INFORMATIQUE THÉORIQUE ET APPLICATIONS

VOLKER DIEKERT
AXEL MÖBUS

Hotz-isomorphism theorems in formal language theory

Informatique théorique et applications, tome 23, nº 1 (1989), p. 29-43
<http://www.numdam.org/item?id=ITA_1989__23_1_29_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 23, n° 1, 1989, p. 29 à 43)

HOTZ-ISOMORPHISM THEOREMS IN FORMAL LANGUAGE THEORY

by Volker DIEKERT ([1]) and Axel MÖBUS ([2])

Abstract. – *A language $L \subseteq X^*$ is called a language with a Hotz-isomorphism if there is a generating grammar $G=(V, X, P, S)$ such that its Hotz group $F(V \cup X)/P$ is canonically isomorphic to $F(X)/L$.*

The main result of the paper states that L is a language with a Hotz-isomorphism if and only if $F(X)/L$ is a finitely presentable group. We also prove an analogous result for Hotz monoids and discuss cancellative quotients.

Further, we investigate the effectiveness of our construction of the grammars which allow the Hotz-isomorphism and the preservation of the type of the grammars. Finally, we prove various undecidability results concerning this construction.

Résumé. – *Un langage $L \subseteq X^*$ est dit un langage avec isomorphisme de Hotz, s'il existe une grammaire $G=(V, X, P, S)$ engendrant L, telle que son groupe de Hotz $F(V \cup X)/P$ est canoniquement isomorphe à $F(X)/L$.*

Le résultat principal de ce travail montre que L est un langage avec isomorphisme de Hotz si et seulement si $F(X)/L$ est un groupe finiment présentable.

On montre un résultat analogue pour les monoïdes de Hotz et on discute des quotients cancellatifs. Nous examinons l'effectivité de notre construction des grammaires qui réalisent l'isomorphisme de Hotz et la préservation du type des grammaires. Finalement nous montrons divers résultats d'indécidabilité en rapport avec la construction.

INTRODUCTION

The equivalence of two (context-free) grammars is undecidable in general. One attempt to overcome this difficulty is to look for algebraic invariants which then can be investigated by means of mathematical methods. Very natural algebraic candidates which one can associate to a given grammar $G=(V, X, P, S)$ are the so-called Hotz monoid $M(G)=(V \cup X)^*/P$ which is

([1]) Institut für Informatik der Technischen Universität München, Arcisstr. 21, D-8000 München 2, R.F.A.
([2]) Rechenzentrum der Universität Düsseldorf, Universitätsstr. 1, D-4000 Düsseldorf 1, R.F.A.

Informatique théorique et Applications/Theoretical Informatics and Applications
0296-1598 89/01 29 15/$3.50/© Gauthier-Villars

the free monoid over $V \cup X$ modulo the productions P and the Hotz group $H(G) = F(V \cup X)/P$ which is the free group over $V \cup X$ modulo P.

In [5] Hotz showed that the Hotz group of a reduced context-free grammar depends on the generated language only. Frougny *et al.* proved in [4] that the Hotz group is in this case canonically isomorphic to the free group over the terminals modulo the generated language. Thus, two equivalent grammars have canonically isomorphic Hotz groups. Now, groups have a lot of computable invariants which can be used to get information about the language. For example, we may compute cohomology groups. One obtains thereby lower bounds for the number of productions which are needed to generate the language, *see* [2] for applications of this. Languages with a Hotz-isomorphism (on group level) are recursively enumerable languages $L \subseteq X^*$ where there is a generating grammar G such that the groups $F(X)/L$ and $H(G)$ are canonically isomorphic. Clearly, if a language $L \subseteq X^*$ belongs to this class then $F(X)/L$ is finitely presentable. Here we prove the converse: if $F(X)/L$ is finitely presentable then there is a generating grammar where the Hotz group is canonically isomorphic to $F(X)/L$. If the group $F(X)/L$ is given by any finite set of relations and if we know whether L is empty or not, then we can effectively construct a generating grammar which allows the Hotz-isomorphism. Thus, few information about L is enough to construct such a grammar. In fact, we prove a stronger result which implies that an analogous assertion holds for Hotz monoids, too. We also show that the class of languages with a Hotz-isomorphism is closed under inverse surjective homomorphisms whereas this class was known not to be closed under inverse homomorphisms in general.

In paragraph 3 we investigate Hotz-isomorphism theorems for cancellative monoids. In paragraph 4 we discuss the relationship between the type of the language and the type of the grammar which gives the Hotz isomorphism. Our construction yields a grammar which is not context-sensitive even if we start with a regular grammar, in general. However, as long as the grammar is context-free, it is well known that a suitable grammar of the same type is obtainable simply by reducing. For context-sensitive languages L we do only have an analogous result of type preserving, if the collapsing group $F(X)/L$ is finite. For $F(X)/L$ infinite there are very simple examples of context-sensitive languages which have a Hotz-isomorphism but no generating context-sensitive grammar which yields it.

In Part II we show that for a given grammar without any further information the following questions will be undecidable in general: Does the grammar itself have a Hotz-isomorphism? Does the grammar produce a language with a Hotz-isomorphism? However, in some sense the first question is simpler than the second. We shall prove that the first one is semi-decidable, whereas the second question is not. Finally, we solve another open problem: we show that there is no algorithm which transforms every given grammar into an equivalent grammar with a Hotz-isomorphism, if such a grammar exists and which may produce any garbage otherwise.

The present paper is an extended version of our paper presented at the conference STACS 88 at Bordeaux. The paragraphs 3 and 4 are new, so is Theorem 18 and the proof of Theorem 17 which is much simpler than the original one given in the proceedings. We would like to acknowledge a remark of a referee of STACS 88, which stimulated the work in paragraph 4. We also thank Klaus-Jörn Lange for helpful discussions.

NOTATIONS

For the algebraic concepts used in the paper the reader is referred to [1]. The language theoretic background may be found in any textbook on formal languages, e. g. in [6].

By X we denote a finite alphabet. A grammar (over X) is a quadruple $G=(V, X, P, S)$ where V is a set of variables with $V \cap X=\varnothing$, $P \subseteq (V \cup X)^* \times (V \cup X)^*$ is a finite set of productions, and $S \in V$ is an axiom. The generated language of a grammar G is $L(G)=\{w \in X^* \mid S \overset{*}{\underset{G}{\Rightarrow}} w\}$. By a language we always understand subsets $L \subseteq X^*$ of the form $L=L(G)$ for some grammar G. If M is a monoid and $R \subseteq M \times M$ is a set of relations then M/R means the quotient of M by the congruence generated by R. If $R=L \times L$ for some subset $L \subseteq M$ then we shall write M/L instead of $M/L \times L$. The cancellative quotient of a monoid M is $C(M)=M/Q$ where Q is the intersection of all congruences having the property that for all $u, v, x, y \in M$ it holds: if uxv and uyv are congruent, then x and y are congruent. Finally, if X is any set then $F(X)$ means the free group over X. By $\mathbf{Z}$ we denote the group of integers.

I. ALGEBRAIC PART

I.1. The main Theorem

Let $G=(V, X, P, S)$ be a grammar and $L=L(G)$ its generated language. We associate the following algebraic objects: $M(G)=(V \cup X)^*/P$ and

$H(G)=F(V \cup X)/P$, where $M(G)$ is called the Hotz monoid of G and $H(G)$ is called the Hotz group. There is a natural mapping of X to these objects which factorizes through the quotient X^*/L. More generally, by an X-monoid (X-group resp.) we mean a monoid (group resp.) M together with a mapping $i_M : X \to M$. A morphism of X-monoids is a homomorphism $f : M \to N$ such that $i_N = f\, i_M$. Also, if M is an X-monoid and $f : M \to N$ is any homomorphism to a monoid N then N will be viewed as an X-monoid with respect to the composition $f\, i_M$. We shall speak of N as an X-monoid without specifying f, if f is clear from the context. The central example of X-monoids and X-morphisms is given in the following commutative diagram where all arrows are canonical:

$$\begin{array}{ccc} X^*/L & \xrightarrow{\quad i \quad} & M(G) \\ \downarrow & & \downarrow \\ C(X^*/L) & \xrightarrow{\quad i_C \quad} & C(M(G)) \\ \downarrow & & \downarrow \\ F(X)/L & \xrightarrow{\quad i_H \quad} & H(G) \end{array}$$

Although we are mainly interested in the case when $i_H : F(X)/L \to H(G)$ is an isomorphism, we work with the morphisms i and i_C, too. We therefore distinguish three levels. We say that a grammar G has a Hotz-isomorphism on monoid level (cancellative level resp., group level resp.) if the X-morphism i (i_C resp., i_H resp.) above is an isomorphism. A language $L \subseteq X^*$ is called a language with a Hotz-isomorphism on monoid level (cancellative level resp., group level resp.) if there is a generating grammar with the corresponding property. Context-free languages (in fact, more generally, homomorphic images of sentential form languages) are languages with a Hotz-isomorphism on cancellative level, *see* [4], Prop. 3 ([2], Thms. 3, 4 resp.), but there are regular languages without Hotz-isomorphism on monoid level, *see* [3], Rem. 1.2. If the level is clear we sometimes omit it and we shall speak of Hotz-isomorphisms only.

The main result of the paper is stated in the next theorem. Its proof will follow from Theorem 4 and Corollary 9 below and is given in paragraph 2.

THEOREM 1: *Let $L \subseteq X^*$ be a language and let $f: X^*/L \rightarrow M$ be a homomorphism to a finitely presented monoid $M = Y^*/R$ with Y and R finite. Then there exists a grammar G generating L with Hotz-monoid X-isomorphic to M. [This means $L = L(G)$ and there is an isomorphism $\varphi: M(G) \rightarrow M$ of monoids with $\varphi i = f$.] Furthermore the construction of G is effective, provided we know any generating grammar for L and we know whether L is empty or not.*

From Theorem 1 we may directly deduce Hotz-isomorphism theorems on two levels:

COROLLARY 2: *A language $L \subseteq X^*$ is a language with a Hotz-isomorphism on monoid level if and only if X^*/L is finitely presentable.*

COROLLARY 3: *A language $L \subseteq X^*$ is a language with a Hotz-isomorphism on group level if and only if $F(X)/L$ is a finitely presentable group.*

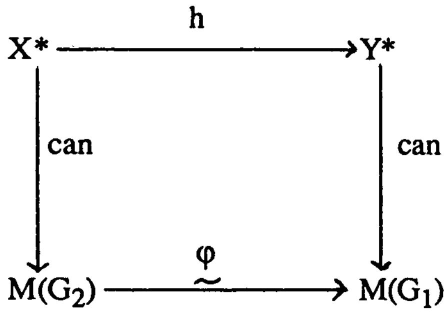
Proofs: Corollary 2 is immediate with $f = id_{X^*/L}$. For Corollary 3 observe that if $F(X)/L$ is finitely presentable as group then $F(X)/L$ is a finitely presentable monoid. Further if $M(G)$ is isomorphic to $F(X)/L$ then we have $M(G) = H(G)$. □

Remark: Corollary 3 was first stated and proved in an unpublished manuscript of the second author, [8]. It answers an open question stated in [2]. The effectiveness of Theorem 1 carries over to Corollaries 2 and 3 in the following sense: Say X^*/L ($F(X)/L$ resp.) is given by some finite set of relations over X and L is given by some generating grammar. Then we can effectively construct a grammar which generates L and with a Hotz-isomorphism on monoid level (group level resp.), provided we know whether L is empty or not.

During the proof of Theorem 1 we shall see that the type of the constructed grammar has, *a priori*, no connection with the type of the language. In paragraph 4 we shall discuss this problem in more detail.

I.2. Operations on languages and the proof of the Main Theorem

THEOREM 4: *Let $h: X^* \rightarrow Y^*$ be a homomorphism and G_1 be a grammar over Y. Then we effectively find a grammar G_2 over X generating $h^{-1}(L(G_1))$ and an isomorphism φ between $M(G_2)$ and $M(G_1)$ such that the following diagram* *commutes:*



Proof: Let $G_1=(V, Y, P, S)$ be a grammar over Y. We may assume $X\cap(V\cup Y)=\emptyset$. Define a grammar $G_2=(V\cup Y, X, P\cup\{h(x)\to x;\ x\in X\}, S)$. It follows $L(G_2)=h^{-1}(L(G_1))$. The morphism $\varphi: M(G_2)\to M(G_1)$ is induced by identity on $V\cup Y$ and by h on X; its inverse is induced by the inclusion $(V\cup Y)\subseteq(V\cup Y\cup X)$. □

Remark: The class of languages with a Hotz-isomorphism is not closed under inverse homomorphism, *see* [2], Th. 6, (ii). Theorem 4 tells us that, at least, the inverse image $h^{-1}(L)$ of a language with a Hotz-isomorphism L has a grammar whose Hotz monoid (Hotz group resp.) is X-isomorphic to that of L. But Theorem 4 tells us more. Although, in general, $F(X)/h^{-1}(L)$ is not isomorphic to $F(Y)/L$ even for surjective h, we can prove that the class of languages with a Hotz-isomorphism on cancellative level (on group level resp.) is closed under inverse images of surjective homomorphisms.

COROLLARY 5: *Let $h: X^*\to Y^*$ be a surjective homomorphism and $L\subseteq Y^*$ be a language with a Hotz-isomorphism on cancellative level (on group level resp.). Then $h^{-1}(L)$ is a language with a Hotz-isomorphism on the same level.*

Proof: Since h is surjective, for all $y\in Y$ there is a letter $x\in X$ with $h(x)=y$. Thus we may define a mapping $g: Y\to X$ such that $hg=id_Y$. Let $X_1=\text{alph}(h^{-1}(L))$ and $Y_1=\text{alph}(L)$ be the alphabets of letters occurring in the corresponding languages. It follows from the formula $hg=\text{id}_Y$ that the restriction of h to X_1^* is surjective onto Y_1^*. We next show that $C(X_1^*/h^{-1}(L))$ is X_1-isomorphic to $C(Y_1^*/L)$. Let $x\in X_1$ and $uxv\in h^{-1}(L)$. Then $gh(u)gh(x)gh(v)\in h^{-1}(L)$ and $gh(u)gh(v)\in h^{-1}(L)$. Since $C(X_1^*/h^{-1}(L))$ is cancellative, we see that x and $gh(x)$ describe the same element in $C(X_1^*/h^{-1}(L))$.

In particular, g induces a surjective homomorphism $g': C(Y_1^*/L) \to C(X_1^*/h^{-1}(L))$. Again, by the formula $hg = \mathrm{id}_Y$ we see that g', and hence $h': C(X_1^*/h^{-1}(L)) \to C(Y_1^*/L)$ is bijective. This proves the desired isomorphism. [It follows that $F(X_1)/h^{-1}(L)$ and $F(Y_1)/L$ are isomorphic, too]. The assertion of Corollary 5 now follows directly from Theorem 4 and the next easy lemma:

LEMMA 6: *Let $L \subseteq X_1^*$ be a language and $X_1 \subseteq X$. Then the language $L \subseteq X_1^*$ is a language with a Hotz-isomorphism over X_1 on any level if and only if $L \subseteq X^*$ is a language with a Hotz-isomorphism over X on the same level.*

Proof: Let $X_2 = X \setminus X_1$ then $X^*/L = X_1^*/L * X_2^*$, $(C(X^*/L) = C(X_1^*/L) * X_2^*$, and $F(X)/L = F(X_1)/L * F(X_2)$, respectively). If $G_1 = (V, X_1, P, S)$ is a grammar for $L \subseteq X_1^*$ with a Hotz-isomorphism over X_1 then $G = (V, X, P, S)$ is equivalent to G_1 and it has a Hotz-isomorphism over X. Conversely, let $G = (V, X, P, S)$ be any grammar for $L \subseteq X^*$ with a Hotz-isomorphism over X then $G_1 = (V \cup X_2 \cup \{Z\}, X_1, P \cup \{Z \to A \mid \lambda;\ A \in X_2\}, S)$ where Z is a new symbol, is equivalent to G and it has a Hotz-isomorphism over X_1. □

LEMMA 7: *Let G_1, G_2 be grammars over X. Then we effectively find a grammar over X generating the intersection $L(G_1) \cap L(G_2)$, whose Hotz-monoid is X-isomorphic to the direct product $M(G_1) \times M(G_2)$.*

Proof: Let X', X'' be two copies of X. Replacing the occurrences of $x \in X$ in the productions of G_1 and G_2 by the corresponding $x' \in X'$ and $x'' \in X''$ respectively, we obtain grammars

$$G_1' = (V_1, X', P_1', S_1), \qquad G_2'' = (V_2, X'', P_2'', S_2)$$

which generate copies of the original languages $L(G_1)$ and $L(G_2)$. We may now assume that the sets V_1, V_2, X, X', X'' are pairwise disjoint. Define the grammar $G = (V, X, P, S)$ by $V = V_1 \cup V_2 \cup X' \cup X'' \cup \{S\}$ where S is a new symbol and

$$P = \{S \to S_2 S_1\} \cup P_1' \cup P_2'' \cup \{vu \to uv;\ u \in V_1 \cup X',\ v \in V_2 \cup X''\} \cup \{x' x'' \to x;\ x \in X\}.$$

Clearly, $L(G) = L(G_1) \cap L(G_2)$. Further, $M(G)$ is canonically isomorphic to $M(G_1) \times M(G_2)$ where the X-morphism to $M(G_1) \times M(G_2)$ is induced by the diagonal $x \to (x, x)$, $x \in X$. □

THEOREM 8: *Let $G = (V, X, P, S)$ be a grammar and $p: M(G) \twoheadrightarrow Y^*/R$ be a surjective X-homomorphism onto a monoid, where Y and R are finite and* *effectively given. Then there exists effectively a grammar G' which generates $L(G)$ such that its Hotz monoid is X-isomorphic to Y^*/R.*

Proof: Since Y^*/R is finitely presented there exists a finite set $Q \subseteq (V \cup X)^* \times (V \cup X)^*$ such that p induces an X-isomorphism $(V \cup X)^*/P \cup Q \cong Y^*/R$. This set $Q \subseteq (V \cup X)^* \times (V \cup X)^*$ is effectively given as follows:

We may lift $p: M(G) \to Y^*/R$ to a homomorphism $p': (V \cup X)^* \to Y^*$. Since R is finite, we find for each $y \in Y$ a word $s(y) \in (V \cup X)^*$ such that y and $p' s(y)$ denote the same element in Y^*/R. We extend s to a homomorphisms: $Y^* \to (V \cup X)^*$. Then we may take

$$Q = \{ (z, sp'(z)); z \in (V \cup X) \} \cup \{ (s(u), s(v)); (u, v) \in R \}.$$

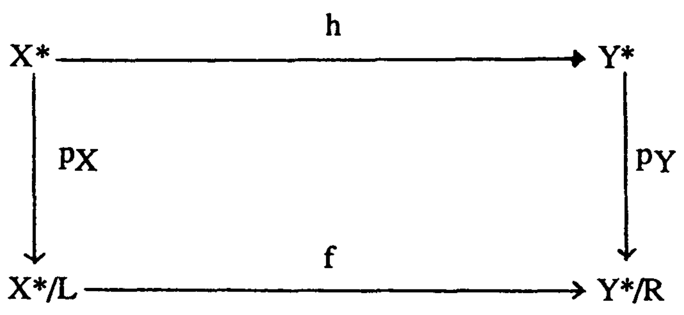
Define a grammar $G' = (V \cup Q, X, P \cup \{q \to q_1 \mid q_2;\ q \in Q,\ q = (q_1, q_2)\}, S)$.

Then $L(G') = L(G)$ and $M(G')$ is canonically X-isomorphic to $(V \cup X)^*/P \cup Q$ and hence X-isomorphic to Y^*/R. □

COROLLARY 9: *Let G_1, G_2 be grammars over X such that $L(G_1) \subseteq L(G_2)$. Then we effectively find a grammar which generates $L(G_1)$ and which has a Hotz monoid X-isomorphic to $M(G_2)$.*

Proof: Since $L(G_1) = L(G_1) \cap L(G_2)$ we find, by Lemma 7, a grammar which generates $L(G_1)$ and whose Hotz monoid is X-isomorphic to the direct product $M(G_1) \times M(G_2)$. Since $M(G_2)$ is a quotient of $M(G_1) \times M(G_2)$ the result follows by Theorem 8. □

Proof of Theorem 1: Let $L \subseteq X^*$ be a language generated by some grammar and $f: X^*/L \to M$ be a homomorphism to a finitely presentable monoid M. Write $M = Y^*/R$ with Y and R finite. Let $h: X^* \to Y^*$ be any homomorphism such that the following diagram with the canonical projections p_X and p_Y commutes:



Let $v \in X^*$ be any word such that $f p_X(L) \subseteq \{p_Y h(v)\}$. (Note, if $L = \emptyset$ then we can take any word and if $L \neq \emptyset$ then we can take any word $v \in L$. If L is known to be non-empty then such a word can be found effectively.)

Define a grammar G_1 by $G_1 = (\{S\}, Y, \{S \to h(v)\} \cup R \cup R^{-1}, S)$. Obviously, $M(G_1)$ is canonically X-isomorphic to Y^*/R, further it holds $p_Y^{-1}(f(p_X(L))) \subseteq L(G_1)$.

By Theorem 4 there is another grammar G_2 which generates $L_2 = h^{-1}(L(G_1))$ such that $M(G_2)$ is X-isomorphic to $M(G_1)$. Hence, $M(G_2)$ is X-isomorphic to Y^*/R. Since $L \subseteq p_X^{-1}(f^{-1}(f(p_X(L))) = h^{-1}(p_Y^{-1}(f(p_X(L))) \subseteq L_2$ we find, by Corollary 9, a grammar G with $L(G) = L$ such that $M(G)$ is X-isomorphic to $M(G_2)$. This is the final step in the proof of Theorem 1. $\square$

I.3. Hotz-isomorphisms on cancellative level

Corollaries 2 and 3 gave purely intrinsic characterizations for languages with a Hotz-isomorphism on monoid level and on group level. For the cancellative level we have the following result.

THEOREM 10: *A language $L \subseteq X^*$ is a language with a Hotz isomorphism on cancellative level if and only if the canonical mapping $X^*/L \to C(X^*/L)$ factorizes $X^*/L \to M \to C(X^*/L)$ where M is a finitely presentable monoid.*

Proof: If G is a grammar for L with a Hotz-isomorphism on cancellative level then we can take $M = M(G)$. Conversely, let $X^*/L \to M \to C(X^*/L)$ be any factorization where M is a finitely presentable monoid. We can write $M = Y^*/R$ where Y and R are finite. For each $y \in Y$ there is a word $v_y \in X^*$ such that y and v_y have the same image in $C(X^*/L)$.

Define $M' = Y^*/(R \cup \{(y, v_y) \mid y \in Y\})$ then M' is finitely presented and we obtain a factorization $X^*/L \overset{f}{\twoheadrightarrow} M' \overset{g}{\twoheadrightarrow} C(X^*/L)$ where f and g are surjective. Now, $C(M')$ is X-isomorphic to $C(X^*/L)$ and the result follows by Theorem 1. $\square$

Another characterization of languages with a Hotz-isomorphism on cancellative level can be given in terms of homomorphic images of sentential form languages. This class of languages is denoted by $\hat{H}IS$ and it was studied for example in Jantzen/Kudlek [7].

THEOREM 11: *Let $L \subseteq X^*$ be a language and $p_c: X^* \to C(X^*/L)$ be the canonical projection. Then L is a language with a Hotz-isomorphism on* *cancellative level if and only if there is a language* $\hat{L} \in \hat{H}IS$ *such that* $L \subseteq \hat{L} \subseteq p_c^{-1} p_c(L)$.

Proof: If $L \subseteq \hat{L} \subseteq p_c^{-1} p_c(L)$ with $\hat{L} \in \hat{H}IS$ then

$$C(X^*/L) = C(X^*/\hat{L})$$

and $\hat{L}$ has a Hotz-isomorphism on cancellative level by [2], Thms. 3,4. Hence, L has a Hotz-isomorphism by Corollary 9. For the other direction we may assume $L \neq \varnothing$. Let $G = (V, X, P, S)$ be a grammar for L with a Hotz-isomorphism on cancellative level $i_c: C(X^*/L) \xrightarrow{\sim} C(M(G))$. For each $A \in V$ choose $v_A \in X^*$ such that $i_c(v_A) \equiv A$ in $C(M(G))$. Consider $\hat{G} = (V, X, P \cup \{A \to v_A \mid A \in V\}, S)$. Then $\hat{G}$ is a so-called terminating grammar and $L(\hat{G}) \in \hat{H}IS$ by [2], Lemma 4. It holds $L \subseteq L(\hat{G}) \subseteq p_c^{-1} p_c(L)$, since $L \neq \varnothing$. □

Open Questions: If $L \subseteq X^*$ is a language with a Hotz-isomorphism on cancellative level then $C(X^*/L)$ is the cancellative quotient of a finitely presentable monoid. Does the converse hold? A positive answer would yield a nicer characterization of these languages than Theorem 10.

Another open question is the following: Let $L \subseteq X^*$ be a language with a Hotz isomorphism on group level. Is L such a language on cancellative level? The answer to the last question is "yes", provided $L^i \cap L^j \neq \varnothing$ for some $0 \leqq i < j$. [Since then $C(\text{alph}(L)^*/L)$ is a group]. In particular, this holds if $\lambda \in L$.

I.4. Context-sensitive grammars

In this paragraph we are concerned with the type of the grammar which can give a Hotz-isomorphism. If $L \subseteq X^*$ is a context-free language then any reduced context free grammar for L has a Hotz-isomorphism on cancellative level, *see* [4], Prop. 3. The situation for context-sensitive languages is completely different. Firstly, there are languages such as $\{a^n b^n c^n d^n \mid n \geqq 0\}$ without a Hotz-isomorphism on any level, *see* [3], Rem. 2.5. Secondly, we shall see that there are context-sensitive languages with a Hotz-isomorphism on monoid level, but they do not have any generating context-sensitive grammar (with erasing) with a Hotz-isomorphism on group level. Such an example is given by the language $\{a^n b^n c^n \mid n \geqq 0\} \cup \{cba\}$, *see* Example 13 (i) below.

THEOREM 12: *Let H be a finitely presentable group, $h \in H$ be a group element, $f: X^* \to H$ be a homomorphisms, $R \subseteq X^*$ be a regular subset, and $L = R \cap f^{-1}(h)$.*

If the group H is X-isomorphic to $F(X)/L$ and L is not context-free then L is a language with a Hotz-isomorphism on group level but there is no context-sensitive grammar (with erasing) for L with a Hotz-isomorphism.

Proof: Clearly, L is recursively enumerable; by Corollary 3 L is a language with a Hotz-isomorphism. Assume $G = (V, X, P, S)$ would be a grammar for L with a Hotz-isomorphism where all rules in P have the following form (context sensitive with erasing):

$$\gamma A \delta \to \gamma\beta\delta,\ A \in V,\ \gamma, \beta, \delta \in (V \cup X)^*.$$

Replace all these rules by the corresponding context-free productions $A \to \beta$. We obtain a context-free grammar $\hat{G}$ with $H(G) = H(\hat{G})$ and $L \subseteq L(\hat{G})$. Since f factorizes $X^* \to X^*/L(\hat{G}) \to H(\hat{G}) \to H$, we have $L \subseteq L(\hat{G}) \subseteq f^{-1}(h)$. Therefore $L = R \cap L \subseteq R \cap L(\hat{G}) \subseteq R \cap f^{-1}(h) = L$; hence $L = R \cap L(\hat{G})$ is context-free, contradicting the assumption. □

Example 13: (*i*) Let $X = \{a, b, c\}$, $f: X^* \to \mathbf{Z} \times \mathbf{Z}$ defined by $f(a) = (1, 0)$, $f(b) = (0, 1)$, $f(c) = (-1, -1)$, and $L = (a^* b^* c^* \cup \{cba\}) \cap f^{-1}(0, 0)$. Then we have $L = \{a^n b^n c^n \mid n \geqq 0\} \cup \{cba\}$ is context-sensitive but not context-free. Since $X^*/L = C(X^*/L) = F(X)/L = \mathbf{Z} \times \mathbf{Z}$, the language L has a Hotz isomorphism on monoid level but by Theorem 12, even on group level, no Hotz-isomorphism is given by a context-sensitive grammar with erasing for L. (Instead of $\mathbf{Z} \times \mathbf{Z}$ we could use any finitely presented quotient group of X^* which is context-sensitive but not context-free.)

(ii) Let $X = \{a, b, c\}$, $H = F(X)/\{(aca, c), (bcb, c), (ab, ba)\}$, $f: X^* \to H$ the canonical mapping and

$$L = f^{-1} f(c) = \{w_1 c w_2;\ w_1, w_2 \in \{a, b\}^*,\ |w_1|_x = |w_2|_x,\ x \in \{a, b\}\}$$

where $|w|_x$ means the number of occurrences of a letter x in a word w. Then Theorem 12 applies to the context-sensitive language L. □

We do not know whether a context-sensitive language $L \subseteq X^*$ has a context-sensitive grammar with a Hotz-isomorphism if $F(X)/L$ is a context-free group [i. e., if $F(X)/L$ has a free subgroup of finite index]. But we have a positive result for $F(X)/L$ finite.

THEOREM 14: *Let $L \subseteq X^*$ be a language such that $F(X)/L$ is a finite group. Then there is a context-sensitive grammar with erasing (and without erasing if L is context-sensitive), for L with a Hotz-isomorphism on group level.*

Proof: Without restriction we have $X \neq \emptyset$. Let $H = F(X)/L$. Since H is a finite group, we have $L \neq \emptyset$. Let $q: X^* \to H$ the canonical morphism onto H and $h_0 = q(w) \in H$ for some $w \in L$.

Let $G = (V, X, P, S)$ be a grammar for L. It is well-known that we can assume all productions to have the following form:

$$\gamma A \delta \to \gamma \beta \delta, \qquad A \in V, \quad \gamma, \beta, \delta \in (V \cup X)^*.$$

We extend $q: X^* \to H$ to a homomorphism $q: ((V \times H) \cup X)^* \to H$ by $q(A, h) = h$ for $(A, h) \in V \times H$. We also define a homomorphism $p: ((V \times H) \cup X)^* \to (V \cup X)^*$ by $p(A, h) = A$ and $p(x) = x$ for $(A, h) \in V \times H$, $x \in X$. With the help of these mappings we define the productions of a new grammar $G' = (V \times H, X, P', (S, h_0))$ as follows:
Put

$$\gamma' A' \delta' \to \gamma' \beta' \delta' \in P' \text{ for } A' \in V \times H, \gamma', \beta', \delta' \in ((V \times H) \cup X)^*$$

if and only if:

$$p(\gamma') p(A') p(\delta') \to p(\gamma') p(\beta') p(\delta') \in P \qquad \text{and} \qquad q(A') = q(\beta').$$

It is easily proved by induction on the length of a derivation that $(S, h_0) \overset{*}{\underset{G'}{\Rightarrow}} w' \in ((V \times H) \cup X)^*$ if and only if $S \overset{*}{\underset{G}{\Rightarrow}} p(w')$ and $q(w') = h_0$. Hence, $L(G) = L$.

It is seen from the form of the productions that q induces a homomorphism $\bar{q}: H(G') \to H$ such that the composition $H = F(X)/L \overset{\text{can.}}{\to} H(G') \overset{\bar{q}}{\to} H$ is the identity. Therefore, the mapping $i_H: F(X)/L \to H(G')$ is injective. As shown in [2], Thm. 8 we can effectively find an equivalent subgrammar of G' with a Hotz group X-isomorphic to $F(X)/L$. □

II. UNDECIDABILITY RESULTS

The following results hold for Hotz-isomorphism on any level. The proofs are given for the monoid level, only. The other levels are handled completely analogously. The property of languages whether they are languages with a

Hotz-isomorphism is clearly non-trivial; there are languages with a Hotz-isomorphism, e. g. the empty language, and context-sensitive languages without any Hotz-isomorphism, such as $L=\{a^n b^n c^n d^n \mid n \geqq 0\}$. Thus, by the Theorem of Rice, *see* e. g. [6], Th. 8.6, it is undecidable whether a grammar generates a language with a Hotz-isomorphism. Similarly, it is undecidable whether the grammar itself allows a Hotz-isomorphism. To see this, let $G=(V, X, P, S)$ be any grammar with $X \neq \varnothing$. Let Z be a new symbol and set $G'=(V \cup \{Z\}, X, P \cup \{Z \to A \mid x \mid \lambda;\ A \in V,\ x \in X\}, S)$.

We have $L(G)=L(G')$ and $M(G')=\{1\}$. The grammar G' has a Hotz-isomorphism if and only if $X^*/L(G)$ is the trivial monoid. This is undecidable, again by the Theorem of Rice. Thus, both questions whether a grammar generates a language with a Hotz-isomorphism and whether a grammar has a Hotz-isomorphism are undecidable. However, the second question is semi-decidable whereas the first one is not.

THEOREM 15: *The set of grammars with a Hotz-isomorphism is recursively enumerable.*

Proof: Let $G=(V, X, P, S)$ be a grammar. First notice that it is semi-decidable whether the canonical mapping $i: X^*/L(G) \to M(G)$ is surjective. Indeed, the answer to this is given by the value of the following semi-decidable predicate:

$$\forall A \in V, \quad \exists n \geqq 1, \quad \exists \alpha_i \in (V \cup X)^*,$$
$$i=1, \ldots, n-1, \quad \exists\, \alpha_n \in X^*: A \underset{G}{\Leftrightarrow} \alpha_1 \underset{G}{\Leftrightarrow} \cdots \underset{G}{\Leftrightarrow} \alpha_n.$$

Let us show that it is also semi-decidable whether the morphism $i: X^*/L(G) \to M(G)$ has a left-inverse. For a function $f: V \to X^*$ let $f_x: (V \cup X)^* \to X^*/L(G)$ which is induced by $f(A)$ for $A \in V$ and by the identity on X. Now, the morphism i has a left-inverse if and only if there is a function $f: V \to X^*$ such that $f_x(l)=f_x(r)$ in $X^*/L(G)$ for all productions $(l, r) \in P$. Since the defining relations for $X^*/L(G)$ are recursively enumerable and P is finite, the existence of such a function f is semi-decidable.

The Theorem follows because the conjunction of two semi-decidable questions is semi-decidable again. □

PROPOSITION 16: *Neither the set of grammars generating a language with a Hotz-isomorphism nor its complement is recursively enumerable.*

Proof: This is an easy consequence of the well-known containment property of recursively enumerable sets of languages, *see* [6], Lemma 8.2. □

We finish the paper by showing two other undecidability results.

THEOREM 17: *Let $X \neq \emptyset$. There is no algorithm of the following type:*

It takes a grammar G over X as input and if $L(G)$ is a non-empty language with Hotz-isomorphism then it produces an output grammar G' such that $L(G)=L(G')$ and $H(G')$ is X-isomorphic to $F(X)/L(G)$, otherwise the algorithm may produce any garbage or it may even fail to terminate.

In fact, we shall prove a stronger result, namely there is no such algorithm which is required to work correctly only on grammars producing finite non-empty languages $L \subseteqq X^*$ where X^*/L is trivial or the group with two elements $\mathbf{Z}/2\,\mathbf{Z}$.

Proof: Let G be any grammar with terminal alphabet X. Choose any $x \in X$. Then we effectively find a grammar G_1 such that

$$L(G_1)=(L(G) \cap \{x\}) \cup (X \setminus \{x\}) \cup \{\lambda, x^2\}.$$

Of course, $F(X)/L(G_1)=\{1\}$ if $x \in L(G)$ and $F(X)/L(G_1) \cong \mathbf{Z}/2\,\mathbf{Z}$ otherwise. Assume there would be such an algorithm as above. Since $L(G_1)$ is finite and non-empty, we would obtain a grammar G_1' with $H(G_1')=\{1\}$ if $x \in L(G)$ and $H(G_1') \cong \mathbf{Z}/2\,\mathbf{Z}$ otherwise.

Since $H(G_1')$ is an effectively given finite presentation of a finite abelian group, we can decide whether this group is trivial or not. Thus, we could decide whether $x \in L(G)$ or not. □

THEOREM 18: *Let $X \neq \emptyset$. There is no algorithm of the following type:*

It takes a grammar G over X as input and if $F(X)/L(G)=F(G)=F(X)$ (i. e., $L(G)$ contains at most one element) then it produces an output grammar G' such that $L(G)=L(G')$ and $H(G')$ is X-isomorphic to $F(X)$, otherwise it may produce any garbage or it may even fail to terminate.

Proof: Choose any two different words $v_1, v_2 \in X^*$, and grammars G_1, G_2 such that $L(G_1)=\{v_1\}$ and $L(G_2)=\{v_2\}$.

Now, let G be any grammar over X. We effectively find a grammar G_{12} such that $L(G_{12})=L(G) \cap \{v_1, v_2\}$. Assume there would be an algorithm as above. Then the following happens: If $L(G_{12}) \neq \{v_1, v_2\}$ then $F(X)/L(G_{12})=F(X)$ and we find a grammar G_{12}' such that $H(G_{12}')$ is X-isomorphic to $F(X)$. Knowing its existence, we effectively find this X-isomorphism; and we can compute the image of the axiom in $F(X)$. If this image is not v_1 then $L(G) \neq \{v_1\}$, if it is v_1 then $L(G) \neq \{v_2\}$.

If $\{v_1, v_2\} \subseteqq L(G)$ this algorithm may produce any garbage or it may fail to terminate. However, since the question $\{v_1, v_2\} \subseteqq L(G)$ is semi-decidable, we could find another algorithm which always terminates with values 1 or 2 on input G such that the output 1 implies $L(G) \neq \{v_1\}$ and the output 2 implies $L(G) \neq \{v_2\}$. Let $f(G)=1$ or 2 according to the output of this algorithm. Then $G \to G_{f(G)}$ is a computable function. By the well-known recursion theorem, *see* [6], Thm. 8.18, we have $L(G)=L(G_{f(G)})$ for some grammar G. But by the definition of the grammars G_1 and G_2 this is impossible. □

Remark: For the sake of completeness, let us mention that in the exotic case $X=\emptyset$ there exist such algorithms as demanded in Theorems 17 or 18. (Note, however, that the emptiness problem is still undecidable for grammars over an empty set of terminals.) Slightly more general, if G is given and we know any word $w \in X^*$ such that $L(G) \subseteq \{w\}$ [but we may ignore whether $L(G)$ is empty or not], then we effectively find an equivalent grammar G' which has a Hotz-isomorphism. This may be seen by the proof to Theorem 1.

Note added in proof: The problem state preceding Theorem 14, has a positive answer: this theorem can be generalized to context-free $F(X)/L$. For context-sensitive grammars (with or without erasing) G_1, G_2 there exists a context-sensitive grammar (with or without erasing) G, such that $L(G)=L(G_1) \cap L(G_2)$ and $C(M(G)) \cong C(M(G_2))$. [However $C(M(G)) \cong C(M(G_1)) \times C(M(G_2))$ cannot be achieved in general. This "context-sensitive version of Lemma 7" is wrong even on group level.]

We conclude a context-sensitive version of Corollary 9 on cancellative level, and from this we obtain Theorem 14 for context-free groups.

REFERENCES

1. A. Clifford and G. Preston, *The Algebraic Theory of Semigroups*, Amer. Math. Soc., Vol. I, 1961; Vol. II, 1967.
2. V. Diekert, *Investigations on Hotz Groups for Arbitrary grammars*, Acta Informatica, Vol. 22, 1986, pp. 679-698.
3. V. Diekert, *On some variants of the Ehrenfeucht Conjecture*, Theoret. Comp. Sci., Vol. 46, 1986, pp. 313-318.
4. C. Frougny, J. Sakarovitch and E. Valkema, *On the Hotz Group of a Context-free Grammar*, Acta Informatica, Vol. 18, 1982, pp. 109-115.
5. G. Hotz, *Eine neue Invariante für Kontext-freie Sprachen*, Theoret. Comp. Sci, Vol. 11, 1980, pp. 107-116.
6. J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory*, Languages and Computation, Addison-Wesley, Massachusetts, 1979.
7. M. Jantzen and M. Kudlek, *Homomorphic Images of Sentential Form Languages Defined by Semi-Thue System*, Theoret. Comp. Sci., Vol. 33, 1984, pp. 13-43.
8. A. Möbus, *On Languages with a Hotz-isomorphism*, Manuscript, 1986.