Some Decidability Results about Regular and Push Down Translations

K. Culik II
Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada

Research Report CS-78-09

January 1978

Some Decidability Results about

Regular and Push Down Translations

K. Culik II

Introduction

We will show a number of problems about finite transducers and push down transducers to be decidable. Our main tool is a result established in [3], namely that for a given context free language and two homomorphisms h_1, h_2, it is decidable whether h_1 and h_2 are string by string equal on L, i.e. whether $h_1(w) = h_2(w)$ for all $w \in L$. We show some generalizations of this result but mainly we are concerned with decision problems about translations.

The equivalence problem for (arbitrary) finite transducers, even for nondeterministic gsm machines, is known to be undecidable; the same problem for deterministic gsm machines is known to be decidable. We show that it is decidable whether a given (arbitrary) finite transducer defines a functional translation and that the equivalence problem for such transducers is decidable. This remains so even when we restrict the comparison to the inputs from a given context free language. We show that the equivalence between an unambiguous push down tranducer and a functional finite transducer is decidable. Another of the problems shown to be decidable is the identity problem for (arbitrary) finite transducers, i.e. the problem of deciding whether a given regular translation is an identity on its domain.

We assume that the reader is familiar with the fundamental theory of formal languages, specifically with the notions of finite and push-down tranducers as defined in [1]. We will use the notation of [1] unless stated otherwise. The translation defined by transducer M is denoted by t_M. The translations defined by finite and push-down transducers are called regular (rational) and push-down translations, respectively. The family of these translationa are denoted by RT and PDT, respectively. For translations $t_1 \subseteq \Sigma^* \times \Gamma^*$ and $t_2 \subseteq \Gamma^* \times \Delta^*$, the composition of t_1 and t_2 is denoted by $t_1 \circ t_2$ and defined as

$$t_1 \circ t_2 = \{(x, z) : (x, y) \in t_1 \text{ and } (y, z) \in t_2 \text{ for some } y \in \Gamma^*\} .$$

A translation $t \subseteq \Delta_1^* \times \Delta_2^*$ is said to be homomorphically characterized by a language L if there are two homomorphisms $h_i : \Sigma^* \to \Delta_i^*$, $i = 1,2$, so that $t = \{(h_1(w), h_2(w)) : w \in L\}$. The family of translations homomorphically characterized by languages from family $\mathcal{L}$ is denoted by $H(\mathcal{L})$. The families of regular and context free languages are denoted by REG and CFL, respectively. The following two results are well-known, c.f. [1].

<u>Theorem 1</u> RT = H(REG) .

<u>Theorem 2</u> PDT = H(CFL) .

<u>Definition</u> A translation $f \subseteq \Sigma^* \times \Delta^*$ is said to be <u>functional</u> if $(x, y) \in f$ and $(x, z) \in f$ implies $y = z$. Then we write $y = f(x)$.

In [3] the notions of equivalence and ultimate equivalence of homomorphisms and gsm mappings on a language were introduced. We will consider arbitrary functional translations.

Definition Let L be a language over alphabet Σ and f_1, f_2 be functional translations from Σ^* to Δ^*. We say that

(i) f_1 and f_2 are equivalent on L iff $f_1(w) = f_2(w)$ for all w in Σ^*.

(ii) f_1 and f_2 are ultimately equivalent on L if there is only a finite number of w in L such that $f_1(w) \neq f_2(w)$.

For a family of effectively specified languages $\mathcal{L}$ and a family of effectively specified functional translations $\mathcal{T}$ we have two decision problems. We speak about the problem of $\mathcal{T}$ (ultimate) equivalence for $\mathcal{L}$. An instance of such a problem is given by two translations f_1, f_2 in $\mathcal{T}$ and a language L in $\mathcal{L}$, and asks the question: are f_1 and f_2 (ultimately) equivalent on L ?

The problem of homomorphism equivalence for regular languages was shown to be decidable in [2] and in [3] the following stronger result was obtained:

Theorem 3 The problem of homomorphism (ultimate) equivalence for CFL is decidable.

We will show that the problem of homomorphism (ultimate) equivalence for $\mathcal{L}$ is equivalent to the identity problem for the family of translations $H(\mathcal{L})$.

Definition We say that a translation $t \subseteq \Sigma^* \times \Sigma^*$ is a restriction of the identity translation iff there is a language $L \subseteq \Sigma^*$ so that $t = \{(w, w) : w \in L\}$.

For a family of effectively specified translations we have a decision problem, namely whether a given translation is a restriction of the identity. We speak about the ultimate identity problem if we check whether a given translation is a restriction of identity up to a finite number of exceptions.

Theorem 4 Let $\mathcal{L}$ be a family of languages. The (ultimate) identity problem for the family of translations $H(\mathcal{L})$ is decidable iff the problem of homomorphism (ultimate) equivalence for $\mathcal{L}$ is decidable.

Proof Given t in $H(\mathcal{L})$ we can find L in $\mathcal{L}$ and homomorphisms h_1, h_2 so that $t = \{(h_1(w), h_2(w) : w \in L\}$. Translation t is (ultimately) a restriction of identity iff h_1 and h_2 are (ultimately) equivalent on L.

□

Corollary 1 The (ultimate) identity problem for families RT and PDT is decidable.

Proof By Theorems 1, 2 and 4.

□

We will consider decision problems about functional regular translations. First, we show that the property of functionality for finite transducers is decidable.

Theorem 5 Given a finite tranducer M, it is decidable whether the translation t_M is functional.

Proof Consider translation $t = t_M^{-1} \circ t_M$. Clearly, t is a restriction of identity iff t_M is functional. Since the family RT is effectively closed under composition (see [1]) we can check the functionality of t_M by Corollary 1.

□

Note that the proof of Theorem 5 cannot be extended to PDT since they are not closed under composition. Actually, from the undecidability of the emptiness of intersection of two context free languages (see [4]) it follows immediately that the functionality for PDT is undecidable.

Corollary 2 Given a finite transducer M it is decidable whether the translation t_M is one-to-one.

Proof By Theorem 5 and the well known fact that regular translations are effectively closed under inversion.

□

Lemma 1 Assume that L is a family of languages with the following properties:

(i) L is effectively closed under functional regular translations;

(ii) the emptiness problem for L is decidable;

(iii) the problem of homomorphism (ultimate) equivalence for L is decidable.

Then the problem of functional regular translations (ultimate) equivalence for L is decidable.

Proof In the proof of Theorem 3.4 in [3] only the functionality of deterministic gsm mappings was used. Hence, the same proof works for our lemma. The modification for ultimate equivalence is obvious.

□

Theorem 6 The problem of functional regular translations (ultimate) equivalence for CFL is decidable.

Proof By Theorem 3 and Lemma 1.

□

Definition A push-down transducer is said to be unambiguous if for each input string there is at most one computation. A push-down translation is unambiguous if it is defined by a unambiguous push-down transducer.

Note, that a push-down transducer is unambiguous iff its underlying push-down automaton (obtained by ignoring outputs) is unambiguous [4] and t is semantically unambiguous in terminology of [1]. Functionality for push-down transducers is not an effective property since unambiguity for push-down automata (context-free grammars) is undecidable [4]. However, each deterministic push-down transducer is unambiguous, so our results also hold for this effective subclass of PDT. Note that deterministic PDT are properly included in unambiguous PDT which in turn are properly included in functional PDT.

<u>Theorem 7</u> Given an unambiguous push-down transducer P and a functional finite transducer M it is decidable whether $t_P = t_M$.

<u>Proof</u> We first check whether $\mathrm{dom}\ t_P = \mathrm{dom}\ t_M$. This is decidable by [5, Theorem IV 5.5] since $\mathrm{dom}\ t_p$ is an unambiguous CFL and $\mathrm{dom}\ t_M$ a regular set (both effectively given). If the domains match we continue as follows. We consider the translation $t = t_M^{-1} \circ t_P$. Since the inverse of a regular translation is again in RT [1], and since the family PDT is closed under composition (from either side) with regular translation, we have $t \in \mathrm{PDT}$ (effectively). Clearly, $t_M(w) = t_P(w)$ for each w in the common domain iff t is a restriction of identity. Hence, the proof is completed by Corollary 1.

□

Note that we can modify the proof of Theorem 7 to show that the "ultimate equivalence problem" (agreement up to a finite number of exceptions) between an unambiguous push-down transducer and a functional finite transducer is decidable.

Now, we generalize the notion of the (ultimate) equivalence of functional translations on a language.

<u>Definition</u> We say that two functional translations $t_1 : \Sigma^* \to \Gamma^*$ and $t_2 : \Delta^* \to \Gamma^*$ are <u>equivalent modulo relation</u> (translation) $R \subseteq \Sigma^* \times \Delta^*$, written $t_1 \equiv t_2 \bmod R$, iff $t_1(x) = t_2(y)$, for all $(x, y) \in R$. For two families of translations T_1 and T_2 we have the problem of T_1 (ultimate) equivalence for T_2.

Note, that t_1 and t_2 are equivalent on L iff $t_1 \equiv t_2 \bmod \{(w, w) : w \in L\}$.

Lemma 2 Let $\mathcal{T}$ be a family of functional translations effectively closed under composition (from left) with homomorphisms, and $\mathcal{L}$ be a family of languages. The problem of $\mathcal{T}$ (ultimate) equivalence for $H(\mathcal{L})$ is decidable iff the problem of $\mathcal{T}$ (ultimate) equivalence for $\mathcal{L}$ is decidable.

Proof

(1) Consider t_1, t_2 in $\mathcal{T}$ and $L \in \mathcal{L}$. Clearly, t_1 and t_2 are (ultimately) equivalent on L iff $t_1 \equiv t_2 \bmod \{(w, w) : w \in L\}$.

(2) Consider t_1, t_2 in $\mathcal{T}$ and $R \in H(\mathcal{L})$. By the definition of $H(\mathcal{L})$, there is $L \in \mathcal{L}$ and homomorphisms h_1, h_2 so that $R = \{(h_1(w), h_2(w)) : w \in L\}$. Let $t_i' = h_i \circ t_i$ for $i = 1,2$. By our assumptions t_1' and t_2' are in $\mathcal{T}$ (effectively). Clearly, $t_1 \equiv t_2 \bmod R$ iff t_1' and t_2' are equivalent on L .

□

Theorem 8 The problem of functional regular translations (ultimate) equivalence for PDT is decidable.

Proof By Lemma 2 and Theorem 6.

□

Since homomorphism is a special case of a functional regular translation we have:

Corollary 3 The problem of homomorphism (ultimate) equivalence for PDT is decidable.

References

[1] A.V. Aho and J.D. Ullman, The Theory of Parsing, Translation, and Compiling, Vol. I: Parsing, Prentice-Hall, 1972.

[2] K. Culik II and I. Fris, "The decidability of the equivalence problem for DOL-systems", Information and Control 35 (1977), 20-39.

[3] K. Culik and A. Salomaa, "On the decidability of homomorphism equivalence for languages", J. Comput. System Sci., to appear.

[4] J.E. Hopcroft and J.D. Ullman, Formal Languages and Their Relations to Automata, Addison Wesley, 1969.

[5] A. Salomaa and M. Soittola, Automata-Theoretic Aspects of Formal Power Series, to appear.