Reverse Twin Shuffles *

Joost Engelfriet
Department of Computer Science, Leiden University
P.O.Box 9512, 2300 RA Leiden, The Netherlands
e-mail: engelfri@wi.leidenuniv.nl

It is shown in [ER] that every recursively enumerable language is the image of the twin shuffle language under a dgsm mapping (where a dgsm is a deterministic generalized sequential machine with accepting states). The twin shuffle language is $L_{\mathrm{TS}} = \{w \in \{0, 1, \overline{0}, \overline{1}\}^* \mid \overline{\pi}(w) = \pi(w)\}$, where $\pi, \overline{\pi} : \{0, 1, \overline{0}, \overline{1}\}^* \to \{0, 1\}^*$ are the homomorphisms with $\pi(0) = 0$, $\pi(1) = 1$, $\pi(\overline{0}) = \pi(\overline{1}) = \lambda$ (the empty string), $\overline{\pi}(\overline{0}) = 0$, $\overline{\pi}(\overline{1}) = 1$, $\overline{\pi}(0) = \overline{\pi}(1) = \lambda$. In other words, L_{TS} consists of all strings that are obtained by shuffling x with its "twin" $\overline{x}$, for every $x \in \{0, 1\}^*$. If w is such a shuffle, then $\pi(w) = \overline{\pi}(w) = x$. As an example, $\overline{0}\overline{1}010\overline{0}\overline{0}\overline{0}$ is in L_{TS}.

A natural question is whether or not the same result also holds for "reverse twins" x and $\overline{x}^R$ (where y^R is the usual mirror image of a string y). The answer is yes. This means that, from the point of view of generating the recursively enumerable languages, the direction of the twins is not important, which supports the robustness of the twin shuffle concept. This result is in fact shown in Theorem 4.5 of [Bra], using nondeterministic finite state transductions (or, rational transductions) rather than dgsm mappings. Thus, the following theorem slightly strengthens the result of [Bra].

Let $L_{\mathrm{RTS}} = \{w \in \{0, 1, \overline{0}, \overline{1}\}^* \mid \overline{\pi}(w) = \pi(w)^R\}$ be the *reverse twin shuffle language*. For instance, $\overline{0}\overline{0}010\overline{1}0\overline{0}$ is in L_{RTS}. In [Bra], L_{RTS} is denoted $\mathrm{Merge}_{1,1}(\{0, 1\}^*)$.

Theorem. For every recursively enumerable language L there is a dgsm mapping γ such that $\gamma(L_{\mathrm{RTS}}) = L$.

Proof. Since dgsm mappings are closed under composition, it suffices to show the existence of a dgsm mapping γ such that $\gamma(L_{\mathrm{RTS}}) = L_{\mathrm{TS}}$. It is impossible for a dgsm to reverse the direction of one of the twins. The idea

*This is a revised version of a note on page 144 of EATCS Bulletin 60 of the proof is to let the dgsm select twins x that are palindromes, and in particular of the form $x = y\#y^R$, with $y \in \{0,1\}^*$, and 0, 1, # coded as 01, 10, 00, respectively.

Let γ be the mapping that has the regular language D as domain, with

$$D = \{01, 10, \overline{01}, \overline{10}\}^* 00\overline{00}\{01\overline{01}, 10\overline{10}\}^*,$$

and such that $\gamma(h(u)00\overline{00}v) = u$ for all strings $u, v \in \{0, 1, \overline{0}, \overline{1}\}^*$, where h is the homomorphism with $h(0) = 01$, $h(1) = 10$, $h(\overline{0}) = \overline{01}$, and $h(\overline{1}) = \overline{10}$.

It is easy to see that γ is a dgsm mapping. Thus it remains to show that $\gamma(L_{\mathrm{RTS}}) = L_{\mathrm{TS}}$. Consider a string $w = h(u)00\overline{00}v$ in $L_{\mathrm{RTS}} \cap D$. Since $\overline{\pi}(w) = \pi(w)^R$, the specific form of D (with the "marker" $00\overline{00}$ in the middle) implies that $\overline{\pi}(h(u)) = \pi(v)^R$ and $\overline{\pi}(v) = \pi(h(u))^R$. Since $\overline{\pi}(v) = \pi(v)$ (because $v \in \{01\overline{01}, 10\overline{10}\}^*$), this implies that $\overline{\pi}(h(u)) = \pi(h(u))$. Hence $h(\overline{\pi}(u)) = h(\pi(u))$ and so, since h is injective, $\overline{\pi}(u) = \pi(u)$, i.e., $\gamma(w) = u \in L_{\mathrm{TS}}$. This shows that $\gamma(L_{\mathrm{RTS}}) \subseteq L_{\mathrm{TS}}$. In the other direction, if u is a string in $\{0, 1, \overline{0}, \overline{1}\}$ such that $\overline{\pi}(u) = \pi(u)$, consider the string $w = h(u)00\overline{00}v$ where v is the (unique) element of $\{01\overline{01}, 10\overline{10}\}^*$ such that $\pi(v) = \pi(h(u))^R$. Then $w \in D$, $\gamma(w) = u$, and $w \in L_{\mathrm{RTS}}$ because $\overline{\pi}(w) = \overline{\pi}(h(u))00\overline{\pi}(v) = \pi(h(u))00\pi(v) = (\pi(h(u))00\pi(v))^R = \pi(w)^R$. This ends the proof.

For which mappings $f : \{0,1\}^* \to \{0,1\}^*$ is the "f-twin shuffle language" $L_f = \{w \in \{0, 1, \overline{0}, \overline{1}\}^* \mid \overline{\pi}(w) = f(\pi(w))\}$ a generator of the recursively enumerable languages?

References

[Bra] F.-J.Brandenburg; Representations of language families by homomorphic equality operations and generalized equality sets, Theoret. Comp. Sci. 55 (1987), 183-263

[ER] J.Engelfriet, G.Rozenberg; Fixed point languages, equality languages, and representation of recursively enumerable languages, J. of the ACM 27 (1980), 499-518