UNIVERSITY OF TURKU
Faculty of Science

Dissertation Review

Author: Ing. Josef Rukavička
Title: Enumeration of Factors in Special Languages
Reviewer: Jarkko Peltomäki, Ph.D., University of Turku, Finland

Content and Contributions

The dissertation Enumeration of Factors in Special Languages of Mr. J. Rukavička (hereby "the author") concerns mainly enumeration problems and extension and transitivity problems in formal language theory. The thesis consists of an introduction to nine scientific papers by the author including the papers as an appendix. Five of the papers concern rich words and palindromes; the four other topics of closed and privileged words, transition property of power-free languages, de Bruijn graphs, and dissection of languages each correspond to one paper. These topics are only loosely related, or related by a rather general theme, so it could be argued that the dissertation is a combination of research on five different topics with focus on rich words and palindromes. Out of the nine papers two are published in peer-reviewed international journals, four in peer-reviewed conference proceedings, and two papers are currently unpublished. The publication venues are respected and well-known journals/conference proceedings. The results are presented and proved rigorously adhering to the usual mathematical standards.

The presented research solves (not always completely) research questions proposed by other scientists. Regarding rich words, the author shows in [[Ru02]] that the number of rich words grows subexponentially. This confirms a version of a conjecture of Guo, Shallit, and Shur from 2016. This is very interesting as quite often languages of interest exhibit exponential growth as happens in the case of power-free words which are also studied in the dissertation. Moreover, this indicates that palindromic richness is strictly a low-complexity phenomenon. This proof of subexponential growth shows that a rich word has subexponential factor complexity but does not provide any bounds. This is remedied in the author's other significant investigation [[Ru03]] where he provides an explicit upper bound for the factor complexity of a rich word. This is a key result in the author's proof on how to decide if two finite rich words can be factors of a common word found in [[Ru04]]. This results in a rather theoretical decision algorithm but provides first answers to this transition problem proposed in 2017 by Pelantová and Starosta based on a 2014 work of Vesti. The author also provides new results on extensions of rich words [[Ru05]] (unpublished) improving on previous results of Vesti and studies the structure of infinite words with bounded palindromic

length [[Ru08]] related to a 2012 conjecture of Frid et al. These latter problems are worth studying, but they are in my opinion not as significant as the preceding ones. In my opinion, the author has attacked several open problems in the theory of rich words and palindromes and made good progress on them. Rich words were introduced slightly over ten years ago, and the research interest has waned during the past few years (perhaps because others could not make progress), but I nevertheless find them a topical and interesting subject.

In the paper [[Ru06]], the author considers a problem of Peltomäki which asks for an upper bound for the number of privileged words. The author finds such a bound for the number of closed words and this yields an upper bound for the number of privileged words as well. Privileged and closed words have attracted much attention recently, and they have connections to rich words.

The paper [[Ru07]] solves partially an old problem of Restivo and Salemi proposed in 1985. In the paper, the author proves that an extension word always exists for a pair of left-extendable and right-extendable α-power free words when $\alpha \geq 2$ and at least 3 letters are used. This is a significant advance as previously the problem had only been solved for $\alpha = 2^+$ and 2 letters (Restivo-Salemi 1985) and $\alpha = 3$ and arbitrarily many letters (Restivo-Salemi 1985, Petrova-Shur 2019). Given that these problems have been open for so long, it is rather surprising that the author's proof does not use any deep results of combinatorics on words. Even though the problem itself is somewhat old, power-free words is one of the central topics in combinatorics on words, so it is likely that the results are well-received in the community.

The paper [[Ru01]] considers a problem of Stanley on a bijective proof of a formula concerning de Bruijn graphs. The author's construction gives a new solution to the problem. I find this paper to be disconnected from the other topics of the dissertation as it is mainly concerned with graphs, not words. I find Stanley's problem to be a good combinatorial question, but I do not find it as significant as the other problems considered in the dissertation. The final (unpublished) paper [[Ru09]] considers yet another problem: dissection of languages. The notion of dissection was introduced in 2013 by Yamakami and Kato, and they showed various properties of this notion and proved that several well-known classes of formal languages are dissectible by regular languages. The author proves that a language whose growth is bounded by a tetration function is dissectible by an intersection of context-free languages. I find the notion of dissectibility interesting, but I remark that that the topic is very disconnected from the main focus of the dissertation.

Presentation of the Results

The English used is adequate and the overall quality of the text is good. The main results are clearly stated in the text. The dissertation does not contain an explicit goal, and this can be

detected as lack of cohesion in the included topics. There are indeed two topics (De Bruijn graphs and dissection of languages) that do not fall under the themes of enumeration and extension and transitivity questions. In my opinion, omitting these topics would have improved the cohesion without significantly reducing the merits of the dissertation.

There is some criticism to be made about the style of the dissertation. It seems that the author prefers to write succinctly with minimum amount of explanatory text and references. Without adequate amount of signposting, it is difficult for the reader to decipher the main points of the proofs and techniques. Explanation is attempted in the introduction but, to be honest, it is vague and hard to understand even after reading the papers. The proofs surely speak for themselves, but I think there is a missed opportunity of communication: what exactly makes the proofs work and which parts could be improved? I would have been interested in reading discussion on what sort of improvements would be needed and where in order to improve the presented bounds. I also find it surprising that often the introduction part of the dissertation leaves out parts of the introductions of the actual papers. I would expect the introduction to be more comprehensive with additional explanations and examples. In fact, the introduction contains only two or three short examples to clarify the main concepts; there are however several examples and figures in the papers. This succinctness also leads to lack of motivation. I doubt a potential reader would be happy if he has to read the actual papers and their references to dig up a historical perspective on the problems. I recommend that the author takes helping out his readers more seriously in his future works. Surely it can be unmotivating work, but this likely wins some readers and citations in the long run.

Overall Evaluation

Despite the above criticism, my overall evaluation is that the author's dissertation adequately presents new, valid, and significant scientific results. The dissertation shows without a doubt that the author is capable of conducting independent research especially since he has authored all of the papers by himself. I have no comment on the methods used. I recommend the thesis for defense.

Questions

- Regarding your upper bound on the number of palindromes in a rich word, how could the bound be improved? What are the crucial parts of your argument where an improved analysis would lead to a better overall bound? Do you have an idea what the optimal upper bound would be?
- Regarding dissection of languages, you show that every language with growth bounded by a tetration function is dissectible by an intersection of context-free languages. The intersection of

context-free languages is not necessarily context-free, so what can you say about the intersection here? What language class does it belong to? Do you know of a language that is not dissectible by an intersection of context-free languages?

Remarks

I have found the following issues.

- At the end of p. 21 there is an intended block stating Stanley's problem. The indentation and typesetting make it seem like a direct quote from Stanley, but it is not (or I was unable to find it). I recommend to indicate this to avoid any misinterpretation. Also: B(n) is the set of all binary de Bruijn sequences of *degree n*.
- On l. 4 of p. 22, the author writes "we solved this open problem of Stanley…". This is true, but the author was not the first to do this. This should definitely be edited (with appropriate references) to convey the fact that this problem was first solved in 2009 and the author provides a new proof.

Turku 21th April 2021

Jarkko Peltomäki, Ph.D.
Department of Mathematics and Statistics
20014 University of Turku
Turku, Finland