AN ANNOTATED BIBLIOGRAPHY OF PUMPING

ANTON WIJHOLT

Faculty of Science
Informatics Department
Nijmegen University
The Netherlands

FIRST EDITION
JUNE 1982
YABBER—Yet Another Bibliography: Pumping Lemma's.

AN ANNOTATED BIBLIOGRAPHY OF PUMPING
(First Edition, June 1982)

Anton Wijholt
Faculty of Science, Department of Informatics
Nijmegen University, The Netherlands

INTRODUCTION

This bibliography contains references to papers which are concerned with 'pumping lemma's' (iteration theorems, intercalation theorems, uvwxy-theorems) for context-free and several other kinds of grammars and languages. Pumping lemma's describe necessary properties of certain classes of languages. Pumping lemma's have been obtained for different classes of languages. Therefore they can be used to show that a certain language does not belong to a given class since the language does not satisfy the pumping lemma for that class of languages. Hence, instead of proving such results with ad hoc arguments, comparatively routine proofs can be given.

Probably the first time pumping arguments for context-free grammars have been used was by Scheinberg[B37] in 1960 (observed by Greibach[B23]). Scheinberg shows that the language \( a^n b^n a^n \mid n > 0 \) is not context-free. In 1961 (Bar-Hillel, Perles and Shamir[B3]) the pumping lemma for context-free grammars was presented.

In this annotated bibliography we distinguish between the following five categories:
YABBER-Yet Another Bibliography: Pumping Lemma's.

A. Pumping and regular grammars.

B. Pumping and context-free grammars.

C. Pumping and 'deterministically parsable' grammars.

D. Pumping and other rewriting systems.

E. Related topics (selective).

In the following diagram the growth of interest in pumping lemma's is shown.

<table>
<thead>
<tr>
<th>YEAR</th>
<th>NUMBER OF REFERENCES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1959-1960</td>
<td>***</td>
</tr>
<tr>
<td>1961-1962</td>
<td>***</td>
</tr>
<tr>
<td>1963-1964</td>
<td>**</td>
</tr>
<tr>
<td>1965-1966</td>
<td>***</td>
</tr>
<tr>
<td>1967-1968</td>
<td>********</td>
</tr>
<tr>
<td>1969-1970</td>
<td>********</td>
</tr>
<tr>
<td>1971-1972</td>
<td>********</td>
</tr>
<tr>
<td>1973-1974</td>
<td>********</td>
</tr>
<tr>
<td>1975-1976</td>
<td>*******************</td>
</tr>
<tr>
<td>1977-1978</td>
<td>*******************</td>
</tr>
<tr>
<td>1979-1980</td>
<td>*******************</td>
</tr>
<tr>
<td>1981-1982</td>
<td>*******************</td>
</tr>
</tbody>
</table>

ACKNOWLEDGEMENTS

I'm grateful to K. Culik II, J. Engelfriet, J. v. Leeuwen, A. Lindenmayer, G. Rozenberg and Th.A. Zoethout for pumping references into this bibliography. Corrections and additions are welcome and should be sent to the author.
A. PUMPING AND REGULAR GRAMMARS

Rabin and Scott[A4] introduced a pumping property for regular sets. This "Classical Pumping Lemma" for regular sets gives a necessary condition on the set of strings of a regular language. In a weak form this property says that if a language L is regular then there exists a constant p such that any word z in L with |z| ≥ p can be written as z = uvw such that 0 < |v| < p and such that for all i ≥ 0, uv^iw is in L.

This pumping property is not a sufficient condition, that is, the conditions of the property do not imply regularity. In Kioke[B26] a non-regular language is given which satisfies the conditions of the property. In Jaffe[A2], Yehudai[A8], Ehrenfeucht, Parikh and Rozenberg[A1], Stanat and Weiss[A7] and Sommerhalder[A5,A6] necessary and sufficient conditions are given which imply regularity.

The paper of Kintala and Wotschke[A3] gives a nice example of the use of the pumping lemma for context-free grammars in order to obtain results for regular sets.

REFERENCES


B. PUMPING AND CONTEXT-FREE GRAMMARS

Bar-Hillel, Perles and Shamir's Characterization

Bar-Hillel, Perles and Shamir[B3] introduced a theorem which has become known as the "Classical Pumping Lemma" for context-free languages. This lemma describes a necessary condition on the set of strings of a context-free language. Therefore it can be used to show that certain languages are not context-free. In a weak form this classical lemma says that for any context-free language there is a constant n, such that if z is in L and |z| > n, then we may write z = uvwx such that 1. |vwx| > 0, 2. |vwx| < n, and 3. for all i > 0, uv^iwx^iy is in L. A similar lemma can be formulated for the linear languages (see the exercises in e.g. [B24]).
Ogden's Lemma

Ogden's Lemma obtains a slightly stronger result ("Ogden's Lemma"), that is, there exist non-context-free languages which satisfy the "Classical Pumping Lemma" while they do not satisfy "Ogden's Lemma". Moreover, inherent ambiguity of certain context-free languages can be shown using "Ogden's Lemma". Ogden's result is also valid for the sentential forms of a context-free grammar (see Harrison[24]). In a weak form Ogden's Lemma can be given as follows. Let L be a context-free language. Then there is a constant n, such that if z is in L and we mark any n or more positions of z "distinguished", then we can write z = uvwxy, such that 1. either each of u, v and w contains a distinguished position or each of w, x and y contains a distinguished position, 2. vwx has at most n distinguished positions, and 3. for all i > 0, uv^iwx^iy is in L.

In Bader and Moura[2] a further generalization of Ogden's lemma is given. A non-context-free language is presented which satisfies Ogden's lemma but which does not satisfy this generalization.

Sokolowski's Criterion

Sokolowski[38] introduced an 'almost' pumping result for context-free languages. A criterion is given which can be used to show that a certain language is not context-free. An example of a non-context-free language is given which satisfies the conditions of the "Classical Pumping Lemma", but which does not satisfy his criterion. Sokolowski's criterion is the following. If L \subseteq \Sigma^* is context-free, then for every subset I of \Sigma containing at least two distinct letters (say a and b) and for every u_1, u_2, u_3 in \Sigma*, if \{u_1xu_2xu_3 \mid x \in \Sigma^+\} \subseteq L, then there exist two different words x', x'' in \Sigma^+, such that u_1x'u_2x''u_3 is in L. Sokolowski raises the question whethe-
er this condition is not only necessary but also sufficient. In Nijholt[828] it is shown that this is not the case.

Interchange Lemma's

Ogden, Ross and Winklmann[832] prove an "Interchange Lemma" which gives a necessary condition for languages to be context-free. Similarly as in the case of Sokolowski's criterion their lemma can be used to handle non-context-free languages of the form $uvwv$. Their lemma is a generalization of a technique used in Ross and Winklmann[836] to show that repetitive strings are not context-free. In Ehrenfeucht and Rozenberg[85] the same problem has been discussed.

Pumping Languages

The "Classical Pumping Lemma" for context-free languages does not characterize the context-free languages. That is, if for a certain language the conclusions of the lemma are satisfied then it is not necessarily the case that the language is context-free. A similar remark holds for "Parikh's Theorem" (Parikh[833], B34), van Leeuwen[827]), an other result which can be used to show that certain languages are not context-free. In Goldstine[822] the "Classical Pumping Lemma" is strengthened and then used to give a simplified proof of "Parikh's Theorem". Wise[841] presents a "Strong Pumping Lemma" (which uses sentential forms), which does characterize the context-free languages. Moreover, in this paper it is mentioned that it is not known whether "Ogden's Lemma" characterizes the context-free languages.

The family of languages which satisfy the conditions of the "Classical Pumping Lemma" has been studied by Horvath[825], Cislaru and Paun[815], Coardos[816] and Paun[835]. In Kloze[826] more general classes of 'pumping'
languages are studied. Boasson[B12] and Boasson and Horvath[B13] show that various types of non-context-free languages can be constructed which satisfy "Ogden's Lemma".

In the following diagram we have summarized some of the remarks on pumping languages. REG is short for regular languages, CFL for context-free languages, Scott-Rabin-like stands for those languages which satisfy the conditions of the pumping lemma for regular languages, Ogden-like stands for languages which satisfy the conditions of Ogden's lemma and Bar-Hillel-like for languages which satisfy the conditions of the "Classical Pumping Lemma". The class of languages satisfying the generalized version of Ogden's lemma (Bader and Moura[B2]) is situated between CFL and Ogden-like. All the inclusions which are shown are proper.
Iterative Pairs

Iterative pairs have been introduced by Boasson[B7]. A factorization $z = xuwy$ of a sentence $z$ in $L$ is said to be an iterative pair if for all $i \geq 1$ $xu^iwy^i$ is in $L$, with $|uv| > 0$. With the help of iterative pairs Boasson has studied pumping properties of some subclasses of the context-free languages (see [B6,B8,B9]) and relations between iterative pairs of two context-free languages (see [B11]). Beauquier[B5] studies similar relations for systems of iterative pairs. One well-known result of Boasson[B7] is the following. Let $L$ be a context-free language. If for each iterative pair $xuwy$ in $L$ also $xu^iwy^i$ is in $L$, then $L$ is regular. This result is considerably strengthened in Ehrenfeucht and Rozenberg[B20] by allowing $i \geq 0$ rather than $i \geq 1$ in the definition of an iterative pair.

The question whether a context-free language is in fact a regular language has been studied by various authors. Chomsky[B14] introduced the notion of of this question. Ehrenfeucht and Rozenberg[B19,B20] give other conditions on context-free grammars which 'enforce' regularity. Culik II et al[B17] are also concerned with properties which make it possible to distinguish between the regular and the non-regular component of a context-free language.

Programming Languages

The "Classical Pumping Lemma" for context-free languages has been used (Floyd[B21]) to show that ALGOL 60 is not context-free. Pumping lemma's are not the only way to show that a (programming) language is not context-free. Other useful tools are generalized sequential machine mappings or intersection with regular sets. Sokolowski[B36] introduced his criterion to show that declarations in a programming language lead to non-context-free languages.
REFERENCES


YABBER—Yet Another Bibliography: Pumping Lemma's.


TABBED—Yet Another Bibliography: Pumping Lemma's.


C. PUMPING AND 'DETERMINISTICALLY PARSAble' GRAMMARS

From the point of view of parsing the classical subclasses of the context-free grammars are the classes of LL(k), simple precedence and LR(k) grammars. LR(k) grammars generate exactly the deterministic languages. LL(k) and simple precedence grammars generate proper subclasses of the deterministic languages. For LL(k) grammars this has been shown in Rosenkrantz and Stearns[C14] and, e.g.,
van Leeuwen[C11]. A rigorous proof of this fact can be
given with pumping lemma's for LL(k) grammars, cf.
Beatty[C1,C2,C3]. In Nijholt[C12] it is shown, among oth-
ers, that the conditions in the LL-Iteration Theorem of
Beatty are not sufficient to imply the LL(k) property of a
language. The proper inclusion of simple precedence
languages in the class of deterministic languages has been
shown by Fischer[C4], Rosenkrantz, Lewis and Stearns[C13]
and, with the help of pumping lemma's, Krevner and
Yehudai[C10]. These lemma's for LL(k) and simple prece-
cedence languages can also be used to show that certain
features in programming languages are not LL(k) or simple
precedence.

Ogden[B29] has given a proof for a pumping lemma
for deterministic languages using a pushdown automaton.
Harrison and Havel[C5,C6] have given improved and machine
independent proofs of pumping lemma's for deterministic
and strict deterministic languages. King[C8] continues
Harrison and Havel's work by presenting pumping lemma's
for a hierarchy of strict deterministic languages, includ-
ing a lemma for the class of simple deterministic
languages (cf. Korenjak and Hopcroft[C9]). In
Sakarovitch[C15] deterministic iterative pairs are intro-
duced and then pumping properties for deterministic
languages are established. Heilbrunner[C7] generalizes
the results of Beatty and Harrison and Havel by giving
pumping lemma's for LL-regular and LR-regular languages.

REFERENCES

C1. J.C. Beatty. Iteration theorems for LL(k) languages.
Proc. of the Ninth Annual Symp. on Theory of Comput-


C11. J. van Leeuwen. An elementary proof that a certain context-free language is not LL(k), and a generalization. Notes, 1972.
YABBER—Yet Another Bibliography: Pumping Lemma's.


D. PUMPING AND OTHER REWRITING SYSTEMS

Many of the properties which are known for the class of context-free languages have been generalized or converted into properties for other classes of languages generated by rewriting systems. In this section we have collected papers on pumping properties for less classical rewriting systems. Results are given for:

- indexed languages [D7]
- simple matrix languages [D8]
- coupled languages [D20]
- tree languages [D1,D2,D15,D16,D19]
- graph languages [D9,D12]
- Lindenmayer systems [D3,D4,D5,D6,D11,D13,D14,D18]
- derivation trees [D10]
- power series [D17]

REFERENCES


YABBER—Yet Another Bibliography: Pumping Lemma's.


E. RELATED TOPICS (selective)

REFERENCES


YABBER—Yet Another Bibliography: Pumping Lemma's.