information on all possible topics and not just a selection as was the case for our studies. Once this issue is solved, the add-on can be disseminated via the Chrome Webstore.

However, we do not need to stay in the Google universe. There are several environments which could just as well benefit from similar functionality, e.g. online LaTeX platforms or editors in content management systems.

8.1.3 Extensions of the Writing Support Add-on

We said before that our add-on can be considered a toolbox of writing tools and therefore it is only natural that it should be possible to adapt, extend, and exchange them. Below, you will find some ideas for improvements of the current features and for new functionality.

Search Tool

We said that the Writing Support add-on can be used with different information providers. However, we would particularly encourage embedding a provider which provides personalised results. Ideally, items that a user has stored beforehand, e.g. in a Mendeley[1] or zotero[2] library, are then also included.

Furthermore, the only option for keeping interesting documents at hand is to bookmark them. For long texts, where a lot of sources are used, this is bound to be cumbersome at some point, for the bookmarked items cannot be clustered. Therefore, we propose to add a tagging system where the user can add a number of tags to the documents and then filter them accordingly whenever necessary. Similarly, attaching notes to documents makes it easier to locate relevant information for the subsection the writer is currently working on.

Finally, citing papers could be simplified by automatically copying subsections of text into the document – including a correctly formatted reference, of course.

Planning Tool

To increase the writer's efficiency, the planning tool could be extended with preset structure templates for different types of articles. One example would be a template including the required subsection headers for a specific conference or journal. Then, a document skeleton is automatically generated.

Typing Tool

A first suggestion for the typing tool is a small extension of the current context display: currently the add-on shows the immediate surroundings of a term in question. If an item is particularly interesting, however, there is no more information available so far. Here, a detail-on-demand strategy could be applied and further information and a longer subsection of the text could be displayed, for instance when hovering over the item.

[1] https://mendeley.com
[2] https://www.zotero.org

Besides that, including a dictionary would be very beneficial for struggling L2 writers. To achieve a learning effect, the words that had to be looked up could even be stored and used as a basis for language training.

Revision Tool

The revision tool contains a small selection of text metrics we deemed useful for a quick revision. But if you look at the measures other revision aids provide, you will find that there is a very long list of other options available. We suggest adding more of these metrics to the add-on, with an option to deactivate them one by one in order to avoid cluttering and information overload.

Secondly, the revision tool is currently best suited for evaluating the entire document. We have seen that revision does not only happen globally, but also locally. Instead, we could also offer our analysis on a paragraph or section basis and thus achieve higher granularity.

General Extensions

Particular potential, however, lies in the meaningful combination of our tools, across different processes. At the moment, for example, the user can define a number of keywords as the main topics of her text in the planning tool and in the revision tool, we extract key concepts from the written text. Ideally, these two sets of concepts would be compared automatically.

8.1.4 Personalisation

We have talked about the individual differences observed in writing behaviour and it seems only adequate that writing support should be tailored to individual needs and preferences. The simplest method is probably what was already mentioned as an extension of the revision tool: providing an extensive number of features for each tool which can then be activated and deactivated by the user, for instance the aforementioned text metrics or different information sources for the search tool. A number of commonly used options would be preselected to make it easier to get started, in particular as a non-expert user.

In addition to that, if at some point in the future it is possible to automatically – and reliably – detect writing processes, the next step is to integrate this into the add-on (cf. Chapter 5). This will enable context-aware interaction with the various tools. However, in the user study we already saw that writers did not always select the tool intended for a process, which is a significant hurdle for automatic tool selection.

8.1.5 Further Areas of Development

Google Docs is often used for collaboratively working on a document. Collaborative Writing leads to a number of new questions – how to agree on a structure? Who is responsible for which subsection? Which aspects were already covered and how were they expressed? Do the subsections match each other in terms of style? A concrete example for collaboration support is an extension of the typing tool: say a writer is working on a paragraph dealing with a specific concept someone else has already mentioned in a different part of the document. Then this subsection could be brought up as a context example showing how keywords related to that topic were employed before.

8.2 Summary

In this thesis, we introduced a comprehensive writing tool which supports all stages of the composition process. Its concept was developed based on results we gathered through extensive literature research and a formative study on writing processes, where we identified needs and opportunities for support.

When designing the tool, particular focus was laid on the cognitive psychology underlying the writing activities. We included a number of features for *searching*, *planning*, *translating*, and *reviewing*. Implementation as a Google Docs add-on made it possible to integrate the tool directly into the user's working environment.

The add-on contains a set of tools for each writing process. In addition to commonly used features, we also included new ideas like the examples of words in context and the extraction of the document structure based on the paragraph styles.

A user study confirmed the potential of the tool, but also revealed some problems with our implementation, the possibilities for interaction with Google Docs add-on, and writing tools in general. Considering also the very positive feedback gathered during a number of demos, this is a great motivation to continue working on this topic and to extend the add-on accordingly.

List of Abbreviations

API	Application Programming Interface
DOI	Digital Object Identifier
GAS	Google Apps Script
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
JSON	JavaScript Object Notation
KT	Knowledge Technologies
L2	Second Language
LTM	Long-Term Memory
REST	Representational State Transfer
TF-IDF	Term Frequency - Inverse Document Frequency
URL	Unified Resource Locator
WM	Working Memory

Bibliography

Denis Alamargot, Lucile Chanquoy, and Maria Chuy. L'élaboration du contenu du texte: De la mémoire à long terme à l'environnement de la tâche. *Psychologie française*, 50(3):287–304, 2005.

Denis Alamargot, Gilles Caporossi, David Chesnet, and Christine Ros. What makes a skilled writer? Working memory and audience awareness during text composition. *Learning and Individual Differences*, 21(5):505–516, 2011.

Rui Alexandre Alves, São Luís Castro, Liliana De Sousa, and Sven Strömqvist. Influence of typing skill on pause-execution cycles in written composition. *Writing and cognition: Research and applications*, 2007.

Alan Baddeley. Working memory. *Science*, 255(5044):556, 1992.

Anne Becker. A review of writing model research based on cognitive processes. *Revision: History, theory, and practice*, pages 25–49, 2006.

James Britton and others. The Development of Writing Abilities (11-18). 1975.

Scott A. Crossley, Kristopher Kyle, and Danielle S. McNamara. The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior research methods*, 48(4):1227–1237, 2016.

Christophe Dansac and Denis Alamargot. Accessing referential information during text composition: When and why? *Knowing what to write: Conceptual processes in text production*, pages 76–97, 1999.

Cecilia di Sciascio, Vedran Sabol, and Eduardo E. Veas. uRank: Exploring Document Recommendations through an Interactive User-Driven Approach. In *IntRS@ RecSys*, pages 29–36, 2015.

Carol Sue Englert, Xiuwen Wu, and Yong Zhao. Cognitive tools for writing: Scaffolding the performance of students through technology. *Learning Disabilities Research & Practice*, 20(3):184–198, 2005.

Xiangmin Fan. *Scalable Teaching and Learning via Intelligent User Interfaces*. PhD thesis, University of Pittsburgh, 2017.

Ralph P. Ferretti, William E. Lewis, and Scott Andrews-Weckerly. Do goals affect the structure of students' argumentative writing strategies? *Journal of Educational Psychology*, 101(3):577, 2009.

Rudolph Flesch. A new readability yardstick. *Journal of applied psychology*, 32(3): 221, 1948.

Linda Flower and John R. Hayes. A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4):365, December 1981. ISSN 0010096X. doi: 10.2307/356600.

Arthur C. Graesser, Danielle S. McNamara, Max M. Louwerse, and Zhiqiang Cai. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods*, 36(2):193–202, 2004.

John R. Hayes. New directions in writing theory. *Handbook of writing research*, 2: 28–40, 2006.

Ronald T. Kellogg. Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory & Cognition*, 15(3):256–266, 1987. ISSN 0090-502X, 1532-5946. doi: 10.3758/BF03197724.

Ronald T. Kellogg. Effectiveness of prewriting strategies as a function of task demands. *The American Journal of Psychology*, pages 327–342, 1990.

Ronald T. Kellogg. Long-term working memory in text production. *Memory & cognition*, 29(1):43–52, 2001.

Mary Lynch Kennedy. The composing process of college students writing from sources. *Written Communication*, 2(4):434–456, 1985.

Adam Kilgarriff and Christiane Fellbaum. *WordNet: An Electronic Lexical Database*. JSTOR, 2000.

C. Michael Levy and Sarah Ransdell. Computer-aided protocol analysis of writing processes. *Behavior Research Methods, Instruments, & Computers*, 26(2):219–223, 1994.

Paul P. Maglio and Christopher S. Campbell. Tradeoffs in displaying peripheral information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 241–248. ACM, 2000.

Cerstin Mahlow and Robert Dale. Production media: Writing as using tools in media convergent environments. *Handbook of Writing and Text Production*, 10: 209, 2014.

Mari Carmen Puerta Melguizo, Lou Boves, and Olga Muñoz Ramos. A proactive recommendation system for writing: Helping without disrupting. *International Journal of Industrial Ergonomics*, 39(3):516–523, 2009.

Mari Carmen Puerta Melguizo, Teresa Bajo, and Olga Gracia Castillo. A Proactive Recommendation System for Writing in the Internet Age. *Journal of Writing Research*, 2(1), 2010.

Thierry Olive. Working memory in writing: Empirical evidence from the dual-task technique. *European Psychologist*, 9(1):32–42, 2004.

Thierry Olive and Ronald T. Kellogg. Concurrent activation of high-and low-level production processes in written composition. *Memory & Cognition*, 30(4):594–600, 2002.

Annie Piolat and Jean-Yves Roussey. Students' drafting strategies and text quality. *Learning and instruction*, 6(2):111–129, 1996.

Annie Piolat, Ronald T. Kellogg, and Fernand Farioli. The triple task technique for studying writing processes: On which task is attention focused? *Current Psychology Letters. Brain, Behavior and Cognition*, 4:67–83, 2001.

Emily Pitler and Ani Nenkova. Revisiting readability: A unified framework for predicting text quality. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 186–195. Association for Computational Linguistics, 2008.

Lia Plakans. The role of reading strategies in integrated L2 writing tasks. *Journal of English for Academic Purposes*, 8(4):252–266, 2009.

Bradley James Rhodes. *Just-in-Time Information Retrieval*. PhD thesis, Massachusetts Institute of Technology, 2000.

Gert Rijlaarsdam and Huub Van den Bergh. Writing process theory. *Handbook of writing research*, pages 41–53, 2006.

D. Gordon Rohman. Pre-writing the stage of discovery in the writing process. *College composition and communication*, 16(2):106–112, 1965.

Marlene Scardamalia and Carl Bereiter. Development of Dialectical Processes in Composition. *Language, Literacy, and Learning in Educational Practice: A Reader*, page 295, 1994.

Marlene Scardamalia, Carl Bereiter, and Rosanne Steinbach. Teachability of reflective processes in written composition. *Cognitive science*, 8(2):173–190, 1984.

Dale H. Schunk and Carl W. Swartz. Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology*, 18(3): 337–354, 1993.

Philip Shaw. Science research students' composing processes. *English for specific purposes*, 10(3):189–206, 1991.

Tony Silva. Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *Tesol Quarterly*, 27(4):657–677, 1993.

Nancy Sommers. Revision strategies of student writers and experienced adult writers. *College composition and communication*, 31(4):378–388, 1980.

Åsa Wengelin, Mark Torrance, Kenneth Holmqvist, Sol Simpson, David Galbraith, Victoria Johansson, and Roger Johansson. Combined eyetracking and keystroke-logging methods for studying cognitive processes in text production. *Behavior research methods*, 41(2):337–351, 2009.

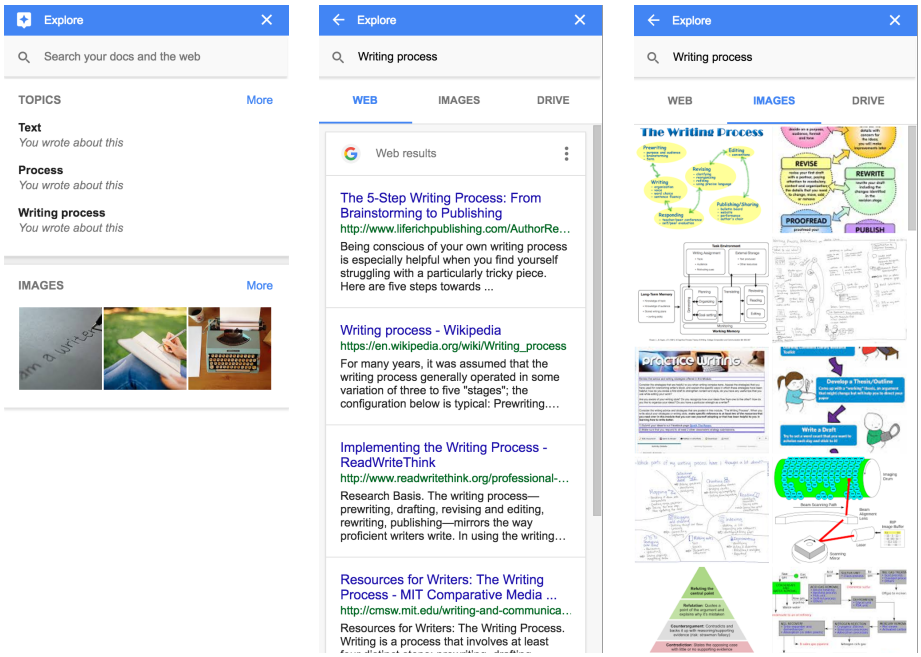
Appendix

Appendix A

Tools

This part of the appendix contains screenshots of the writing tools presented in Section 2.3.

A.1 Search Tools



(a) The main interface **(b)** Web search for keyword(s) **(c)** Image search for keyword(s)

Figure A.1: The Explore tool in Google Docs

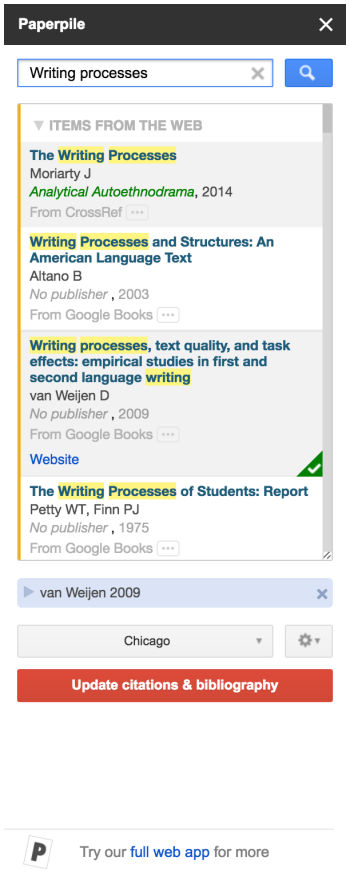


Figure A.2: The Paperpile Google Docs Add-on

A.2 Planning Tools

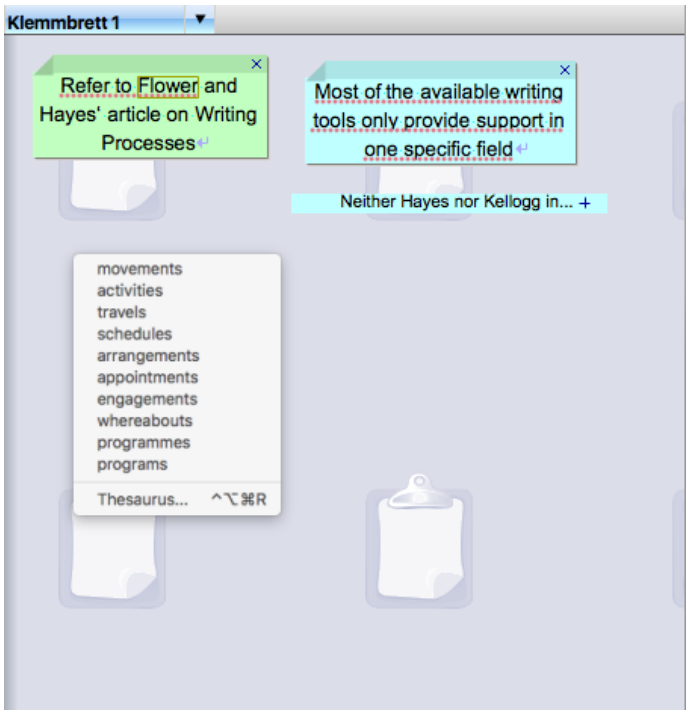


Figure A.3: Papyrus Autor "Klemmbrett"

A.3 Revision Tools

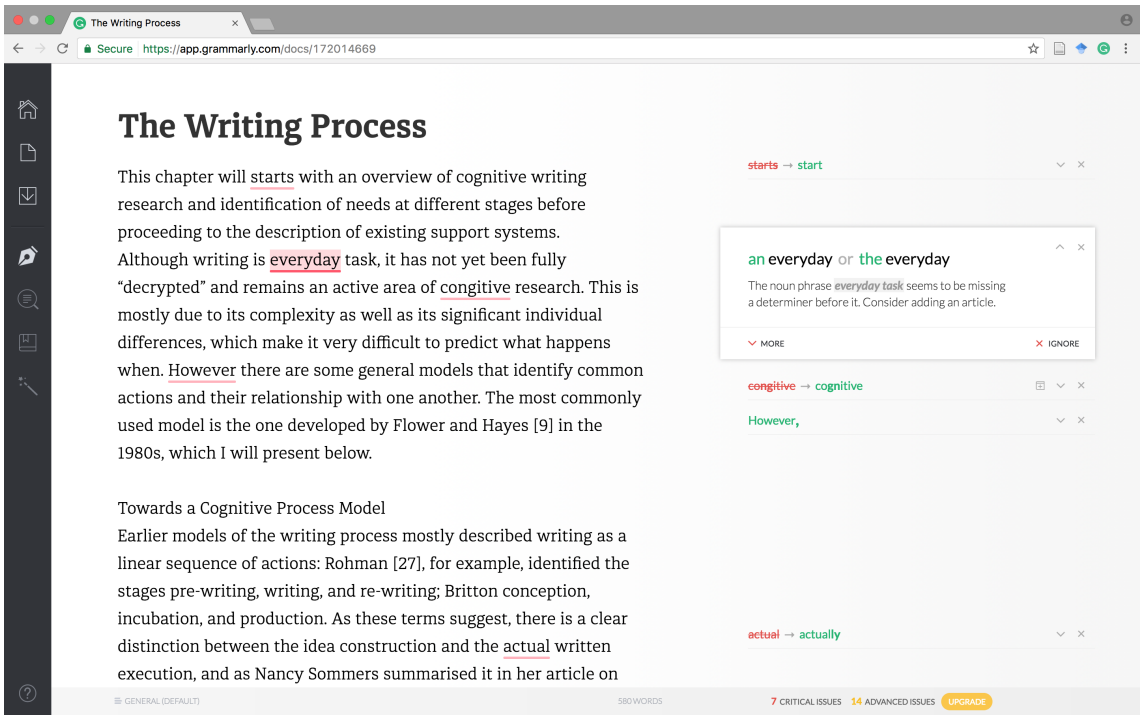
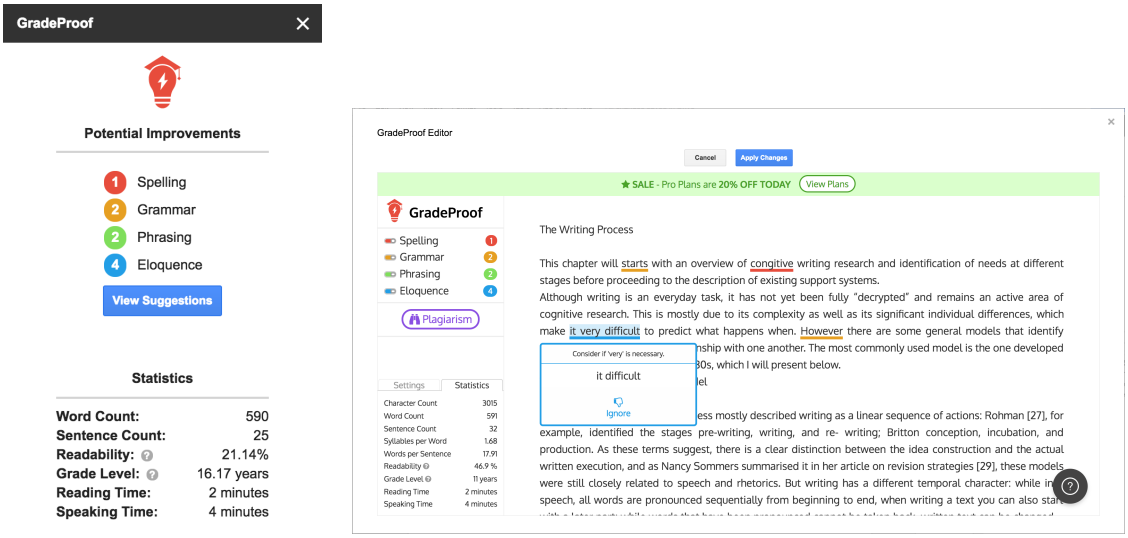
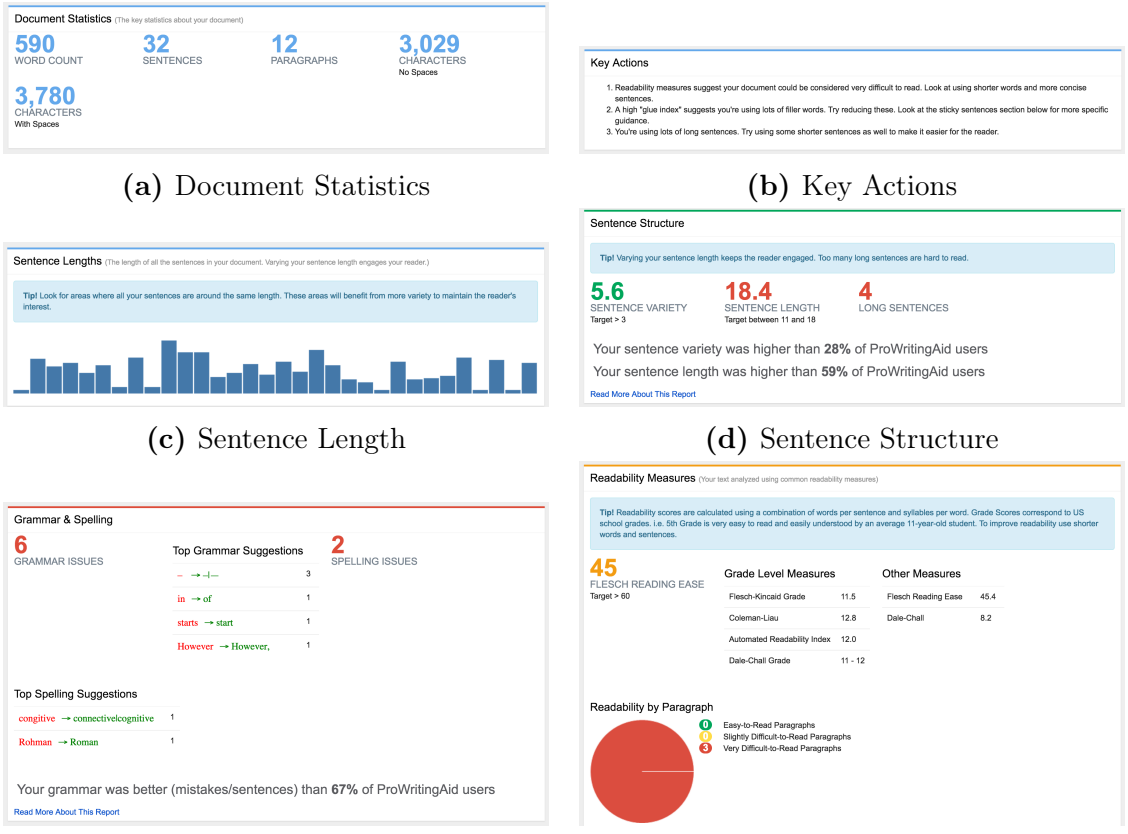


Figure A.4: Browser version of Grammarly



(a) Overview sidebar (b) Popup with detailed suggestions

Figure A.5: The Gradeproof Google Docs Add-on



(a) Document Statistics (b) Key Actions

(c) Sentence Length (d) Sentence Structure

(e) Grammar & Spelling (f) Readability Measures

Figure A.6: The ProWritingAid Google Docs Add-on: selection of measures in the popup "Summary Report"

Appendix B

Technologies and Frameworks

B.1 Node.js and Libraries

Node.js[1] is a JavaScript runtime-environment for network applications. It is designed to be non-blocking through asynchronous event loops. Advantages are its scalability and the fact that there are a number of packages available, e.g. for the quick and easy creation of REST interfaces.

In my add-on, Node.js is used in connection with the restify module[2], which makes it easy to define API endpoints, listen for incoming HTTP requests, to execute the respective functions and return data if required.

Another library that I'm using is natural[3], which provides natural language support like a tokenizer and stemmer. Together with the pos-js part-of-speech tagger[4], this is important for keyword extraction from a text, for example.

As stated in the description of the translating tool, the Writing Support tool uses the WordNet corpus [Kilgarriff and Fellbaum, 2000] for synonym suggestions. WordNet is a "large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept" and was developed at Princeton University. Each synset is assigned a unique identifier. We parsed the file containing all words and their respective synsets and inserted them into a collection of my MongoDB database. Upon synonym retrieval for a term, a query for all words in all the term's synsets is executed. Finally, duplicates as well as the term itself are removed.

[1] https://nodejs.org
[2] http://restify.com
[3] https://github.com/NaturalNode/natural
[4] https://github.com/dariusk/pos-js

B.2 MongoDB

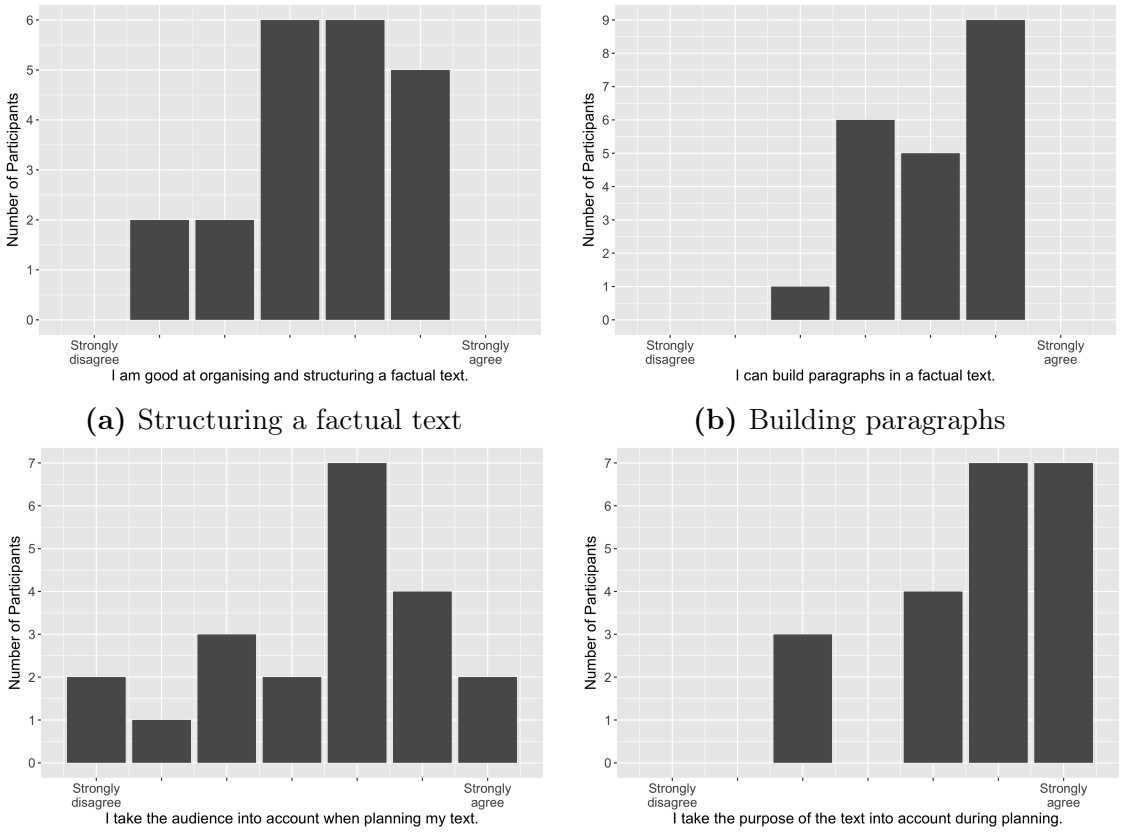
MongoDB[5] is an open-source NoSQL database system in which data items are stored as JSON-style documents. Although the organisation in collections can be used similarly to tables in a SQL database, their structure is not bound to a previously defined table schema but can vary from document to document. Collections can be queried based on document attributes. Each query returns a cursor over the relevant documents. The integration with Node.js is realised through a specialised driver[6]. Queries can be executed through the driver, or directly via the command line in the interactive mongo environment. Mongoexport allows exporting MongoDB collections for use in applications, for example as comma-separated values files.

[5] https://www.mongodb.com
[6] https://mongodb.github.io/node-mongodb-native

Appendix C

Results of the User Study

C.1 Pre-Study Questionnaire



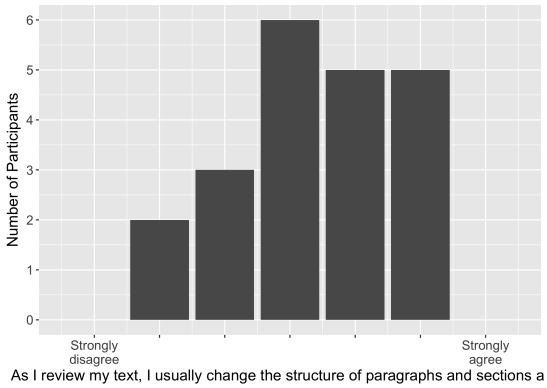
(a) Structuring a factual text

(b) Building paragraphs

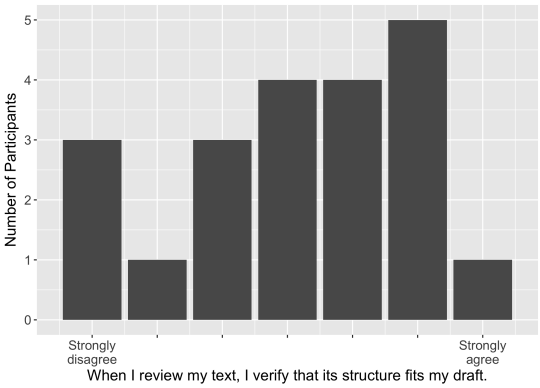
(c) Audience awareness

(d) Accounting for text purpose

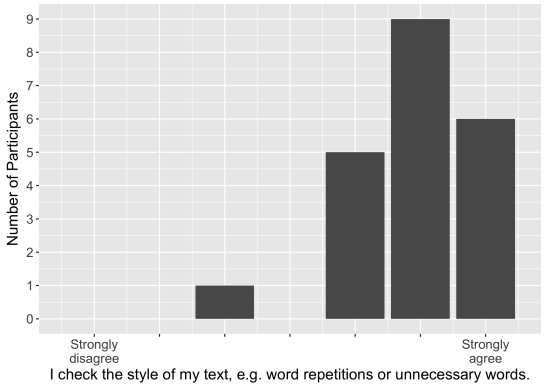
Figure C.1: Self-assessed writing skills and writing behaviour



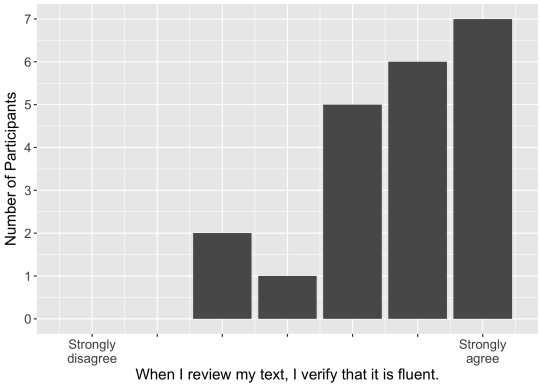
(a) Changes during revision



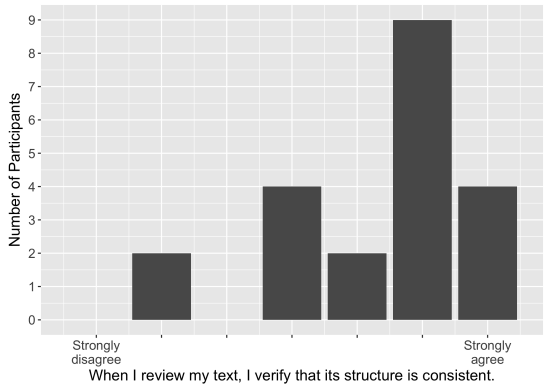
(b) Returning to the draft



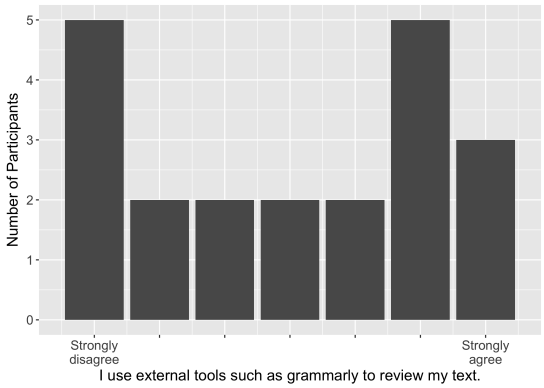
(c) Reviewing the style



(d) Reviewing the structure



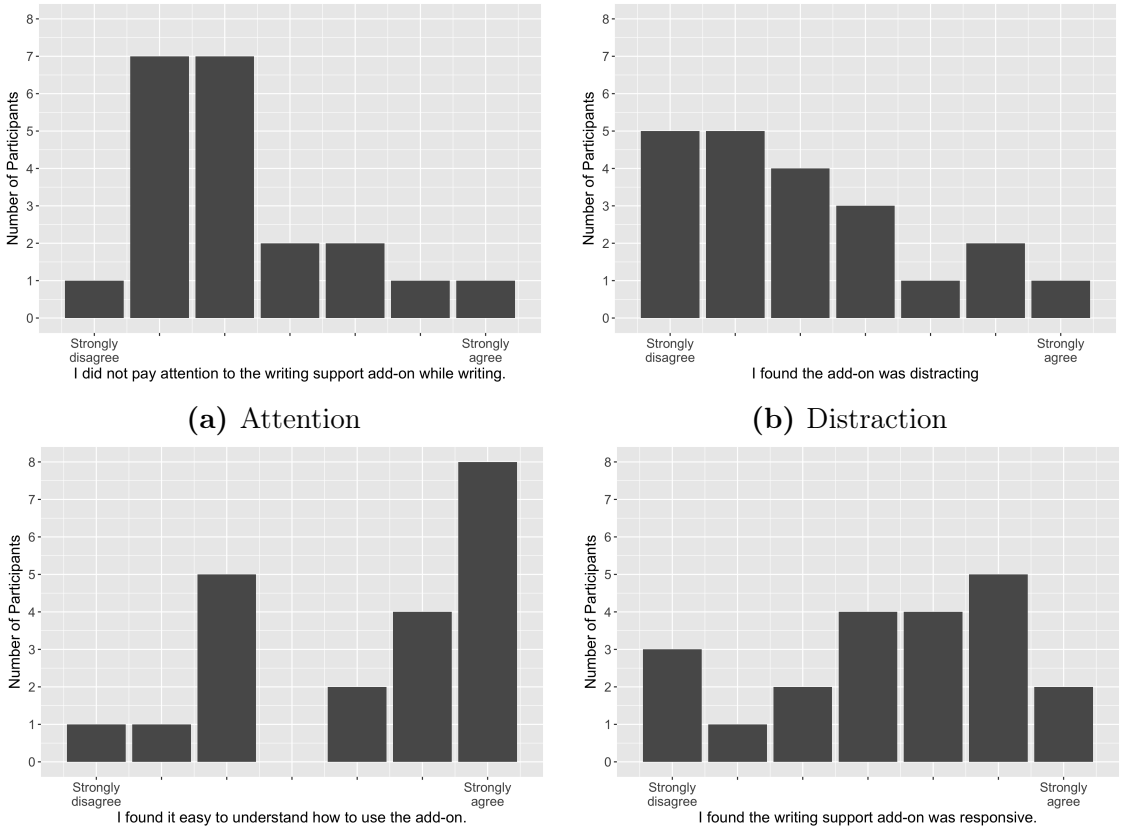
(e) Reviewing the consistency



(f) Using external revision tools

Figure C.2: Self-assessed revision behaviour

C.2 Post-Study Questionnaire



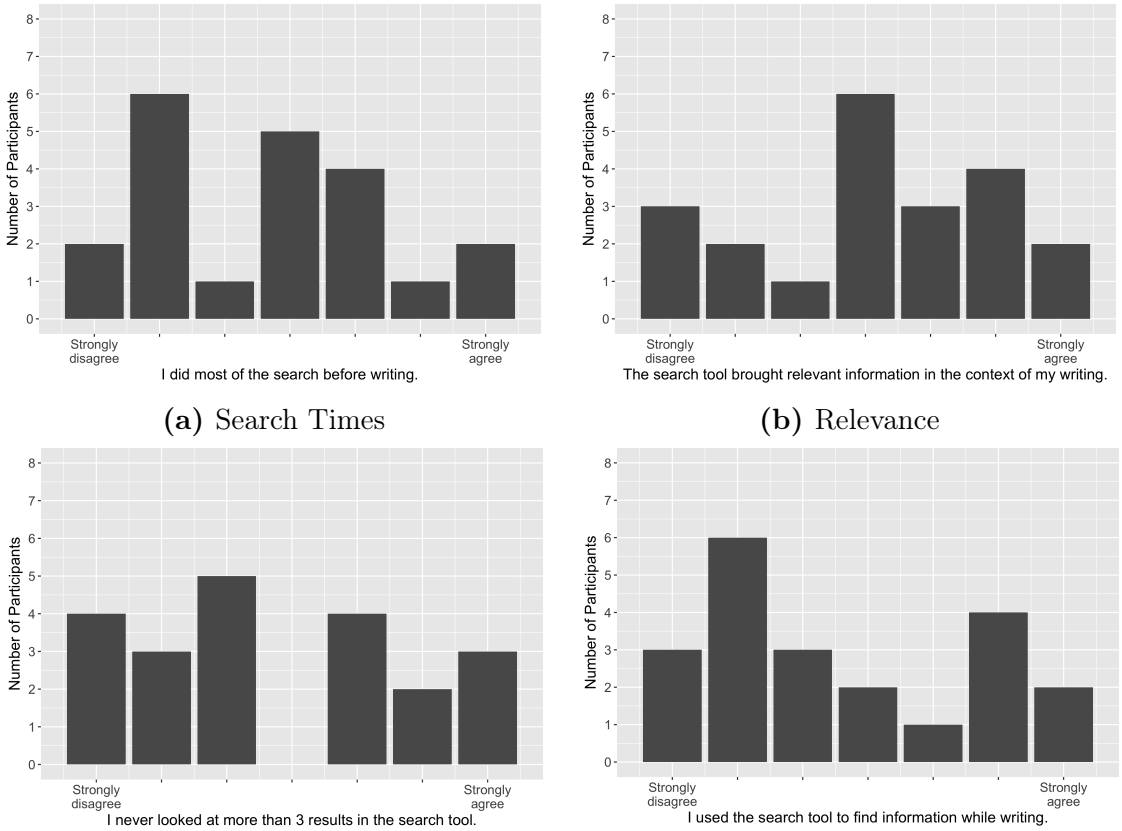
(a) Attention

(b) Distraction

(c) Understandability

(d) Responsiveness

Figure C.3: Overall rating of the add-on



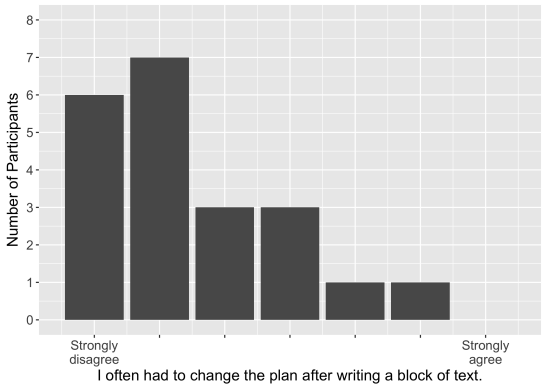
(a) Search Times

(b) Relevance

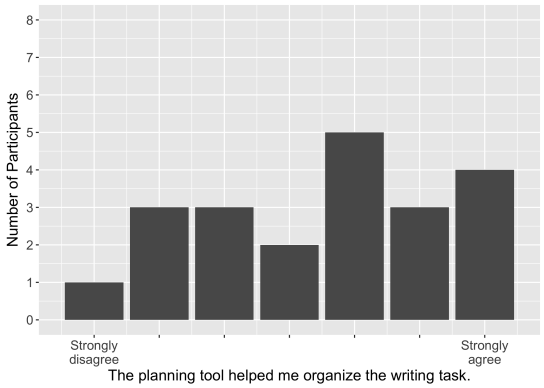
(c) Items considered

(d) Usage

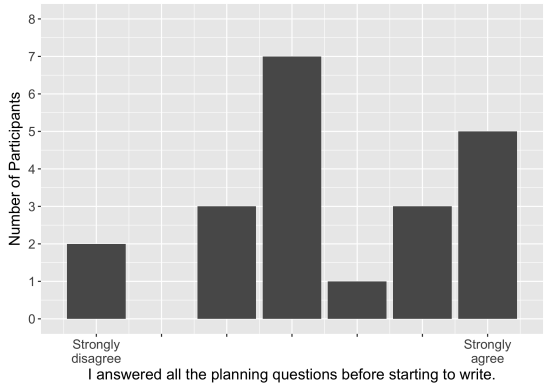
Figure C.4: Searching and the search tool



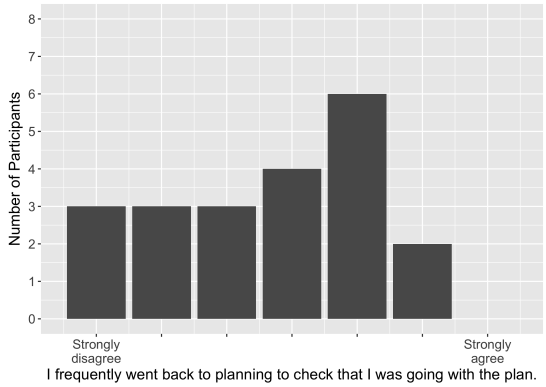
(a) Changing plans



(b) Helpfulness

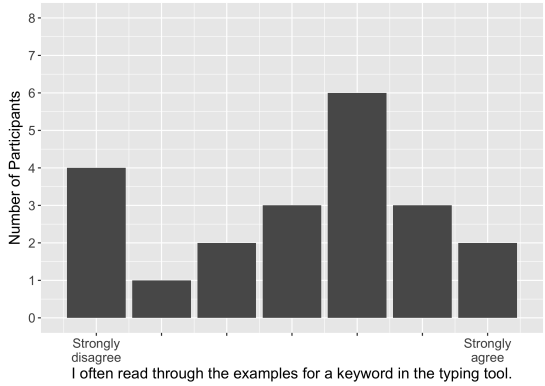


(c) Answered questions

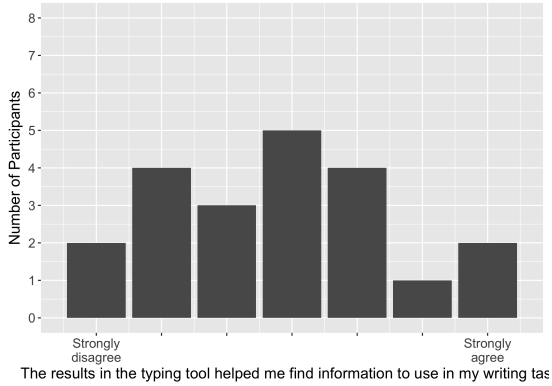


(d) Checking plans

Figure C.5: Planning and the planning tool

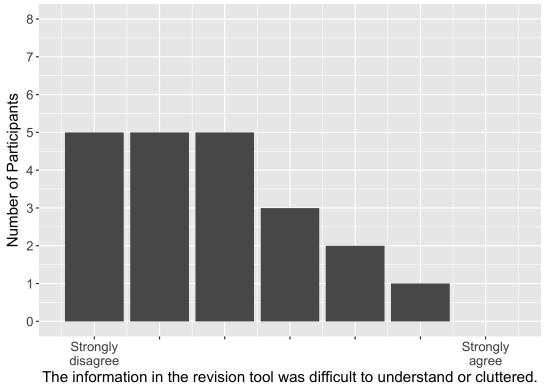


(a) Reading items

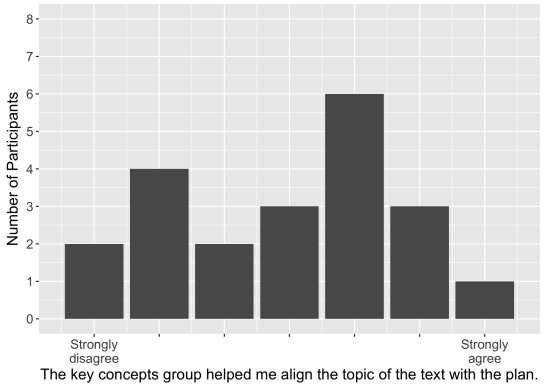


(b) Helpfulness

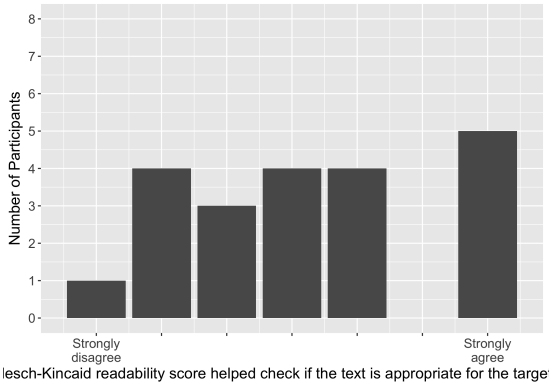
Figure C.6: Typing tool



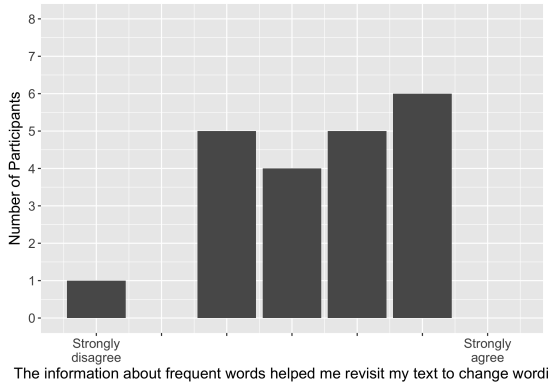
(a) Cluttering



(b) Key Concepts



(c) Audience



(d) Wording

Figure C.7: Revision tool