

SWAN – Scientific Writing AssistaNt

A Tool for Helping Scholars to Write Reader-Friendly Manuscripts

<http://cs.joensuu.fi/swan/>

Tomi Kinnunen* Henri Leisma Monika Machunik Tuomo Kakkonen Jean-Luc Lebrun

Abstract

Difficulty of reading scholarly papers is significantly reduced by *reader-friendly* writing principles. Writing reader-friendly text, however, is challenging due to difficulty in recognizing problems in one's own writing. To help scholars identify and correct potential writing problems, we introduce SWAN (Scientific Writing AssistaNt) tool. SWAN is a rule-based system that gives feedback based on various quality metrics based on years of experience from scientific writing classes including 960 scientists of various backgrounds: life sciences, engineering sciences and economics. According to our first experiences, users have perceived SWAN as helpful in identifying problematic sections in text and increasing overall clarity of manuscripts.

1 Introduction

A search on “tools to evaluate the quality of writing” often gets you to sites assessing only one of the qualities of writing: its *readability*. Measuring ease of reading is indeed useful to determine if your writing meets the reading level of your targeted reader, but with scientific writing, the statistical formulae and readability indices such as Flesch-Kincaid lose their usefulness.

In a way, readability is subjective and dependent on how familiar the reader is with the specific vocabulary and the written style. Scientific papers are targeting an audience at ease with

a more specialized vocabulary, an audience expecting sentence-lengthening precision in writing. The readability index would require recalibration for such a specific audience. But the need for readability indices is not questioned here. “Science is often hard to read” (Gopen and Swan, 1990), even for scientists.

Science is also hard to *write*, and finding fault with one's own writing is even more challenging since we understand ourselves perfectly, at least most of the time. To gain objectivity scientists turn away from silent readability indices and find more direct help in checklists such as the peer review form proposed by Bates College¹, or scoring sheets to assess the quality of a scientific paper. These organise a systematic and critical walk through each part of a paper, from its title to its references in peer-review style. They integrate readability criteria that far exceed those covered by statistical lexical tools. For example, they examine how the text structure frames the contents under headings and subheadings that are consistent with the title and abstract of the paper. They test whether or not the writer fluidly meets the expectations of the reader. Written by expert reviewers (and readers), they represent them, their needs and concerns, and act as their proxy. Such manual tools effectively improve writing (Chuck and Young, 2004).

Computer-assisted tools that support manual assessment based on checklists require natural language understanding. Due to the complexity of language, today's natural language processing (NLP) techniques mostly enable computers to deliver shallow language understanding when the

* T. Kinnunen, H. Leisma, M. Machunik and T. Kakkonen are with the School of Computing, University of Eastern Finland (UEF), Joensuu, Finland, e-mail: tkinnu@cs.joensuu.fi. Jean-Luc Lebrun is an independent trainer of scientific writing and can be contacted at jllebrun@me.com.

¹<http://abacus.bates.edu/~ganderso/biology/resources/peerreview.html>

vocabulary is large and highly specialized – as is the case for scientific papers. Nevertheless, they are mature enough to be embedded in tools assisted by human input to increase depth of understanding. SWAN (*Scientific Writing AssistaNt*) is such a tool (Fig. 1). It is based on metrics tested on 960 scientists working for the research Institutes of the Agency for Science, Technology and Research (A*STAR) in Singapore since 1997.

The evaluation metrics used in SWAN are described in detail in a book written by the designer of the tool (Lebrun, 2011). In general, SWAN focuses on the areas of a scientific paper that create the first impression on the reader. Readers, and in particular reviewers, will always read these particular sections of a paper: title, abstract, introduction, conclusion, and the headings and subheadings of the paper. SWAN does *not* assess the overall quality of a scientific paper. SWAN assesses its fluidity and cohesion, two of the attributes that contribute to the overall quality of the paper. It also helps identify other types of potential problems such as lack of text dynamism, overly long sentences and judgmental words.

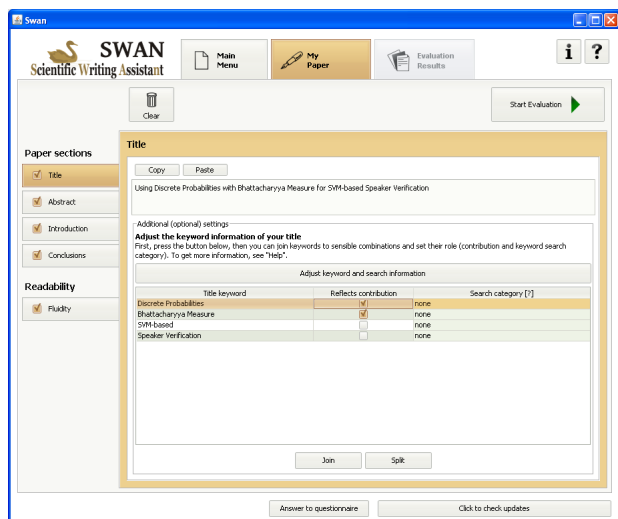


Figure 1: Main window of SWAN.

2 Related Work

Automatic assessment of student-authored texts is an active area of research. Hundreds of research publications related to this topic have been published since Page’s (Page, 1966) pioneering work on automatic grading of student essays. The research on using NLP in support of writing scientific publications has, however, gained much less attention in the research community.

Amadeus (Aluisio et al., 2001) is perhaps the system that is the most similar to the work outlined in this system demonstration. However, the focus of the Amadeus system is mostly on non-native speakers on English who are learning to write scientific publications. SWAN is targeted for more general audience of users.

Helping our own (HOO) is an initiative that could in future spark a new interest in the research on using of NLP for supporting scientific writing (Dale and Kilgarriff, 2010). As the name suggests, the shared task (HOO, 2011) focuses on supporting non-native English speakers in writing articles related specifically to NLP and computational linguistics. The focus in this initiative is on what the authors themselves call “domain-and-register-specific error correction”, i.e. correction of grammatical and spelling mistakes.

Some NLP research has been devoted to applying NLP techniques to scientific articles. Paquot and Bestgen (Paquot and Bestgen, 2009), for instance, extracted keywords from research articles.

3 Metrics Used in SWAN

We outline the evaluation metrics used in SWAN. Detailed description of the metrics is given in (Lebrun, 2011). Rather than focusing on English grammar or spell-checking included in most modern word processors, SWAN gives feedback on the core elements of any scientific paper: *title*, *abstract*, *introduction* and *conclusions*. In addition, SWAN gives feedback on *fluidity* of writing and paper structure.

SWAN includes two types of evaluation metrics, *automatic* and *manual* ones. Automatic metrics are solely implemented as text analysis of the original document using NLP tools. An example would be locating judgemental word patterns such as *suffers from* or locating sentences with passive voice. The manual metrics, in turn, require user’s input for tasks that are difficult – if not impossible – to automate. An example would be highlighting title keywords that reflect the core contribution of the paper, or highlighting in the abstract the sentences that cover the relevant background.

Many of the evaluation metrics are strongly inter-connected with each other, such as

- Checking that abstract and title are consistent; for instance, frequently used abstract keywords should also be found in the title;

and the title should not include keywords absent in the abstract.

- Checking that all title keywords are also found in the paper structure (from headings or subheadings) so that the paper structure is self-explanatory.

An important part of paper quality metrics is assessing text *fluidity*. By fluidity we mean the ease with which the text can be read. This, in turn, depends on how much the reader needs to memorize about what they have read so far in order to understand new information. This memorizing need is greatly reduced if consecutive sentences do not contain rapid change in topic. The aim of the text fluidity module is to detect possible topic discontinuities within and across paragraphs, and to suggest ways of improving these parts, for example, by rearranging the sentences. The suggestions, while already useful, will improve in future versions of the tool with a better understanding of word meanings thanks to WordNet and lexical semantics techniques.

Fluidity evaluation is difficult to fully automate. Manual fluidity evaluation relies on the reader’s understanding of the text. It is therefore superior to the automatic evaluation which relies on a set of heuristics that endeavor to identify text fluidity based on the concepts of *topic* and *stress* developed in (Gopen, 2004). These heuristics require the analysis of the sentence for which the Stanford parser is used. These heuristics are perfectible, but they already allow the identification of sentences disrupting text fluidity. More fluidity problems would be revealed through the manual fluidity evaluation.

Simply put, here *topic* refers to the main focus of the sentence (e.g. the subject of the main clause) while *stress* stands for the secondary sentence focus, which often becomes one of the following sentences’ topic. SWAN compares the position of topic and stress across consecutive sentences, as well as their position inside the sentence (i.e. among its subclauses). SWAN assigns each sentence to one of four possible fluidity classes:

1. **Fluid:** the sentence is maintaining connection with the previous sentences.
2. **Inverted topic:** the sentence is connected to a previous sentence, but that connection only becomes apparent at the very end of the sentence (“The cropping should preserve all critical points. Images of the same size should also be kept by the cropping”).

3. **Out-of-sync:** the sentence is connected to a previous one, but there are disconnected sentences in between the connected sentences (“The cropping should preserve all critical points. The face features should be normalized. The cropping should also preserve all critical points”).
4. **Disconnected:** the sentence is not connected to any of the previous sentences or there are too many sentences in between.

The tool also alerts the writer when transition words such as *in addition*, *on the other hand*, or even the familiar *however* are used. Even though these expressions are effective when correctly used, they often betray the lack of a logical or semantic connection between consecutive sentences (“The cropping should preserve all critical points. However, the face features should be normalized”). SWAN displays all the sentences which could potentially break the fluidity (Fig. 2) and suggests ways of rewriting them.

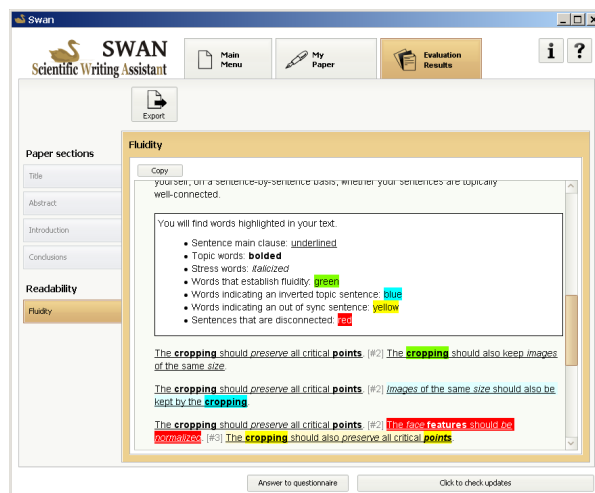


Figure 2: Fluidity evaluation result in SWAN.

4 The SWAN Tool

4.1 Inputs and outputs

SWAN operates on two possible evaluation modes: *simple* and *full*. In simple evaluation mode, the input to the tool are the title, abstract, introduction and conclusions of a manuscript. These sections can be copy-pasted as plain text to the input fields.

In full evaluation mode, which generally provides more feedback, the user provides a full paper as an input. This includes semi-automatic import of the manuscript from certain standard

document formats such as TeX, MS Office and OpenOffice, as well as semi-automatic structure detection of the manuscript. For the well-known Adobe's portable document format (PDF) we use state-of-the-art freely available *PDFBox* extractor². Unfortunately, PDF format is originally designed for layout and printing and not for structured text interchange. Most of the time, simple copy & paste from a source document to the simple evaluation fields is sufficient.

When the text sections have been input to the tool, clicking the *Evaluate* button will trigger the evaluation process. This has been observed to complete, at most, in a minute or two on a modern laptop. The evaluation metrics in the tool are straight-forward, most of the processing time is spent in the NLP tools. After the evaluation is complete, the results are shown to the user.

SWAN provides constructive feedback from the evaluated sections of your paper. The tool also highlights problematic words or sentences in the manuscript text and generates graphs of sentence features (see Fig. 2). The results can be saved and reloaded to the tool or exported to html format for sharing. The feedback includes tips on how to maintain authoritativeness and how to convince the scientist reader. Use of powerful and precise sentences is emphasized together with strategical and logical placement of key information.

In addition to these two main evaluation modes, the tool also includes a manual fluidity assessment exercise where the writer goes through a given text passage, sentence by sentence, to see whether the next sentence can be predicted from the previous sentences.

4.2 Implementation and External Libraries

The tool is a desktop application written in Java. It uses external libraries for natural language processing from Stanford, namely Stanford POS Tagger (Toutanova et al., 2003) and Stanford Parser (Klein and Manning, 2003). This is one of the most accurate and robust parsers available and implemented in Java, as is the rest of our system. Other external libraries include Apache Tika³, which we use in extracting textual content from files. JFreeChart⁴ is used in generating graphs

and XStream⁵ in saving and loading inputs and results.

5 Initial User Experiences of SWAN

Since its release in June 2011, the tool has been used in scientific writing classes in doctoral schools in France, Finland, and Singapore, as well as in 16 research institutes from A*STAR (Agency for Science Technology and Research). Participants to the classes routinely enter into SWAN either parts, or the whole paper they wish to immediately evaluate. SWAN is designed to work on multiple platforms and it relies completely on freely available tools. The feedback given by the participants after the course reveals the following benefits of using SWAN:

1. Identification and removal of the inconsistencies that make clear identification of the scientific contribution of the paper difficult.
2. Applicability of the tool across vast domains of research (life sciences, engineering sciences, and even economics).
3. Increased clarity of expression through the identification of the text fluidity problems.
4. Enhanced paper structure leading to a more readable paper overall.
5. More authoritative, more direct and more active writing style.

Novice writers already appreciate SWAN's functionality and even senior writers, although evidence remains anecdotal. At this early stage, SWAN's capabilities are narrow in scope. We continue to enhance the existing evaluation metrics. And we are eager to include a new and already tested metric that reveals problems in how figures are used.

Acknowledgments

This work of T. Kinnunen and T. Kakkonen were supported by the Academy of Finland. The authors would like to thank Arttu Viljakainen, Teemu Turunen and Zhengzhe Wu in implementing various parts of SWAN.

References

- [Aluisio et al.2001] S.M. Aluisio, I. Barcelos, J. Sampaio, and O.N. Oliveira Jr. 2001. How to learn the many "unwritten rules" of the game of the academic discourse: a hybrid approach based on critiques and cases to support scientific writing. In

²<http://pdfbox.apache.org/>

³<http://tika.apache.org/>

⁴<http://www.jfree.org/jfreechart/>

⁵<http://xstream.codehaus.org/>

- Proc. IEEE International Conference on Advanced Learning Technologies*, Madison, Wisconsin, USA.
- [Chuck and Young2004] Jo-Anne Chuck and Lauren Young. 2004. A cohort-driven assessment task for scientific report writing. *Journal of Science, Education and Technology*, 13(3):367–376, September.
- [Dale and Kilgarriff2010] R. Dale and A. Kilgarriff. 2010. Text massaging for computational linguistics as a new shared task. In *Proc. 6th Int. Natural Language Generation Conference*, Dublin, Ireland.
- [Gopen and Swan1990] George D. Gopen and Judith A. Swan. 1990. The science of scientific writing. *American Scientist*, 78(6):550–558, November-December.
- [Gopen2004] George D. Gopen. 2004. *Expectations: Teaching Writing From The Reader's perspective*. Longman.
- [HOO2011] 2011. HOO - helping our own. Webpage, September. <http://www.clt.mq.edu.au/research/projects/hoo/>.
- [Klein and Manning2003] Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proc. 41st Meeting of the Association for Computational Linguistics*, pages 423–430.
- [Lebrun2011] Jean-Luc Lebrun. 2011. *Scientific Writing 2.0 – A Reader and Writer's Guide*. World Scientific Publishing Co. Pte. Ltd., Singapore.
- [Page1966] E. Page. 1966. The imminence of grading essays by computer. In *Phi Delta Kappan*, pages 238–243.
- [Paquot and Bestgen2009] M. Paquot and Y. Bestgen. 2009. Distinctive words in academic writing: A comparison of three statistical tests for keyword extraction. In A.H. Jucker, D. Schreier, and M. Hundt, editors, *Corpora: Pragmatics and Discourse*, pages 247–269. Rodopi, Amsterdam, Netherlands.
- [Toutanova et al.2003] Kristina Toutanova, Dan Klein, Christopher Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proc. HLT-NAACL*, pages 252–259.