Master Parisien de Recherche en Informatique

Course 2.16 – Finite automata based computation models

26 november 2019 — Exam (1) — Part B

Books and computers forbidden — Lecture and personal notes allowed.
This part should be written on separate test papers.

The two exercices are independent and can be done in any order.

Remark: the original version had several minor errors, in **3.**, **2.** and **9.**, that are fixed in this version. This changed $\cup\{\varepsilon\}$ into $\setminus\{\varepsilon\}$ everywhere, although almost no student noticed that there was a subtlety about the empty word.

1 Probabilistic automata

1. Show that if L is a regular language then there exists a probabilistic automaton $\mathcal{A}$ such that for every word w, $\mathbb{P}_{\mathcal{A}}(w)$ is 1 if $w \in L$ and 0 otherwise.

2. Show that if L is stochastic language and L' is a regular language then $L \cap L'$ and $L \setminus L'$ are stochastic.

Recall that L is λ-stochastic if there a probabilistic automaton $\mathcal{A}$ such that $L = \mathcal{L}_{\mathcal{A}}(\lambda) = \{w : \mathbb{P}_{\mathcal{A}}(w) > \lambda\}$. The goal is to show that one can change the cut-point arbitrarily:

Lemma 1. *If L is a stochastic language and $\lambda \in (0,1)$, then L is $\lambda-$stochastic.*

Let $\lambda \in (0,1)$ and L be a $\eta-$stochastic language for some $\eta \in [0,1]$.

3. Show Lemma 1 when $\eta \in \{0, \lambda, 1\}$.

4. Show Lemma 1 when $0 < \lambda < \eta$. *Hint: build $\mathcal{B}$ such that $\mathbb{P}_{\mathcal{B}}(w) = \frac{\lambda}{\eta}\mathbb{P}_{\mathcal{A}}(w)$.*

5. Let $\alpha \in (0,1)$, show that there exists an automaton $\mathcal{B}$ such that $\mathbb{P}_{\mathcal{B}}(w) = \alpha\mathbb{P}_{\mathcal{A}}(w) + (1-\alpha)$. Assume that $\eta < \lambda < 1$ and show that there is a value of $\alpha \in (0,1)$ such that $L = \mathcal{L}_{\mathcal{B}}(\lambda)$. Conclude.

2 Generalized automata

Definition 2. A *generalized automaton (GA)* is a tuple $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$ where A is a finite alphabet, Q is a finite set of states, $S \in \mathbb{R}^{1\times Q}$ is the initial (row) vector, $\mu(a) \in \mathbb{R}^{Q\times Q}$ is a matrix of transition weights for every $a \in A$ and $T \in \mathbb{R}^{Q\times 1}$ is the final (column) vector. The weight of a word w is $\mathbb{P}_{\mathcal{A}}(w) = S\mu(w)T$. For any cut-point $\lambda \in \mathbb{R}$, we define the cut-point language $\mathcal{L}_{\mathcal{A}}(\lambda) = \{w : \mathbb{P}_{\mathcal{A}}(w) > \lambda\}$. For any real number λ and language L, we say that L is a *generalized $\lambda-$language* if there exists a GA $\mathcal{A}$ such that $L = \mathcal{L}_{\mathcal{A}}(\lambda)$. We say that L is a *generalized language* if it is a generalized $\lambda-$language for some $\lambda \in \mathbb{R}$.

Note that contrary to probabilistic automata, the matrices and vectors are not stochastic anymore and we can use negative and bigger than one values, such as -1 or $3/2$. This also applies to cut-points. Clearly a probabilistic automaton is a GA and a stochastic language is a generalized language. The goal is to show that the converse is also true.

1. Show that any generalized language is a generalized $0-$language, *i.e.* of the form $\mathcal{L}_{\mathcal{A}}(0)$ for some $\mathcal{A}$.

2. Let $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$ be GA of dimension n. Let $\mathbf{0}_n$ denote the all-zero column vector of length n and $\mathbf{0}_n^T$ its transpose. Consider the GA $\mathcal{B} = \langle A, Q, S', \mu', T'\rangle$ defined by

$$S' = \begin{bmatrix}\mathbf{0}_n^T & 0 & 1\end{bmatrix}, \qquad \mu'(a) = \begin{bmatrix}\mu(a) & \mu(a)T & \mathbf{0}_n \\ \mathbf{0}_n^T & 0 & 0 \\ S\mu(a) & S\mu(a)T & 0\end{bmatrix}, \qquad T' = \begin{bmatrix}\mathbf{0}_n \\ 1 \\ 0\end{bmatrix}.$$

Show that for any nonempty word w,

$$\mu'(w) = \begin{bmatrix}\mu(w) & \mu(w)T & \mathbf{0}_n \\ \mathbf{0}_n^T & 0 & 0 \\ S\mu(w) & S\mu(w)T & 0\end{bmatrix}.$$

What happens for the empty word ε ? Show that $\mathcal{L}_\mathcal{B}(0) = \mathcal{L}_\mathcal{A}(0) \setminus \{\varepsilon\}$.

3. We say that a matrix M is *doubly zero sum (DZS)* if every row and column of M sums to zero, *i.e.* $\sum_k M_{ik} = \sum_k M_{kj} = 0$ for every i and j. Show that the product of two DZS matrices is DZS.

4. Given a square matrix $M \in \mathbb{R}^{n\times n}$, show that there exists numbers $\alpha_1, \dots, \alpha_n$, $\beta_1, \dots, \beta_n$ and γ such that

$$\widehat{M} = \left[\begin{array}{c|ccc|c} 0 & 0 & \cdots & 0 & 0 \\ \hline \alpha_1 & & & & 0 \\ \vdots & & M & & \vdots \\ \alpha_n & & & & 0 \\ \hline \gamma & \beta_1 & \cdots & \beta_n & 0 \end{array}\right]$$

is DZS. Why are those numbers uniquely defined?

5. Show that for any two matrices M, N we have $\widehat{MN} = \widehat{M}\widehat{N}$.

6. Show that for every generalized language L, $L \setminus \{\varepsilon\}$ is accepted by some GA with cut-point 0 such that all transition matrices are DZS, and the initial and final vectors have entries in $\{0, 1\}$.

7. Let J denote the $n \times n$ matrix whose entries are all 1. Show that $J^2 = nJ$ and that $JM = MJ = 0$ for any DZS matrix M.

8. Given a DSZ matrix M and $\delta \in \mathbb{R}$, define $\widetilde{M}(\delta) = M + \delta J$. Show that $\widetilde{M}(\delta)\widetilde{M}(\delta') = \widetilde{MN}(n\delta\delta')$ for any two DZS M, N.

9. Let L be some generalized $0-$language recognized by $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$ as in question **6.**. Let $\delta \in \mathbb{R}$ and consider the GA $\mathcal{B} = \langle A, Q, S', \mu', T'\rangle$ defined by

$$S' = \begin{bmatrix}S & S\end{bmatrix}, \qquad \mu'(a) = \begin{bmatrix}\widetilde{\mu(a)}(\delta) & 0 \\ 0 & \delta J\end{bmatrix}, \qquad T' = \begin{bmatrix}T \\ -T\end{bmatrix}.$$

Show that $\mathcal{L}_\mathcal{B}(0) = \mathcal{L}_\mathcal{A}(0) \setminus \{\varepsilon\}$ and that there exists a choice of δ such that all transitions matrices of $\mathcal{B}$ have nonnegative entries. What can we say about the entries of S' and T' ?

10. Show how to construct an automaton $\mathcal{C}$ from $\mathcal{B}$ such that $\mathcal{L}_\mathcal{C}(0) = \mathcal{L}_\mathcal{B}(0)$, the initial vector and all transition matrices of $\mathcal{C}$ are stochastic, and the final vector has entries in $\{-1, 0, 1\}$. *Hint: rescale the entries so that they become small.*

11. We now modify $\mathcal{C}$ into $\mathcal{D}$ by adding 1 to all entries of the final vector. Show that $\mathcal{L}_\mathcal{D}(1) = \mathcal{L}_\mathcal{C}(0)$.

12. Let $\mathcal{D} = \langle A, Q, S, \mu, T\rangle$ be the automaton obtained at the previous question and consider $\mathcal{E} = \langle A, Q, S', \mu', T'\rangle$ defined by

$$S' = \begin{bmatrix}S & 0\end{bmatrix}, \qquad \mu'(a) = \tfrac{1}{3}\begin{bmatrix}\mu(a) & \mu(a)T \\ \mathbf{0}_n^T & 0\end{bmatrix}, \qquad T' = \begin{bmatrix}\mathbf{0}_n & 1\end{bmatrix}.$$

Show that $\mathcal{E}$ is substochastic and that for any word w, $\mathbb{P}_\mathcal{E}(w) = 3^{-|w|}\mathbb{P}_\mathcal{D}(w)$.

13. Explain how to modify $\mathcal{E}$ into a *stochastic automaton* $\mathcal{F}$ such that $\mathbb{P}_\mathcal{F}(w) = \frac{1}{2}\mathbb{P}_\mathcal{E}(w) + \frac{1}{2}(1 - 3^{-|w|})$. Conclude that $\mathcal{L}_\mathcal{F}(\frac{1}{2}) = \mathcal{L}_\mathcal{D}(1)$. What have we shown overall?

14. Show that every generalized language is stochastic. *Hint: show that if L is stochastic then $L \cup \{\varepsilon\}$ is stochastic.*

15. For any word $w = w_1 \cdots w_k$, denote by $w^R = w_k \cdots w_1$ the reverse of w. Define the reverse of a language L by $L^R = \{w^R : w \in L\}$. Show that the reverse image of a stochastic language is stochastic. *Hint: observe that for any two matrices A and B, $(AB)^T = B^T A^T$.*

Solutions to exercises

1. Consider a *complete*[1] *deterministic* finite automaton that recognizes L and transform it into a probabilistic automaton $\mathcal{A}$ by putting a weight of 1 on every edge. This is indeed a probabilistic automaton because it is deterministic (there is a single non-zero entry on every row of every matrix). By construction, for every word w, $\mathbb{P}_{\mathcal{A}}(w) = 1$ if $w \in L$ and 0 otherwise.

2. Apply the previous question to L' to get $\mathcal{B}$. Let $\mathcal{A}$ be such that $L = \mathcal{L}_{\mathcal{A}}(\lambda)$ for some $\lambda \in [0,1]$. We then build the product automaton $\mathcal{C}$ such that $\mathbb{P}_{\mathcal{C}}(w) = \mathbb{P}_{\mathcal{A}}(w)\mathbb{P}_{\mathcal{B}}(w)$. Then there are three cases to consider:

- if $w \in L \cap L'$ then $\mathbb{P}_{\mathcal{C}}(w) = \mathbb{P}_{\mathcal{A}}(w) > \lambda$,
- if $w \notin L'$ then $\mathbb{P}_{\mathcal{C}}(w) = 0 \leqslant \lambda$,
- otherwise $w \in L'$ but $w \notin L$ then $\mathbb{P}_{\mathcal{C}}(w) = \mathbb{P}_{\mathcal{A}}(w) \leqslant \lambda$.

Therefore $\mathcal{L}_{\mathcal{C}}(\lambda) = L \cap L'$. We get the result for $L \setminus L'$ by noticing that $L \setminus L' = L \cap L''$ where L'' is the complement of L', a regular language.

3. If $\eta = \lambda$ then there is nothing to prove. If $\eta = 0$ then L is regular so we can apply question **1.**. By construction, for every word w, $\mathbb{P}_{\mathcal{A}}(w) = 1$ if $w \in L$ and 0 otherwise. It follows that $\mathcal{L}_{\mathcal{A}}(\lambda) = L$ because $\lambda \in (0,1)$. If $\eta = 1$ then $L = \varnothing$ which is trivially $\lambda-$stochastic (take any automaton without any accepting state).

4. Build automaton $\mathcal{B}$ from $\mathcal{A}$ by multiplying the initial distribution by $\frac{\lambda}{\eta} \in (0,1)$: this gives a substochastic automaton that can be made stochastic. Then $\mathbb{P}_{\mathcal{B}}(w) = \frac{\lambda}{\eta}\mathbb{P}_{\mathcal{A}}(w)$ for every word w. It follows that $x \in L$ if and only if $\mathbb{P}_{\mathcal{A}}(w) > \lambda$ if and only if $\mathbb{P}_{\mathcal{B}}(w) > \eta$.

5. The first part is immediate: $\mathcal{B}$ is the convex combination of $\mathcal{A}$ and the automaton with constant probability 1. We then have that $w \in \mathcal{L}_{\mathcal{B}}(\lambda)$ if and only if $\alpha\mathbb{P}_{\mathcal{A}}(w) + (1-\alpha) > \lambda$ if and only if $\mathbb{P}_{\mathcal{A}}(w) > \frac{\lambda+\alpha-1}{\alpha}$. We then choose α so that this number corresponds to L: $\frac{\lambda+\alpha-1}{\alpha} = \eta$ if and only if $\lambda + \alpha - 1 = \alpha\eta$ if and only if $\alpha = \frac{1-\lambda}{1-\eta}$ and this is indeed a number in $(0,1)$ because $\eta < \lambda < 1$.
We have therefore shown Lemma 1 in all possible cases.

---

1. Let $L = \mathcal{L}_{\mathcal{A}}(\lambda)$ be a GA recognized by $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$. Consider the GA $\mathcal{B} = \langle A, Q, S', \mu', T'\rangle$ defined by

$$S' = \begin{bmatrix} S & -\lambda \end{bmatrix}, \qquad \mu'(a) = \begin{bmatrix} \mu(a) & 0 \\ 0 & 1 \end{bmatrix}, \qquad T' = \begin{bmatrix} T \\ 1 \end{bmatrix}.$$

Then we immediately check that for any word w,

$$S'\mu'(w)T' = \begin{bmatrix} S & -\lambda \end{bmatrix} \begin{bmatrix} \mu(w) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} T \\ 1 \end{bmatrix} = S\mu(w)T - \lambda.$$

Therefore $S\mu(w)T > \lambda$ if and only if $S'\mu'(w)T' > 0$.

2. One checks

$$\begin{bmatrix} \mu(w) & \mu(w)T & \mathbf{0}_n \\ \mathbf{0}_n^T & 0 & 0 \\ S\mu(w) & S\mu(w)T & 0 \end{bmatrix} \begin{bmatrix} \mu(a) & \mu(a)T & \mathbf{0}_n \\ \mathbf{0}_n^T & 0 & 0 \\ S\mu(a) & S\mu(w)T & 0 \end{bmatrix} = \begin{bmatrix} \mu(wa) & \mu(wa)T & \mathbf{0}_n \\ \mathbf{0}_n^T & 0 & 0 \\ S\mu(wa) & S\mu(wa)T & 0 \end{bmatrix}$$

and therefore it follows by induction that for any nonempty word w,

$$S'\mu'(w)T' = \begin{bmatrix} \mathbf{0}_n^T & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu(w) & \mu(w)T & \mathbf{0}_n \\ \mathbf{0}_n^T & 0 & 0 \\ S\mu(w) & S\mu(w)T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{0}_n \\ 1 \\ 0 \end{bmatrix} = S\mu(w)T.$$

On the other hand, the empty word has weight $S'T' = 0$ which might be different from its weight ST in $\mathcal{A}$. It follows that $\mathcal{L}_{\mathcal{B}}(0)$ never contains the empty word, even if $\mathcal{L}_{\mathcal{A}}(0)$ did.

3. Let M, N be DZS. Then

$$\sum_i (MN)_{ij} = \sum_i \sum_k M_{ik}N_{kj} = \sum_k N_{kj}\left(\sum_i M_{ik}\right) = 0$$

and similarly for row sums.

[1]For every state q and letter a, there is a transition from q labelled by a.

4. We have no choice (unicity is clear) since every row and column must sum to zero: take $\alpha_i = \sum_{j=1}^n M_{ij}$ and $\beta_j = \sum_{i=1}^n M_{ij}$. Then we must take $\gamma = -\sum_{i=1}^n \alpha_i = -\sum_{j=1}^n \beta_j$ so we need to check that those sums are indeed equal:

$$\sum_{i=1}^n \alpha_i = \sum_{i=1}^n \sum_{j=1}^n M_{ij} = \sum_{j=1}^n \sum_{i=1}^n M_{ij} = \sum_{j=1}^n \beta_j.$$

5. A simple blockwise product shows that

$$\widehat{M}\widehat{N} = \begin{bmatrix} 0 & 0 & 0 \\ \alpha & M & 0 \\ \gamma & \beta & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ \alpha' & N & 0 \\ \gamma' & \beta' & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ * & MN & 0 \\ * & * & 0 \end{bmatrix}$$

for some values $(*)$. And since the product is DZS by **3.**, it must be $\widehat{MN}$ by unicity.

6. Let L be a GA, then by questions **1.** and **2.**, $L \setminus \{\varepsilon\} = \mathcal{L}_{\mathcal{A}}(0)$ for some GA $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$ such that S and T have entries in $\{0,1\}$. Now consider $\mathcal{B} = \langle A, Q, S', \mu', T'\rangle$ where $S' = \begin{bmatrix} 0 & S & 0 \end{bmatrix}$, $\mu'(a) = \widehat{\mu(a)}$ and $T' = \begin{bmatrix} 0 & T & 0 \end{bmatrix}^T$. The transition matrices are DZS by construction and for any word nonempty w,

$$\mathbb{P}_{\mathcal{B}}(w) = S'\mu'(w)T' = S'\widehat{\mu(w)}T' = \begin{bmatrix} 0 & S & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ * & \mu(w) & 0 \\ * & * & 0 \end{bmatrix} \begin{bmatrix} 0 \\ T \\ 0 \end{bmatrix} = S\mu(w)T = \mathbb{P}_{\mathcal{A}}(w)$$

using **5.**, and $\mathbb{P}_{\mathcal{B}}(\varepsilon) = ST = \mathbb{P}_{\mathcal{A}}(\varepsilon)$. Therefore $\mathcal{L}_{\mathcal{B}}(0) = \mathcal{L}_{\mathcal{A}}(0) = L \setminus \{\varepsilon\}$ and clearly S' and T' have entries in $\{0,1\}$.

7. An immediate calculation shows that $J^2 = nJ$. Now if M is DZS, then

$$(JM)_{ij} = \sum_{k=1}^n J_{ik} M_{kj} = \sum_{k=1}^n M_{kj} = 0$$

and similarly for JM.

8. By definition and using **7.**,

$$\widetilde{M}(\delta)\widetilde{M}(\delta') = (M + \delta J)(N + \delta' J) = MN + \delta JN + \delta' MJ + \delta\delta' J^2 = MN + n\delta\delta' J = \widetilde{MN}(n\delta\delta').$$

9. Using question **8.**, check that for every nonempty word w,

$$\begin{aligned} S'\mu'(w)T &= \begin{bmatrix} S & S \end{bmatrix} \begin{bmatrix} \widetilde{\mu(w)}(n^{|w|-1}\delta^{|w|}) & 0 \\ 0 & (\delta J)^{|w|} \end{bmatrix} \begin{bmatrix} T \\ -T \end{bmatrix} \\ &= S\widetilde{\mu(w)}(n^{|w|-1}\delta^{|w|})T - S\delta^n J^n T \\ &= S\mu(w)T + n^{|w|-1}\delta^{|w|} SJT - n^{|w|-1}\delta^{|w|} SJT \\ &= S\mu(w)T. \end{aligned}$$

It follows that $S'\mu'(w)T > 0$ if and only if $S\mu(w)T > 0$. On the other hand, $\mathbb{P}_{\mathcal{B}}(\varepsilon) = S'T' = ST - ST = 0$ so $\varepsilon \notin \mathcal{L}_{\mathcal{B}}(0)$. Therefore $\mathcal{L}_{\mathcal{B}}(0) = \mathcal{L}_{\mathcal{A}}(0) \setminus \{\varepsilon\}$. Clearly for a matrix M, since $\widetilde{M}(\delta) = M + \delta J$, we can ensure that it has nonnegative entries by taking δ large enough. The entries of S are still in $\{0,1\}$ whereas the entries of T' are now in $\{-1,0,1\}$.

10. For every a, $\mu'(a)$ is nonnegative, and S' is also nonnegative. Therefore there exists $\delta > 0$ such that $\delta\mu'(a)$ and $\delta S'$ are substochastic. Now consider $\mathcal{C} = \langle A, Q, S'', \mu'', T''\rangle$ defined by

$$S'' = \begin{bmatrix} \delta S' & \alpha \end{bmatrix}, \qquad \mu''(a) = \begin{bmatrix} \delta\mu'(a) & v_a \\ 0 & 1 \end{bmatrix}, \qquad T' = \begin{bmatrix} T' \\ 0 \end{bmatrix}.$$

where $v_a \in \mathbb{R}^n$ and α are such that S'' and $\mu''(a)$ are stochastic, which is possible by the choice of δ. It follows that for every word w,

$$S''\mu''(w)T'' = \begin{bmatrix} S' & 1 \end{bmatrix} \begin{bmatrix} \delta^{-|w|}\mu'(w) & * \\ 0 & 1 \end{bmatrix} \begin{bmatrix} T' \\ 0 \end{bmatrix} = \delta^{-|w|} S'\mu'(w)T'$$

and therefore $S''\mu''(w)T'' > 0$ if and only if $S'\mu'(w)T' > 0$. By construction, the entries of T'' are in $\{-1,0,1\}$.

11. Let $\mathbf{1}_n$ be the all-one vector of size n, then for every word w,

$$S''\mu''(w)(T''+\mathbf{1}_n) = S''\mu''(w)T'' + S''\mathbf{1}_n = S''\mu''(w)T'' + 1$$

since the $M\mathbf{1}_n = \mathbf{1}_n$ for any stochastic matrix, and $S\mathbf{1}_n = 1$ for a stochastic vector S. Therefore $\mathcal{L}_{\mathcal{D}}(1) = \mathcal{L}_{\mathcal{C}}(0)$.

12. One easily checks by induction that for every word w,

$$S'\mu'(w)T' = \begin{bmatrix} S & 0 \end{bmatrix}, \qquad 3^{-|w|}\begin{bmatrix} \mu(w) & \mu(w)T \\ \mathbf{0}_n^T & 0 \end{bmatrix}\begin{bmatrix} \mathbf{0}_n & 1 \end{bmatrix} = 3^{-|w|}S\mu(w)T.$$

Furthermore, since for every a, $\mu(a)$ is stochastic and T has entries in $\{0,1,2\}$, the sum of each line in $\mu'(a)$ is positive and at most $(1+2)/3 = 1$. Clearly the initial vector S' is stochastic and the final vector has entries in $\{0,1\}$.

13. It is trivial to build a PA such that the probability of every word w is $1-3^{-|w|}$, then by taking the convex combination of this automaton and $\mathcal{E}$ we get the result. It follows that

$$\mathbb{P}_{\mathcal{F}}(w) > \tfrac{1}{2} \quad\Leftrightarrow\quad \tfrac{1}{2}\mathbb{P}_{\mathcal{E}}(w) + \tfrac{1}{2}(1-3^{-|w|}) > \tfrac{1}{2} \quad\Leftrightarrow\quad \mathbb{P}_{\mathcal{E}}(w) > 3^{-|w|} \quad\Leftrightarrow\quad 3^{-|w|}\mathbb{P}_{\mathcal{D}}(w) > 3^{-|w|} \quad\Leftrightarrow\quad \mathbb{P}_{\mathcal{D}}(w) > 1.$$

Overall, we have shown that for every generalized language L, $L \setminus \{\varepsilon\}$ is a stochastic language.

14. Let L be a generalized language: then $L' = L \setminus \{\varepsilon\}$ is stochastic by the previous question. If $L = L'$ then we have shown the result. Otherwise $L = L' \cup \{\varepsilon\}$ is stochastic by question **2.** of the first exercice since $\{\varepsilon\}$ is regular.

15. Let $L = \mathcal{L}_{\mathcal{A}}(\lambda)$ be a stochastic language recognized by $\mathcal{A} = \langle A, Q, S, \mu, T\rangle$. Consider the GA $\mathcal{B} = \langle A, Q, S', \mu', T'\rangle$ defined by

$$S' = T^T \qquad \mu'(a) = \mu(a)^T \qquad T' = S^T.$$

Now we check that for any word $w = w_1 \cdots w_k$,

$$S'\mu'(w)T = T^T\mu(w_1)^T \cdots \mu(w_k)^T S^T = (S\mu(w_n)\cdots\mu(w_1)T)^T = (S\mu(w^R)T)^T = S\mu(w^R)T$$

since the transpose of a real is itself. It follows that L^R is a generalized language and therefore a stochastic language.