A Review of the PhD Thesis of
Jiří Koutný
Entitled
Grammars with Restricted Derivation Trees

The thesis concentrates on studying properties of grammars with derivation trees restricted in several different ways. The main focus is on fundamental properties of these grammars, such as their expressive power or closure properties. Apart from that, syntax analysis based on one of the considered types of grammars is also considered. *The subject of the thesis is therefore undoubtedly in the scope of computer science.* Moreover, despite their long use in computer science, grammars and their properties are still an active research subject studied in papers published at renowned international conferences as well as in established international journals focusing on theoretical computer science. *The research described in the thesis is thus clearly up-to-date.*

The thesis is divided into 10 chapters (some of which are, however, very brief). The original contribution of the author's research is described in Chapters 7, 8, and 9.

In Chapter 7, a new notion of *tree controlled grammars with restricted cuts of derivation trees* is introduced. Subsequently, the generative power of the introduced grammars is characterized. Chapter 8 is devoted to *path tree controlled grammars*. Several new results concerning the generative power, normal forms, and the relation of these grammars to pseudo-knots are presented. The chapter also provides a counter-argument to one earlier result on the generative power of path tree controlled grammars. Chapter 9 concentrates on *grammars controlled by restricting multiple paths* in their derivation trees which is an original direction since earlier works studied restrictions on one path only. The chapter comes with new results on the pumping and closure properties of n-path tree controlled grammars and their expressive power. Moreover, original algorithms for syntax analysis based on these grammars are proposed.

Multiple of the results presented in the thesis are not particularly deep, e.g., the results on the generative power, normal forms, and pseudo-knots presented in Chapter 8, or the closure properties, generative power approximation, and the complexity theoretic arguments concerning syntax analysis in Chapter 9. However, in total, *the original contribution of the results presented in the thesis is significant*.

The contents of the thesis is based on four publications at the local student conference EEICT, one publication at the MEMICS workshop, and—most importantly—on several journal papers: two published (Kybernetika, Theoretical and Applied Informatics), one accepted (Schedae Informaticae), and two submitted (Acta Cybernetica). Although the journals in which the results were published do not really belong among highly ranked journals, they still represent quite valuable publication venues, and *the contents of the thesis can therefore be clearly considered to be sufficiently published*.

The thesis is written in English of a reasonable level. However, despite I am not a native speaker, and my own English is far from perfect, I could still identify quite some errors in English in the thesis (concerning, e.g., the use of articles, which are sometimes over-used and sometimes missing, a quite frequent improper use of "being", etc.).

I appreciate that the author provides precise formal descriptions of most of the studied issues. Nevertheless, some more intuition could sometimes be added. Also, some of the long lists of definitions and theorems in Chapters 3 and 4 seem rather scary. If it is not essential to refer to some definition or theorem by its number, I would prefer to group them and embed them into plain paragraphs. Finally, I do not like the fact that some notions are defined repeatedly. This is especially the case of the definitions in Chapter 6, but also, e.g., Def. 8.1. Chapter 6 is, moreover, rather short, and I do not see a reason for having it as a separate chapter. Finally, sometimes,

I find the text somewhat hard to follow (e.g., “a graph is called a directed graph”, the name “tree controlled grammars under not common n-path control”, the sentence below Definition 7.1, “it holds” in the closure properties on page 68, “can be restricted basically by” on page 75, etc.).

Concerning technical correctness of the text, I have to admit that I was not able to check every tiny detail of each proof. However, I have not found any major problem in any of the parts that I have read in detail. On the other hand, the text contains multiple places that seem to me to contain minor problems. Here is a list of some of such places:

1. In Section 2.1 devoted to sets within an otherwise rather formal text, I would expect at least some note about the existence of formal set theories.

2. The Pentonnen normal form for context-sensitive grammars and the Chomsky normal form for context-free grammars are defined such that they do not allow for having the empty string in a language.

3. Matrix grammars are defined as 5-tuples consisting of 4 elements.

4. In the definition of $GSM_M(u)$ (Def. 3.104), τ is applied on strings u whereas it is defined for single letters only (Def. 3.102). Likewise, the use of u in the definition of rational transduction is not proper.

5. The introduction of Section 4.1 promises “strictly formal definitions”, Def. 4.2 is, however, rather textual.

6. In Definition 7.3, ${}_xM$ should probably be a set, not a sequence.

7. In Definition 8.4, E is sometimes used as a set and sometimes as a an alphabet symbol.

8. Lemma 8.1 excludes empty strings from the languages of the considered grammars. Moreover, the proof of the lemma assumes G to be context-free, but this assumption is not stated in the lemma.

9. In Algorithm 1, M is defined such that $RT_M(R) = R$, but the proof of Theorem 8.3 defines R in a different way.

10. The proof of Theorem 8.7 refers to Theorems 9.3 and 9.4 which is probably wrong.

11. The proof of Theorem 8.9 defines a matrix grammar as a set. Moreover, the definition of M contains several wrong uses of the overline and hat symbols.

12. The paragraph on pumping properties in Section 9.1 states that pumping lemmas can be used for determining that a language does not belong to a certain class of languages only. For regular languages, however, there is a pumping lemma that is both necessary and sufficient (cf. “A necessary and sufficient pumping lemma for regular languages”, Jeffrey Jaffe, ACM SIGACT News,10(2):48–49, 1978).

13. In Definition 9.9, the requirement on $word(p)$ to satisfy the premise of Lemma 3.2 is unclear.

14. In Theorem 9.11, I do not think that a union itself specifies any hierarchy.

15. The assumption that R is generated by an unambiguous linear grammar should be stated in Theorem 9.15, not just in its proof. Also, in the first paragraph of “Top-Down Parsing of nc-n-path-TC(CF,LIN)”, R should be generated by an unambiguous linear grammar.

16. The informal description of the idea of the top-down parsing is rather unclear, especially the end of page 73 and beginning of page 74: E.g., in the phrase “the pointer to the symbol”, which symbol is meant? The best would be to complement the informal description by an algorithm in pseudo-code. Further, it seems to me that the text refers to the “second automaton” as though it was described in the text, but it is not.

17. In the description of the bottom-up parsing on page 75, the reader is sometimes left with the feeling that unit productions are solely of the form $A \rightarrow A$. Moreover, the link of the description of removing unit productions with the stated theorems is not very clear. Finally, stating that there is no derivation of the form $B \Rightarrow^* A$ in G' is not correct since you do not require $A \neq B$.

Further, I would like the author to answer the following questions during his defence:

1. Theorem 9.2 allows $4n - 1$ of the pumped sub-words to be empty. Should not the requirement on the non-emptiness of these sub-words be strengthened?

2. Are you aware of some practical scenarios in which an application of syntax analysis based on the algorithms described in Chapter 9 would help?

To sum up, despite the above presented criticism, *the PhD thesis of Jiří Koutný proves*, in my view, *a sufficient level of scientific erudition of its author*. *The thesis meets the standards needed to award its author the PhD degree.* Therefore, I recommend the thesis to be accepted for defence, and upon its successful completion, Jiří Koutný to be assigned the PhD degree.

Brno, July 31, 2012

Prof. Ing. Tomáš Vojnar, Ph.D.

Faculty of Information Technology
Brno University of Technology
Božetěchova 2, CZ-612 66, Brno