24. *Sur l'extension et la restriction d'un langage associé à l'espace contextuel*

Par Masami ITO
Université de Kyoto-Sangyo

(Comm. by Kinjirô KUNUGI, M. J. A., Feb. 12, 1972)

Dans ce mémoire, nous définissons l'extension et la restriction d'un langage associé à l'espace contextuel (défini par S. Marcus (1)) et nous explorons une structure de l'espace contextuel à l'aide de ces notions.

1. Langage. Nous appelons $\mathcal{L}=(A, L)$ un *langage*. Où, A est un ensemble fini d'éléments nommés mots. L est un sous-ensemble quelconque de A^+, où A^+ est l'ensemble des suites finies, nommées phrases, d'éléments de A.

Exemple. Posons $A=\{a, b, c\}$ et $L=\{a, ab, abc, abca^2b^3, \cdots\}$. $\mathcal{L}=(A, L)$ est alors un langage.

2. Quelque notions nécessaires. Nous introduisons le mot vide ε, jouissant de la propriété $\varepsilon x=x\varepsilon=x$ pour toute phrase $x \in A^+$. En outre, nous écrivons $A^*=A^+ \cup \{\varepsilon\}$.

Nous appelons (x, y) un contexte, où $x, y \in A^*$ et $xy \neq \varepsilon$. Soient z un élément de A^+ et (x, y) un contexte. Si $xzy \in L$, nous disons que la phrase z est admise par le contexte (x, y) dans le langage $\mathcal{L}$.

Considérons $C(x)_{\mathcal{L}}$ (où $x \in A^+$) comme l'ensemble de tous les contextes admettant x, i.e., $C(x)_{\mathcal{L}}=\{(w, z)\,;\, wz \neq \varepsilon, wxz \in L\}$.

3. Distance et diamètre. Nous définirons une distance entre deux phrases quelconques de A^+. Supposons que nous puissions trouver une séquence (x_i) (où $x_i \in A^+$ et $i=0, 1, 2, \cdots, n-1, n$) telle que $x=x_0$, $x_n=y$ et $C(x_i)_{\mathcal{L}} \cap C(x_{i+1})_{\mathcal{L}} \neq \emptyset$ $(i=0, 1, 2, \cdots, n-1)$ pour $x, y \in A^+$ (où $x \neq y$). Nous appelons alors (x_i) $(i=0, 1, 2, \cdots, n-1, n)$ une chaîne de x à y et n une longueur de chaîne. Si nous pouvons trouver une des chaînes de x à y les plus courtes, nous disons que la distance entre x et y est égale à la longueur de cette chaîne et nous employons le symbole $\operatorname{dis}(x, y)_{\mathcal{L}}$ pour représenter la distance entre x et y. Au cas qu'il n'existe pas de telle chaîne, nous disons que la distance entre x et y est infinie et nous écrivons $\operatorname{dis}(x, y)_{\mathcal{L}}=+\infty$. Nous définissons $\operatorname{dis}(x, x)_{\mathcal{L}}=0$ pour toute phrase $x \in A^+$. Il est facile de voir que cette définition remplit les conditions de distance.

Le diamètre du langage $\mathcal{L}$ se présente ainsi:

$$d(\mathcal{L})=\sup_{x, y \in L} \operatorname{dis}(x, y)_{\mathcal{L}}.$$

4. Espace contextuel. Nous appelons un langage *l'espace con-*

textuel, qui a la distance introduite à la manière plus haut.

5. Extension et Restriction d'un langage. Soient $\mathcal{L}=(A, L)$ et $\mathcal{K}=(B, K)$ deux langages. Lorsqu'ils remplissent les quatre conditions suivantes, nous disons que le langage $\mathcal{K}$ est une *extension* du langage $\mathcal{L}$ et que le langage $\mathcal{L}$ est une *restriction* du langage $\mathcal{K}$:

1) $A \subseteq B$; 2) $L \subset K$; 3) $\operatorname{dis}(x, y)_{\mathcal{L}} = \operatorname{dis}(x, y)_{\mathcal{K}}$, *pour toute paire de phrases* $x, y \in L$; 4) $d(\mathcal{L}) < d(\mathcal{K})$.

Nous pouvons avoir immédiatement la proposition suivante:

Proposition. *Soient $\mathcal{L}, \mathcal{K}$ et $\mathcal{J}$ trois langages. Si $\mathcal{K}$ est une extension de $\mathcal{L}$ et que $\mathcal{J}$ soit une extension de $\mathcal{K}$, $\mathcal{J}$ est alors une extension de $\mathcal{L}$.*

6. Existence d'une extension d'un langage. Nous démontrerons le théorème suivant:

Théorème 1. *Soit $\mathcal{L}=(A, L)$ (où $d(\mathcal{L}) < +\infty$) un langage. Pour chaque nombre naturel n, il existe une extension $\mathcal{M}=(B, M)$ de $\mathcal{L}$ telle que $n = d(\mathcal{M}) - d(\mathcal{L})$.*

Démonstration. Il suffit de démontrer qu'il existe une extension $\mathcal{K}$ de $\mathcal{L}$ telle que $d(\mathcal{K}) - d(\mathcal{L}) = 1$, d'après la proposition ci-dessus et par induction.

Soit $d(\mathcal{L}) = m < +\infty$. Nous pouvons trouver alors deux éléments $w, z \in L$ tels que $\operatorname{dis}(w, z)_{\mathcal{L}} = m$. En profitant de ce w, nous pouvons construire un langage $\mathcal{K}=(C, K)$. Pour construire ce langage, nous introduisons un nouveau mot α qui n'appartient pas à A. Nous posons $C = A \cup \{\alpha\}$ et $K = L \cup \{\alpha^p w \alpha^q ; p, q = 1, 2, 3, \cdots\}$. Il n'est pas difficile de vérifier que le langage $\mathcal{K}$ soit une extension du langage $\mathcal{L}$ tel que $d(\mathcal{K}) = m+1$.

Corollaire. *Pour chaque nombre naturel n, il existe un langage $\mathcal{L}$ tel que $d(\mathcal{L}) = n$.*

Démonstration. Posons $\mathcal{M}=(A, A^+)$ (où $A=\{a\}$). Il est aisé de voir que $d(\mathcal{M})=1$. D'après le théorème 1 et par induction, nous avons immédiatement le corollaire.

7. Restriction d'un langage. Il n'est pas facile d'explorer des structures de langages généraux au point de vue de l'espace contextuel. Nous nous bornerons ici à donner un résultat sur un langage spécial.

Théorème 2. *Soit $\mathcal{L}=(A, L)$ (où $A=\{a_i ; i=1, 2, \cdots, n\}$) un langage qui remplit les trois conditions suivantes:*

(A) *Soit x un élément de L. Pour chaque résolution de x, i.e., $x = uyv$ (où $u, v \in A^*, y \in A^+$), y est aussi un élément de L.*

(B) *Pour chaque $x \in L$, il existe deux éléments $u, v \in A^*$ (où $uv \in A^+$) tels que $uxv \in L$.*

(C) *Pour chaque $a_i \in A$ $(i=1, 2, \cdots, n)$, il existe un élément $w(i) \in A^*$ tel que $a_i w(i) a_i \in L$.*

Pour ce langage, nous avons au moins un des deux cas suivants:

(1) $d(\mathcal{L})\leq 2n$.

(2) *Il existe une restriction* $\mathcal{M}$ *de* $\mathcal{L}$ *telle que* $d(\mathcal{M})\leq p(n)$, *où* $p(n)=n$ *pour un nombre pair* n *et* $p(n)=n-1$ *pour un nombre impair* n.

Démonstration. Remarquons d'abord le fait suivant:

Soient x, y deux éléments de A^+. Si dis $(x, y)_{\mathcal{L}}<+\infty$, nous avons alors dis $(x, y)_{\mathcal{L}}\leq 2n$. Car nous pouvons trouver un contexte appartenant à $C(x)_{\mathcal{L}}\cap C(y)_{\mathcal{L}}$ sous la forme (a, ε) ou (ε, a) (où $a\in A$) d'après la condition (A), si $C(x)_{\mathcal{L}}\cap C(y)_{\mathcal{L}}\neq\emptyset$.

Pour $n=1$, nous pouvons vérifier facilement le théorème. Pour le cas général, nous introduisons les signes $\sim$ et $\approx$. Nous écrivons $a_i\sim a_j$, lorsqu'il existe un élément $w(i, j)\in A^*$ tel que $a_iw(i, j)a_j\in L$ ou $a_jw(i, j)a_i\in L$. S'il existe une séquence des nombres naturels (i_p) (où $p=0, 1, 2, \cdots, m-1$) telle que $i=i_0, i_m=j$ et $a_{i_p}\sim a_{i_{p+1}}$ $(p=0, 1, 2, \cdots, m-1)$, nous écrivons $a_i\approx a_j$.

Posons $A_i=\{a_k\,;\, a_i\approx a_k\}$.

Les propriétés de A_i:

1) Pour chaque $i(i=1, 2, \cdots, n)$, nous avons $a_i\in A_i$.

2) Si $a_k\in A_i$, nous avons $a_i\in A_k$.

D'après ces propriétés, il n'existe que l'un des deux cas suivants:

(1) Pour chaque $i(i=1, 2, \cdots, n)$, nous avons $A_i=A$.

(2) Il existe au moins deux nombres naturels i, j tels que $A_i\cap A_j=\emptyset$.

En cas de (1). Soient u, v deux éléments quelconques de L. D'après les propriétés (A) et (B), il existe deux nombres naturels i, j et nous avons le résultat suivant:

1) Nous avons au moins un de $a_iu\in L$ et de $ua_i\in L$.

2) Nous avons au moins un de $a_jv\in L$ et de $va_j\in L$.

D'après $a_i\approx a_j$, il existe une séquence (i_p) $(p=0, 1, 2, \cdots, m)$ telle que $i_0=i, i_m=j$ et $a_{i_p}\sim a_{i_{p+1}}$ $(i_p\neq i_{p+1})$ $(p=0, 1, 2, \cdots, m-1)$. Par une considération brève, nous pouvons supposer préalablement que $w(i_p, i_{p+1})\neq\varepsilon$ $(p=0, 1, 2, \cdots, m-1)$.

Posons maintenant $w'(i_p)=a_i$, si $w(i_p)=\varepsilon$. Posons $w'(i_p)=w(i_p)$, si $w(i_p)\neq\varepsilon$.

Considérons une séquence d'éléments de A^+:

$$u, w'(i_0), w(i_0, i_1), w'(i_1), w(i_1, i_2), \cdots, w(i_{m-1}, i_m), w'(i_m), v.$$

Il n'est pas difficile de voir que cette séquence soit une chaîne de u à v dans le langage $\mathcal{L}$. Nous avons ainsi dis $(u, v)_{\mathcal{L}}<+\infty$.

D'après la remarque donnée au début, nous avons dis $(u, v)_{\mathcal{L}}\leq 2n$. Par conséquent, nous avons $d(\mathcal{L})\leq 2n$.

En cas de (2). Il existe deux nombres naturels i, j tels que $A_i\cap A_j$

$=\emptyset$. Nous pouvons supposer que $|A_i| \leq |A_j|$ (où $|A_k|$ signifie le nombre d'éléments de l'ensemble A_k). Par conséquent, nous avons $|A_i| \leq p(n)/2$. Posons $M^i = \{x; x \in L \cap A_i^+\}$. Il est évident que $M^i \neq \emptyset$. Construisons un langage $\mathcal{M}^i = (A_i, M^i)$. Il n'est pas difficile de vérifier que le langage $\mathcal{M}^i$ remplisse les conditions du théorème 2.
D'après $|A_i| \leq p(n)/2$ et par induction, nous avons le résultat suivant:

1) $d(\mathcal{M}^i) \leq 2 \cdot p(n)/2 = p(n)$, ou
2) Il existe au moins une restriction $\mathcal{N}$ du langage $\mathcal{M}^i$ telle que $d(\mathcal{N}) \leq p(n)/2$.

Nous pouvons vérifier facilement que le langage $\mathcal{M}^i$ remplit les trois premières conditions pour la définition d'une restriction du langage $\mathcal{L}$. De là et par induction, dans tous les cas de 1) et de 2), nous pouvons voir immédiatement que le théorème se réalise.

Références

[1] S. Marcus: Introduction mathématique à la linguistique restructurale. Dunod (1967).

[2] M. Ito: Sur l'extension et la restriction d'un langage associé à l'espace contextuel (en japonais). Jyôhô-shori, Vol. 13 No. 1, 8–12, Information Processing Society of Japan. Tokyo (1972).