Finding the Right Words: Language Technologies to Support Formulation

Otto Kruse, Madalina Chitez, and Christian Rapp

Abstract This chapter explores the ability of digital technologies to provide language support for writers. With such ability, technologies directly intervene into the productive act of language creation, which we refer to by the traditional term *formulation*. Formulation here is defined as the kind of thinking that happens when a writer tries to linearize thought by using language. In written communication, formulation happens during interaction with an inscription tool and is strongly influenced by the kind of technology used. In this chapter, we look into some of the changes in formulation and language crafting that followed the introduction of digital technologies. We attempt to estimate where the developments are heading by addressing four issues: (1) support for the preparation of formulation, (2) real-time support during inscription, (3) support for the choice of words and collocations, and (4) support for language use at the revision stage by automated feedback and intelligent tutoring. The contribution concludes with some thoughts about future directions.

Keywords Formulation support · Inscription · Written communication · Automated feedback

O. Kruse (✉)
School of Applied Linguistics, Zurich University of Applied Sciences, Winterthur, Switzerland
e-mail: otto.kruse@gmx.net; xkso@zhaw.ch

M. Chitez
West University of Timisoara, Timișoara, Romania
e-mail: madalina.chitez@e-uvt.ro

C. Rapp
School of Management and Law, Center for Innovative Teaching and Learning, Zurich University of Applied Sciences, Winterthur, Switzerland
e-mail: christian.rapp@zhaw.ch

O. Kruse et al. (eds.), *Digital Writing Technologies in Higher Education*,
https://doi.org/10.1007/978-3-031-36033-6_28

1 Introduction

No matter through which theoretical lens we look at academic writing, it always involves crafting language so that the resulting text carries the message that an author wants to communicate. Writers know that words matter, and they usually spend time and effort to get the selected words into the right form and order. They know that the same thought can be expressed in different words and languages, but are also aware that meanings may change with every word replaced and every phrase altered. Writing is both expressing meaning and creating meaning (Wrobel, 2002). Some experienced writers even claim that they do not know what they think unless they read what they wrote. The written text feeds back into thinking and expands the options for creating meaning.

Formulation is defined as the kind of thinking that happens around the moment when words are linearized to a language string (see Kruse & Rapp, "Word Processing Software: The Rise of MS Word"; Kruse & Rapp, 2023). When the formulation activity is finished, writers may assume that what they have in mind is what they find expressed in their text. If not, they may change or accommodate their thoughts to what they wrote. Alternatively, they may continue revising the text until it conveys what they have in mind.

Writing tools are enabling technologies, without which the activity of writing cannot happen. Writing technology enables the inscription of letters and words on a writing surface (Baron, 2009; Bazerman, 2013, 2018; Haas, 1996; Mahlow & Dale, 2014; Ong, 1982). Formulation, in contrast, is a purely mental activity which, however, is not independent of inscription, as writers usually develop their text in interaction with the writing tool where they can re-read and re-think what they wrote. If the tool allows, they can revise the wording to reshape or extend their thoughts. It must be noted, however, that behind written formulation we can still see traces of the historically and ontogenetically earlier skill of oral formulation with its own rhetoric, registers, and formulation strategies.

Due to the various capabilities of digital technologies, the writing tools have become active agents in crafting language, making it necessary to re-consider almost anything we knew about formulation. In this chapter, we will look at formulation in the light of the many digital tools that currently make it easier for writers to process language. These technologies started with the development of grammar and spell checkers in the 1970s (see Smith et al., 1984), which, for the first time in media history, enabled writing tools to exert considerable influence on language production and text composition. This was only the first step. Today, Natural Language Processing (NLP) and computational linguistics are advanced enough to not only support but actually appropriately assemble language pieces into different text types such as essays or business news. They are able to master complex formulative, grammatical, and evaluative problems, as has been shown in Part "Writing Analytics and Language Technologies" of this book. We will examine these technological developments from the perspective of formulation support they can provide for writers and will discuss the conceptual changes in formulation theory that the technologies ask for. For any

conclusive deliberations, we have to challenge the traditional rhetoric assumption that language capability is an exclusively human characteristic. Today, language has fallen prey to the computer, and sometimes it seems that computers outwit the human brain with regard to grammatical, rhetorical, and terminological abilities. Then, the questions that arise are: What is it that will remain for the humans? And how can humans make use of the computer in the best way to boost their linguistic performance?

2 Traditional Views of Formulation (the Past)

Formulation, a topic grounded in psycholinguistics, has attracted attention from many researchers throughout the past centuries. One of the starting points was Wilhelm Wundt's (1900) monumental 2-volume work *The Language*, which presented thorough considerations about creating utterances and forming sentences (see Levelt, 2013, for a review). While Wundt treated formulation primarily in relation to the thoughts expressed, Bühler (1927) pictured it in a communicative frame where the sender, the receiver, and the message were specified, all three of them being part of a given context. The message, in his view, is not only tied to the thought of the writer but also to the receiver it is meant for. Thus, the message has representational meaning, but it also conveys the internal state of the sender and may be seen as an appeal to the receiver (see Nerlich & Clarke, 1998). To address formulation, several alternative terms have been used, such as 'sentence production' or 'sentence formation,' 'language production,' or 'language generation,' all referring to the activity of producing meaningful chains of words.

Only a few studies have dealt with written formulation (e.g., Keseling, 1993; Wrobel, 1995, 2002). The cognitive model by Hayes and Flower (Flower & Hayes, 1981; Hayes, 2012; Hayes & Flower, 1980) introduced the term 'translation' for 'formulation.' This model was built on the idea that content is first created cognitively and then translated into language. This assumption led to numerous discussions about how such a process could take place (e.g., Alamargot & Chanquoy, 2001; Fayol, 2016; Fayol et al., 2012; Galbraith, 1999, 2009). Recent revisions of the model (Galbraith, 1999, 2009; Galbraith & Torrance, 2004) have tried to push it into a dual process approach, understanding "writing processes as an interleaving of dispositional content generation and rhetorical structuring" (Galbraith & Torrance, 2004, p. 63). Instead of a successive creation of content and language, as the original model suggested, they proposed a parallel processing of content generation as both cognitive and linguistic. Content, here, is not created first and then translated, but it is assumed that "ideas form as the language is produced" (Torrance, 2016, p. 80). This opens the door to considering language as a part of, if not a leading force in knowledge creation and meaning-making. This shift in perspective would, consequently, afford an additional theoretical step towards an operational view of language telling us how language is related to ideas and what exactly writers do with words and grammar.

Knowledge cannot be constructed and cannot be thought of without language, but it needs a linguistic theory to say how language does this and how writers use language for the expression of intentions and the creation of thought (see more in Kruse & Rapp, 2023). The quality of thinking and writing depends, for instance, on the size of the mental lexicons of the writers, as well as of their mental phrasebooks. Limited word and collocation knowledge make writing difficult and would allow formulation on a basic level only. Words are the building blocks for sentence construction and the basic elements of meaning-making. Faber (2015) sees words and terms as units of specialized knowledge as much as of a specialized language and considers them as "access points to larger knowledge configurations" (p. 14). Common languages and their rich vocabularies form the core of formulation activities and frame the use of special terminologies associated with different professional, cultural, or educational domains.

Even if language generation of adult writers is automatized to a large degree, they still make deliberate and purposeful choices of terminology, word order, phrases, rhetoric, parts of speech, etc. After about 12 years of school, beginner academic writers have enough training to understand a good deal of the linguistic and rhetorical means of text production for transformative writing strategies (Kellogg, 2008). They are aware of linguistic decisions and know of the need to be precise in language use. They also have a mental lexicon large enough to address the most important issues in academic thought but still have to extend their mental lexicons considerably to keep up with the vocabulary and knowledge of their disciplines. Word learning in the disciplines is usually not independent from acquiring knowledge, and it is a rather slow process of familiarizing with words (Wolter, 2022) including their morphological, collocational, semantic, and pragmatic aspects.

3 Current Transformations of Formulation Induced by Technology (the Present)

The new ways of digital writing have made formulation a much more comfortable activity mainly because of improved options for inscription and revision (see Heilmann, "The Beginnings of Word Processing: A Historical Account"; Rapp & Kruse, "Word Processing Software: The Rise of MS Word"; Kruse & Anson, "Plagiarism Detection and Intertextuality Software"). Corpus studies have led to a wide array of information on the lexical, collocational, grammatical, rhetorical, and genre-specific dimensions of texts that can be operationalized for writers, even though not for all languages alike (Chitez & Dinca, "On Corpora and Writing"). The key to language technologies is automatic text analysis, for which a large number of methods have been developed (see Part "Writing Analytics and Language Technologies"). The

items below encompass a birds-eye-view of what language technologies have to offer for formulation support.

- Automated spell, grammar, and style checkers (Cotos, "Automated Feedback on Writing"; Link & Kolovskaia, "Automated Scoring of Writing")
- Sentence completion and word prediction features for real-time text production support (Kruse & Rapp, "Word Processing Software: The Rise of MS Word")
- Synonym finders for word level support (Kruse & Rapp, "Word Processing Software: The Rise of MS Word")
- Built-in corpora along with search tools and query platforms (Chitez & Dinca, "On Corpora and Writing") for inquiries on research language
- Phrase books for collocation level support (Chitez & Dinca, "On Corpora and Writing")
- Rhetorical and discipline-specific automated feedback for genre writing support (Cotos, "Automated Feedback on Writing")
- Intelligent tutors for guided individualized learning (Banawan et al., "The Future of Intelligent Tutoring Systems for Writing"; Cotos, "Automated Feedback on Writing")
- Preparing formulation by idea and concept development (Kruse et al., "Creativity Software and Idea Mapping Technology")
- Key stroke logging for 'behind the scenes' analyses of inscription and revision processes (Wengelin, "Investigating Writing Processes with Keystroke Logging")
- Reference management systems for quoting and evaluating literature (Proske et al., "Reference Management Systems")
- Plagiarism-detection systems for checking intertextuality and relations to other publications (Anson & Kruse, "Plagiarism Detection and Intertextuality Software")
- Argument mining or mapping for argument construction support (Benetos, "Digital Tools for Written Argumentation")
- Information retrieval and knowledge extraction systems connecting automatic content generation with linguistic framing (Benites, "Information Retrieval and Knowledge Extraction for Academic Writing")
- Automatic text generation producing almost perfect linguistic surface structures with minimal human involvement (Delorme Benites et al., "Automated Text Generation and Summarization for Academic Writing").

With these technologies, formulation has become a collaborative human-computer issue and, eventually, more parts of it are done by the machine than by the human writer. Digital technology provides formulation provision on all levels, be it at the word, phrase, grammar, or document level. NLP developments have decoded the productive aspects of language and are advancing the support not only for language usage but also for content development.

In the remainder of this contribution, we will focus on four functional aspects of the technological abilities listed above to demonstrate and discuss in more detail what impact digital technologies can have on formulation processes:

1. Support for preparing and guiding formulation processes: Making writers collect and create meaningful bits of verbalized thought and knowledge either in the form of mind or concept maps without the constraints of linearization, or in the form of notes and summaries.
2. Real-time support during inscription: Aiding inscription with linguistic support such as sentence completion or grammar and spell checkers to unburden writers from elementary constraints of sentence construction.
3. Support for the choice of vocabulary: Search tools to scan through digital corpora which can provide direct access to linguistic information such as word use, collocations, rhetorical choices, or synonyms.
4. Support through automated feedback and intelligent tutoring: Complex analytic tools can offer formulation support for relevant linguistic and rhetorical traits such as cohesion/coherence, focus, style, structure, connectives, moves/steps, and more.

It is worth noting that most of these developments are happening in English; the transfer to other languages is not always a given.

3.1 Support for Preparing Formulation Processes

Writers may follow different strategies regarding the onset of inscription in a writing project. They may prefer to do the reading and note-taking first to acquire enough knowledge for their paper before they start formulating their ideas and developing content. They may also begin writing right away from what they already know and then do the reading. For formulation, not only knowledge about a topic must be available but also knowledge of disciplinary vocabulary. Some of these prerequisites can be acquired before formulation actually takes place, especially if supported by tools for:

3.1.1 Mind and Concept Mapping

These tools operate at the concept level, where concepts are represented by words, expressions, or interconnected words (see Kruse et al., "Creativity Software and Idea Mapping Technology"). They help prepare formulation by singling out the thoughts to be expressed and the relations between those thoughts, which serve as initial framing for the linearized text. In mind mapping, the resulting tree can be transformed into an outline that contains the central ideas/words to be filled in with text. Concept mapping, in contrast, focusses on the interrelation of single thoughts or bits of knowledge including the relation that connects them.

3.1.2 Note Taking and Summarisation

Such tools are built on the idea that the best preparation for a text is to write small texts in advance (e.g., notes summaries), which later can be used either as a flexible basis for the expression of own ideas or as the basis for the literature report (see Pitura, "Digital Note-Taking for Writing"). Although note taking and summarisation are traditional forms of academic work, they have been integrated in new ways of accessing texts and organising text excerpts. Summarized text can be easily reused, thus offering basic textual units for the formulation process.

3.1.3 Annotation and Social Annotation

Annotation tools in general aim to foster the connection between reading and writing by relating own thought with printed text. In social annotation, where several users are involved, annotations can be commented on, answered, or extended (Hodgson et al., "Social Annotation: Promising Technologies and Practices in Writing"). This allows for acquiring a deeper understanding of the topic at hand, along with an extension of the expressive abilities to write about it.

These three types of tools account for the fact that activities focused on preparation of formulation need to intertwine, not separate, the linguistic and content-related elements. There is no abstract preparation by language learning through memorizing certain language features, studying word lists, or trying to remember collocations and phrases. Also, memorizing content is not an effective strategy because formulation means making sentences move, and this kind of sequentiality is created by linguistic elements such as connectives, sentence structures, and grammar.

3.2 Real-Time Support for Formulation Activities During Inscription

Formulation processes are tied to the short time span when words are written down. Writers may prepare this short moment of inscription by activities such as reading, summarising of literature, or thinking ahead of what they might want to say. Still, most decisions are made in the moment of inscription, and, ideally, formulation support has to be squeezed into the short time slot that inscription offers. Supporting formulation activities at the very moment of inscription needs technologies that are fast enough to enter the microprocesses of inscription without distracting writers too much or disturb their thought processes.

3.2.1 Grammar, Style, and Spell Checkers

Grammatical accuracy is an essential aspect of most writings and has both an operational aspect in terms of text construction and a conventional aspect in terms of the compliance to established norms. The fact that digital inscription tools not only passively preserve letters and words but also actively inform the writers about various dimensions of inacurate language use or sentence construction, has been a millennial invention. Still, to this day, grammar checkers are far from being perfect (Cotos, "Automated Feedback on Writing"). Yet, while inadvertently missing out on some aspects of writing, they still have a similar success rate in detecting textual problems compared to teachers. For a thorough discussion of reliability and validity of automatic scoring and text evaluation, see Link and Koltovskaia ("Automated Scoring of Writing") and Cotos ("Automated Feedback on Writing").

The first grammar checker, called Writer's Workbench, was developed in the 1970s (see Smith et al., 1984) but became publicly available only in the mid-1980s. It was created by Lorinda Cherry and Nina McDonald from the Bell Labs and was based on NLP technology. It involved mainly lists of words and lists of common errors that the program marked in the text. Grammar checkers cannot simply rely on the rules of a consistent grammar, which they would "use" or "apply." Languages simply do not work consistently as rule-based; rather, languages exhibit multidimensional usage patterns of which only some are reflected in grammar books. For more technical information on NLP and grammar checkers, see Dale et al. (2000).

3.2.2 Automatic Word Division and Hyphenation

A writing problem that seems to have been solved is word division, as word processors do this automatically, and no decisions have to be made by the writer. In inflexible inscription systems such as typewriters, the number of words fitting into one line was a problem, at least in languages where separation was restricted to syllables. Not only was knowledge of hyphenation necessary, but also the space left for the last word in a line had to be calculated (at least when typewriters were used). Hyphenation works on the basis of word lists, in which the division points are marked and applied when the text approaches the margin. There are also formulas in use for syllable separation when words on the list are missing. Automatically set hyphens are considered 'soft' hyphens in contrast to self-set hyphens which are considered 'hard.' The soft hyphens disappear when the text is reformatted and the hyphenated word does not hit the end of the margin, while the hard hyphens remain in such a case. For formulation activity, automatic hyphenation is another help function freeing the writer from a lower-order concern that, in typewriting, not only demanded constant attention but was also a source of errors.

3.2.3 Autocompletion and Word Prediction

Perhaps the most straightforward approach to supporting inscription processes is autocompletion, which aims at what is essential in formulation: deciding on the next word(s). Autocompletion software offer potential completions to a sentence beginning; they can offer single words or chains of words such as phrases or collocations. It can offer several suggestions from which the writer can choose. This software type is mainly used in mobile devices and for search engines, but it is also an option in Microsoft Windows. Next to a large vocabulary, a collocation dictionary is needed for word prediction. Individual shortcuts for autocompletion are possible so that a user-based dictionary for completion can be created and, for instance, 'thank you' is automatically offered when typing 'tha'. Autocompletion can be based on a general dictionary or on individual word usage built from previous texts.

3.3 Support for the Choice of Vocabulary

Approaches from corpus and computational linguistics provide features allowing to search for appropriate words, word usage, or collocations. These measures are not as immediately tied into the inscription process as the aforementioned ones but need a certain search action on the part of the writer. The action may be as quick as right-clicking the mouse to open the synonym finder, or it may be a more extended action like querying a corpus-based search tool to look for collocations. Let's consider some prominent examples of digital tools offering vocabulary-level support.

3.3.1 Synonym Finders

Synonym finders, as included in Microsoft Word and similar word processors, are good examples of support features for formulation during inscription. They work only on demand and not automatically like autocorrection (once 'activated'). The technology of synonym finders is comparatively simple, having developed from word collections in dictionaries and then made available as searchable electronic documents. All it takes is to choose the right word that corresponds to the text.

In Microsoft Word, right-mouse-clicking on any word launches the synonym function, usually providing five alternative words. When an alternative word is clicked on, it replaces the original word in the text. In some instances, antonyms are also displayed. The thesaurus, which is available from the same menu in Microsoft Word as the synonym finder, has a different organisational form. It is a structured, alphabetically ordered list of interconnected words. Each term is clickable to retrieve a new list of synonyms, so that variations in meaning of similar terms can be readily assessed.

A more complex system of synonym finding is offered as a Microsoft Word add-on by https://www.synonyms.com. It offers more synonyms and antonyms than

Microsoft Word and it is available in six languages. Additionally, this tool offers word usage examples and is more creative in graphically representing word relations. The same company STANDS4 running the synonym finder also offers an abbreviation and acronym finder at http://www.abbreviations.com, which is a useful addition to the synonym finder.

3.3.2 Phrasebooks

Connecting phrasebooks with word processors is uncommon. Phrasebooks have been created for various languages, domains, and research fields. In general, phrasebooks support writers according to the idiomaticity of the domain/genre by offering complex phrases. In academic writing, the Manchester Academic Phrasebank is the best-known tool, pioneering not only in collecting phrases for numerous topics, but also in providing clarity through presentation in tabular format (Davis & Morley, 2015).

A bilingual (German and English) phrasebook is integrated within Thesis Writer (Kruse & Rapp, 2019; Rapp et al., 2022), which is a specialised platform designed to instruct and guide thesis writers focusing on their extended research papers. Thesis Writer offers a template-based outline generator to create a thesis proposal. Each step (e.g., 'state your research question', or 'describe the state of the knowledge to your topic') is supported by a list of 10 commonly used phrases. Additionally, Thesis Writer offers a large, open phrasebook similar to the Manchester Academic Phrasebank. It contains phrases, relevant for thesis writing, distributed into 16 categories related to research writing and 63 sub-categories, each of which corresponding to a particular communicative aim, similar to Swales' (1981) moves and steps. For each of the 63 categories, 20 distinct phrases are presented–all derived from a large corpus of academic research papers, dissertations and expert statements.

3.3.3 Concordancers and Collocation Finders

Corpus linguistics and computational linguistics have contributed several technologies to offer writers real-time support through text-based evidence (Chitez et al., 2015; Cotos, 2017; Cotos et al., 2017; Flowerdew, 2015; Hsieh & Liou, 2009). In second language teaching, providing lexical support directly from corpora forms an important grounding (Sinclair, 1999, 2004) that expands to more complex linguistic phraseology and rhetorical functions (Flowerdew, 2012, 2015).

L2 learners and writers, however, differ from L1 users who, at least passively, know most words and phrases and therefore look for the most appropriate lexical choice rather than considering rhetorical effectiveness. Whilst for L1 writers, synonym finders (for words) and phrasebooks (for phrases) seem advantageous, there is still a need to provide support for special terminology or more complex expressions, which requires individual corpus searches. The values of these offers likely depend on the size and specific focus of available corpora.

One of the most straightforward approaches to provide access to corpus data is offering writers an accessible or integrated concordance tool such as AntConc or ConcApp to search a corpus of selected documents for language use. Thesis Writer (Kruse & Rapp, 2019) has an integrated concordancer, through which users can explore an embedded English and German corpus. Searches can be performed for single words or word connections (collocations), with the tool searching for all instances where the words or collocations have been used in the corpus and then displaying them in a list. The number of words preceding/following the search term can be selected, and writers can check how the respective word/collocation is used in an authentic sample of documents. Whilst little is known about how much such tools are utilised and what their gain is, it seems that users require training in order to profit from them (Hsieh & Liou, 2009).

A far more differentiated collocation finder is offered by Philip Edmond (http://www.just-the-word.com), which presents search results in a clearly arranged tabular format. Here is an example of a search query on the term "risk":

> 'accept risk' (45); 'carry risk' (96); 'concern about risk' (15); 'cover risk' (31); 'involve risk' (59); 'take risk' (680)

The bracketed numbers refer to the number of entries found in the British National Corpus. Collocations with verbs, adjectives, and other nouns are presented separately. Here, a three-page list of collocations only for the word 'risk' provides a systematic account of all word connections. Even though collocations are a main issue in formulation, it is not clear how such a linguistic offer would serve writers without reducing the amount of information to a manageable size.

3.4 Support Through Automated Feedback and Intelligent Tutoring

Many modern language technologies make use of algorithms that can analyse deep structures of texts and, from there, can help generate automated feedback and provide tutoring for writers (see Part "Writing Analytics and Language Technologies" this volume). Such feedback is usually not given during the initial inscription but rather at a later stage when the text or a considerable part of it exists as a draft or seems finished. Revision means to reformulate parts of the text in order to adjust it to various demands of the content, structure, flow, genre, or audience. None of such changes can be accomplished without altering the wording and re-shaping the linguistic surface. For the writer, this kind of revision means to change the perspective from a text producer to a reader and evaluator of the text. Similarly, writing software has to transpose into an educational technology specifying what and how writers should learn. The tools we are looking at below give feedback not only at the language level but address a much broader range of deeper textual issues such as content development, focus, coherence and cohesion, organisation, rhetoric, flow, and structure. Each of them touches upon a different layer of text development and relates differently to language. Feedback

on any of these measures necessarily leads to changes of the wording of the text and forces them to engage in reformulation. In what follows, we give examples of how automated feedback and intelligent tutoring influence formulation. We hope to demonstrate that reformulation needs more attention as a necessary part of revision, as it is connected to a large number of meta-communicative, meta-discursive, or meta-linguistic aspects demanding learning and re-orientation from the writer.

3.4.1 Rhetorical Support: Move Analysis

Move analysis is grounded on Swales' (1981) analysis of research article introductions, which connects the rhetoric of the text (expressions, phrases) with the communicative intentions of the writer (moves and steps) and the overall Introduction-Methods-Results-Discussion (IMRD) structure of the research article. Scholars analysing discourse in the Swalesian tradition identify phrases that serve as functional language, which is distinguishable from content-based language. To communicate effectively, research writers tend to use such functional language to make their intentions clear and avoid ambiguities. Beginning academic writers, on the other hand, are often not aware of the need to compose their research reports based on genre-specific rhetorical elements and instead try to express themselves creatively yet unconventionally.

The most elabora te Automatic Writing Evaluation (AWE) tool using move analysis to instruct writers is the Research Writing Tutor (RWT) (Cotos, 2014; Cotos, "Automated Feedback on Writing"; Cotos et al., 2020). It is based on the evaluations of a carefully collected 900-document research article corpus containing 30 papers from 30 disciplines. The documents were analysed along the categories of the Swalesian move analysis (Cotos, 2018; Swales, 1981) which, for this purpose, was extended beyond the introduction to cover all IMRD/C sections.

The core feature of the RWT is an algorithm that operates based on a collection of functional language (n-grams) related to specific moves/steps, which allows to identify the IMRD/C rhetorical traits, make them visible by color-coding, and generate feedback comments on them. Numerous examples of alternative language choices characteristic of individual moves/steps may be accessed vis a functional concordancer. A similar automatic feedback system is the AcaWriter, which developed move/step-like detection systems for expository and reflective student genres (Knight et al., 2018, 2020; Shibani, "Analytic Techniques for Automated Analysis of Writing").

Both tools, the RWT and the AcaWriter, provide scaffolding features for writers that are built around an automatic detection of phrases and offer support for their selection, interpretation, and eventually replacement. "Scaffolding", here, means that text construction and learning about academic writing are equally involved. Learning about formulation takes place while developing own paper. The pedagogical problem of such help functions for formulation activities is to offer appropriate word combinations without necessarily constraining the rhetoric of the writers. Making all writers use the same wording would be a rather odd practice for a scaffolding

system. Rather, the selection process of wordings is tied to the aims of a particular textual step and can be optimized when it becomes clear what the aim is and which formulative options are available.

3.4.2 Cohesion and Coherence

The concepts of coherence and cohesion offer another opportunity to connect structural aspects of text organization with linguistic text elements such as transition markers, forward and backward references, and connectives. Coherence refers to the logical dimension of thought organization in a text while cohesion denotes the linguistic connectedness between stretches of text (Halliday & Hasan, 2013; Taylor et al., 2019; van Dijk, 1977). Coherence and cohesion depend on each other, and writing usually involves aligning topic development with linguistic organizers of text flow.

A critical element for coherence and cohesion are connectives or connectors. They serve both as syntactical bridges between clauses and as indicators specifying the relationship between thoughts (e.g., causal, temporal, additive, conclusive, conditional, etc.). In academic writing, precise thinking depends on the selection and usage of connectives. What makes them difficult to learn is their sheer number. The web-based multilingual lexical resource at http://connective-lex.info/ lists 142 English, 274 German, 328 French, and 173 Italian connectives (for more information see Stede et al., 2019). Learning to distinguish and use them takes time. Determining the right connective is not a matter of grammar but rather a matter of thought organization.

What can automated feedback do to support the use of connectives and how can it help writers understand deeper levels of coherence? There are a number of tools that focus on cohesion and include connectives; we will refer here to only two. The first is Coh-Metrix (http://cohmetrix.com) by McNamara et al. (2013, 2014), which is an analytical system using algorithms for a high number of different indicators describing linguistic and discourse representations of a text (McNamara & Graesse, 2012; McNamara et al., 2014). These algorithms have been applied in an online tutoring platform called Writing Pal (Banawan et al., "The Future of Intelligent Tutoring Systems for Writing"; Roscoe & McNamara, 2013") which hosts many analytic, tutorial, and gaming features for learning writing strategies. Coh-Metrix provides five different indices to evaluate uploaded text (Dowell et al., 2016, p. 78): narrativity, deep cohesion, referential cohesion, syntactic simplicity, and word concreteness. What do the coherence measures offer? "Deep cohesion" considers the number of different kinds of connectives and conceptual links while referential cohesion refers to the "words and ideas that overlap across sentences and the entire text, forming explicit threads that connect the text for the reader (p. 78)."

The second tool is the Writing Mentor, an NLP-based tool (Burstein et al., 2018) available as a free-of-charge Google Docs add-on (https://mentormywriting.org). It is designed to provide feedback on four relevant essay parameters: convincing, well developed, coherent, and well edited. 'Coherent' is defined as indicating the flow of ideas (highlighting topical words), transition terms, long sentences, pronoun usage,

and titles. Topical words that mark the flow of ideas are highlighted. For 'coherence', the user has to choose from several feedback types such as transition terms, sentence length, section headers, pronoun reference, and indicators of topic development. Feedback is specified for the genres of essay, letter, narrative, and other. Tutorials are connected to each evaluative dimension connected with the respective advice, thus making the transition from automated feedback to intelligent tutoring.

Both tools, RWT and Writing Mentor, may be characterised as a language-awareness tool, directing the writer's attention to relevant linguistic and rhetorical issues and explaining their significance to textual construction principles. Much of the evaluative activity is left to the writer, as are the conclusions for text revision. Tutorial advice for complex linguistic issues such as coherence has limits. Automated feedback does not guide the writers' pen but points at the factors that matter and makes them think about language.

4 Conclusion: What Are the Developments Pointing at (the Future)?

Language assistance and formulation support have received comparatively little attention in writing theory. Although digital language technologies exist for more than 50 years and have been recognized for their rapid and revolutionizing results, their effects on formulation have not been analysed and theorized systematically. Perhaps one of the reasons is the lack of linguistic underpinning in writing theories and, to some extent, the nature of teaching writing in L1 contexts. Although no one would seriously doubt that language skills are necessary for writing, there is no consistent operational language theory that would explain what writers do with language (see Kruse & Rapp, 2023).

Technologies supporting formulation activities have arrived at an advanced stage of development, with many of them now regularly used in word processors or other digital environments, and more are still to come. We can no longer think of writing without these technologies, but we have to accept that the nature of formulation has significantly changed. The inscription environments of today's word processors and editors made formulation a more comfortable task, with few lower-order constraints that formerly occupied a large part of a writer's attention. Newly developed technologies, such as automated feedback, intelligent tutoring, argumentation support, or corpus-based search tools, address the writers' higher-order concerns, particularly when connected to a certain genre or domain of writing. In natural language generation supported by artificial intelligence, formulation may be completely executed by the computer while the writer's activity would be reduced to the control and revision of wording and content. We have to assume that summarization, reformulation, and editing would be executed automatically by digital technologies so that much of the formulation activity will be delegated to the computer.

Acknowledgements The research of Christian Rapp was financially supported by the ZHAW DIZH Fellowship Call 2020.

References

Alamargot, D., & Chanquoy, L. (2001). *Through the models of writing.* Kluwer. https://doi.org/10.1007/978-94-010-0804-4

Baron, D. (2009). *A better pencil. Readers, writers, and the digital revolution.* Oxford University Press.

Bazerman, C. (2013). *A rhetoric of literate action: Literate action* (Vol. 2). WAC Clearinghouse and Parlor Press.

Bazerman, C. (2018). What do humans do best? Developing communicative humans in the changing socio-cyborgian landscape. In S. Logan & W. Slater (Eds.), *Perspectives on academic and professional writing in an age of accountability* (pp. 187–203). Southern Illinois University Press.

Bühler, K. (1927). *Die Krise der Psychologie* (The crisis of psychology). Fischer.

Burstein, J., Elliot, N., Klebanov, B. B., Madnani, N., Schwartz, M., Houghton, P., & Molloy, H. (2018). Writing Mentor™: Writing progress using self-regulated writing support. *Journal of Writing Analytics, 2*, 285–313. https://doi.org/10.37514/JWA-J.2018.2.1.12

Chitez, M., Rapp, C., & Kruse, O. (2015). Corpus-supported academic writing: How can technology help? In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds.), *Critical CALL-Proceedings of the 2015 EUROCALL Conference* (pp. 125–132). Research-publishing.net.

Cotos, E. (2014). *Genre-based automated writing evaluation for L2 research writing: From design to evaluation and enhancement*. Palgrave Macmillan.

Cotos, E. (2017). Language for specific purposes and corpus-based pedagogy. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 248–265). Wiley.

Cotos, E. (2018). Move analysis. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–8). Wiley.

Cotos, E., Huffman, S., & Link, S. (2017). A move/step model for methods sections: Demonstrating rigour and credibility. *English for Specific Purposes, 46*, 90–106.

Cotos, E., Huffman, S., & Link, S. (2020). Understanding graduate writers' interaction with and impact of the Research Writing Tutor during revision. *Journal of Writing Research, 12*(1), 187–232.

Dale, R., Moisl, H., & Somers, H. (Eds.). (2000). *Handbook of natural language processing.* Marcel Dekker.

Davis, M., & Morley, J. (2015). Phrasal intertextuality: The responses of academics from different disciplines to students' re-use of phrases. *Journal of Second Language Writing, 28*, 20–35. https://doi.org/10.1016/j.jslw.2015.02.004

Dowell, N. M., Graesser, A. C., & Cai, Z. (2016). Language and discourse analysis with Coh-Metrix: Applications from educational material to learning Environments at Scale. *Journal of Learning Analytics*, *3*(3), 72–95. https://doi.org/10.18608/jla.2016.33.5

Faber, P. (2015). Frames as a framework for terminology. In H. J. Kockaert & F. Steuers (Eds.), *Handbook of terminology* (part 1) (pp. 14–33). Benjamin.

Fayol, M. (2016). From language to text. The development and learning of translation. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 130–143). Guilford.

Fayol, M., Alamargot, D., & Berninger, V. W. (2012) (Eds.). *Translation of thought to written text while composing. Advancing theory, knowledge, research, methods, tools, and applications.* Psychology Press.

Flower, L. S., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387. https://doi.org/10.2307/356600

Flowerdew, L. (2012). *Corpora and language education.* Palgrave Macmillan.

Flowerdew, L. (2015). Learner corpora and language for academic and specific purposes. In F. Meunier, G. Gilquin, & S. Granger (Eds.), *The Cambridge handbook of learner corpus research* (pp. 465–484). Cambridge University Press. https://doi.org/10.1017/CBO9781139649414

Galbraith, D. (1999). Writing as a knowledge-constituting process. In M. Torrance & D. Galbraith (Eds.), *Knowing what to write: Conceptual processes in text production* (pp. 139–160). Amsterdam University.

Galbraith, D. (2009). Writing as discovery. *British Journal of Educational Psychology, Monograph Series II, 6*, 5–26. https://doi.org/10.1348/978185409X421129

Galbraith, D., & Torrance, M. (2004). Revision in the context of different drafting strategies. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision. cognitive and instructional processes* (pp. 63–76). Springer. https://doi.org/10.1007/978-94-007-1048-1_5

Haas, C. (1996). *Writing technology. Studies on the materiality of literacy.* Routledge.

Halliday, M. A. K., & Hasan, R. (2013). *Cohesion in English.* Routledge.

Hayes, J. R. (2012). Modelling and remodelling writing. *Written Communication, 29*(3), 369–388. https://doi.org/10.1177/0741088312451260

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Erlbaum.

Hsieh, W.-M., & Liou, H.-C. (2009). A case study of corpus-informed online academic writing for EFL graduate students. *CALICO Journal, 26*(1), 28–47. https://doi.org/10.1558/cj.v26i1.28-47

Kellogg, R. T. (2008). Training writing skills: A cognitive development perspective. *Journal of Writing Research, 1*(1), 1–26. https://doi.org/10.17239/jowr-2008.01.01.1

Keseling, G. (1993). *Schreibprozess und Textstruktur. Empirische Untersuchungen zur Produktion von Zusammenfassungen* [Writing process and text structure. Empirical studies about the production of summaries]. Niemeyer.

Knight, S., Buckingham Shum, S., Ryan, P., Sándor, Á., & Wang, X. (2018). Designing academic writing analytics for civil law student self-assessment. *International Journal of Artificial Intelligence in Education, 28*(1), 1–28. https://doi.org/10.1007/s40593-016-0121-0

Knight, S., Shibani, A., Abel, S., Gibson, A., Ryan, P., Sutton, N., Wight, R., Lucas, C., Sándor, Á., Kitto, K., Liu, M., Vijay Mogarkar, R., & Buckingham Shum, S. (2020). Acawriter: A learning analytics tool for formative feedback on academic writing. *Journal of Writing Research, 12*(1), 141–186. https://doi.org/10.17239/jowr-2020.12.01.06

Kruse, O., & Rapp, C. (2019). Seamless writing: How the digitisation of writing transforms thinking, communication, and student learning. In C.-K. Looi, L.-H. Wong, C. Glahn, & S. Cai (Eds.), *Seamless learning: Perspectives, challenges and opportunities* (pp. 191–208). Springer.

Kruse, O., & Rapp, C. (2023). What writers do with language. New inscription technologies require a fresh look at formulation theory. In P. M. Rogers, D. Russell, P. Carlino, & J. M. Marine (Eds.), *Writing as a human activity: Implications and applications of the work of Charles Bazerman* (pp. 366–392). The WAC Clearinghouse / University Press of Colorado. https://doi.org/10.37514/PER-B.2023.1800.2.15

Levelt, W. J. M. (2013). *A history of psycholinguistics. The pre-Chomsky era.* Oxford University.

Mahlow, C., & Dale, R. (2014). Production media: Writing as using tools in media convergent environments. In E.-M. Jakobs & D. Perrin (Eds.), *Handbook of writing and text production* (pp. 209–230). DeGruyter Mouton.

McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavioral Research, 45*, 499–515. https://doi.org/10.3758/s13428-012-0258-1

McNamara, D. S., & Graesse, A.C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. IGI Global.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix. A writing strategy tutoring system called the Writing Pal*. Cambridge University.

Nerlich, B., & Clarke, D. D. (1998). The linguistic repudiation of Wundt. *History of Psychology, 1*(3), 179–204. https://doi.org/10.1037/1093-4510.1.3.179

Ong, W. J. (1982). *Orality and literacy. The technologizing of the word*. Routledge.

Rapp, C., & Kruse, O. (2020). Thesis Writer 2.0 - a system supporting academic writing, its instruction and supervision. In C. Müller Werder & J. Erlemann (Eds.), *Seamless Learning – lebenslanges, durchgängiges Lernen ermöglichen* [Seamless learning—Enabling lifelong continuous learning] (pp. 235–240). Waxmann. https://doi.org/10.31244/9783830992448

Rapp, C., Kruse, O., & Ott, J. (2022). Thesis Writer: Digitale Anleitung zum wissenschaftlichen Arbeiten, Schreiben und Denken [Thesis Writer: Digital guidance for scientific work, writing and thinking]. In B. Dilger, J. Erlemann, C. Müller, & C. Rapp (Eds.), *Seamless Learning - Grenz- und kontextübergreifendes Lehren und Lernen in der Bodenseeregion* [Seamless learning—Teaching and learning across borders and contexts in the Lake Constance region] (pp. 147–175). Springer. https://doi.org/10.1007/978-3-658-34698-0_7

Roscoe, R. D., & McNamara, D. S. (2013). Writing pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology, 105*(4), 1010–1025. https://doi.org/10.1037/a0032340

Sinclair, J. (1999). A way with common words. In H. Hasselgård & S. Oksefjell (Eds.), *Out of corpora. Studies in Honour of Stig Johansson: Language and computers* (Vol. 26, pp. 157–180). Rodopi.

Sinclair, J. (2004). *How to use corpora in language teaching*. Benjamins.

Smith, C. R., Kiefer, K. E., & Gingrich, P. S. (1984). Computers come of age in writing instruction. *Computers and the Humanities, 18*, 215–224. https://doi.org/10.1007/BF02267225

Stede, M., Scheffler, T., & Mendes, A. (2019). Connective-lex: A web-based multilingual lexical resource for connectives. *Discours, 24*, Article 10098. https://doi.org/10.4000/discours.10098

Swales, J. M. (1981). *Aspects of article introductions.* University of Michigan.

Taylor, K. S., Lawrence, J. F., Connor, C. M., & Snow, C. E. (2019). Cognitive and linguistic features of adolescent argumentative writing: Do connectives signal more complex reasoning? *Reading and Writing, 32*, 983–1007. https://doi.org/10.1007/s11145-018-9898-6

Torrance, M. (2016). Understanding planning in text production. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 72–87). Guilford.

Van Dijk, T. A. (1977). Connectives in text grammar and text logic. In T. A. van Dijk & J. S. Petofi (Eds.), *Grammars and descriptions* (pp. 11–63). De Gruyter.

Wolter, B. (2022). Comparing the L1and L2 mental lexicon. A depth of individual word knowledge model. *Studies in Second Language Acquisition, 23*(1), 41–69. https://doi.org/10.1017/S0272263101001024

Wrobel, A. (1995). *Schreiben als Handlung. Überlegungen und Untersuchungen zur Theorie der Textproduktion* [Writing as action. Thoughts and investigations towards a theory of text production]. Niemeyer. https://doi.org/10.1515/9783110917468.

Wrobel. A. (2002). Schreiben und Formulieren. Prätext als Problemindikator und Lösung [Writing and formulation. Pretext as problem indicator and solution]. In D. Perrin, I. Böttcher, O. Kruse, & A. Wrobel (Eds.), *Schreiben. Von intuitiven zu professionellen Schreibstrategien* [Writing. From intuitive to professioneal writing strategies] (pp. 83–96). Westdeutscher.

Wundt, W. (1900). *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von* Sprache, *Mythos und Sitte. Bd. I. Erster Teil und Bd. 2 Zweiter Teil: Die Sprache* [Cultural Psychology. An investigation into the laws of the evolution of language, myth and custom. Vol. 1: first part and Vol. 2: second part: Language]. Kroner.

Otto Kruse is a retired professor of the School of Applied Linguistics at the Zurich University of Applied Sciences in Winterthur, Switzerland. He was the head of his department's writing center. A psychologist by education, he worked in clinical psychology, social work and applied linguistics. His expertise in the field of writing is connected to the teaching of writing, intercultural aspects of writ-ing, critical thinking, and the development of writing technologies. Together with Christian Rapp, he created "Thesis Writer", a writing platform supporting students with their theses and dissertations.

Madalina Chitez is a Senior Researcher in Applied Corpus Linguistics at the West University of Timisoara, Romania. She is the founder and president of the COHDUS research centre (Centre for Corpus Related Digital Approaches to Humanities) which aims at conducting applicative research that uses linguistic data for interdisciplinary studies.

Christian Rapp heads the Educational Technology Team at the Center for Innovative Teaching and Learning, School of Management and Law, Zurich University of Applied Sciences (ZHAW). He has coordinated various international R&D projects, among others "Seamless Writing: Expanding Tech-nologies to Support Thesis Writing" (EU-Interreg). He is a fellow of the Digitalization Initiative of the Zurich Higher Education Institutions (DIZH) and a board member of the European Association for the Teaching of Academic Writing (EATAW). Together with Otto Kruse, they established "Thesis Writer", an online platform that supports students with writing their theses and dissertations (http://www.thesiswriter.eu/).