t.a. informations

REVUE INTERNATIONALE du TRAITEMENT AUTOMATIQUE du LANGAGE

1968 — NUMÉRO 1

Sommaire

pages

LINGUISTIQUE MATHÉMATIQUE ET TRADUCTION AUTOMATIQUE

La deuxième conférence internationale sur le traitement automatique des langues (Grenoble, 23-25 août 1967) 1

A. Baille : Les Systèmes logiques et les modèles sémantiques de langues naturelles 5

G. Veillon : Description du langage pivot du système de traduction automatique du C.E.T.A. 8

M. Coyaud : Analyse et Classification automatiques en Chinois 18

L. Negrescu, P. Schveiger : Un modèle d'analyse et de syntèse de la déclinaison 29

BIBLIOGRAPHIE

Comptes rendus de T.A. Informations et du Groupe d'Études sur l'Information Scientifique 33

ÉDITIONS KLINCKSIECK / PARIS

Publié avec le concours du C N R S

HUÁNG ZÀIJŪN : Cóng jiǎwén, jīnwén liàngcíde yīngyòng, kaŏchá hànyŭ liàngcíde qĭyúan yŭ fāzhăn. Zhōngguó yŭwén, n° 6, pp. 432-441 (1964).
JAHONTOV, S. E. : Drevnekitajskij jazyk. ANSSSR, Moscou (1965).
LÍNG QUÁNFANG : Yòng yŭ rénde gè tĭ liàngcí. Yŭyán zhīshí, n° 6 (1955).
LIÚ SHÌRÚ : Wèi Jì Năn Bĕi cháo liàngcíde yánjīu. Zhōnghúa shūjú. Pékin (1965).
LIÚ SHÌSHENG : Tán liàngcí. Yŭyán xuéxí, n° 10 (1957).
WANG Lì (Wang Liaoyi) : Yoŭguan rén wù hé xíngwéide xūcí. Yŭ wén xuéxí cóngshū, Zhōnggúo qíngnían chūbănshè, Pékin (1955).
ZHŌNG BENKANG : Shìwù hé liàngcí. Yŭyán zhīshí, n° 8 (1959).

Numérales d'autres langues.

DELOUSTAL, R. : Des déterminatifs en annamite, BEFO, XIV, 5 (1914).
EMENAU, M. B. : Studies in Vietnamese Grammar. Univ. of Calif. Press, Berkeley (1951).
HAAS, M. R. : The use of numeral classifiers in Thai, Language, XVIII, p. 201 sq. (1942).
HAGUENHAUER, Ch. : Morphologie du japonais moderne, Klincksieck, Paris (1951).
HOSPITALIER, J. J. : Grammaire laotienne, Imprimerie Nationale, Paris (1937).
KOUTSOUDAS, A. : Writing Transformational Grammars. Mc Graw-Hill, New York (1966).
NAGANUMA, N. : Basic Japanese Course, Tokyo (1951).

UN MODÈLE D'ANALYSE ET DE SYNTHÈSE DE LA DÉCLINAISON

par Liviu NEGRESCU et Paul SCHVEIGER *

Dans le présent article, on construit un modèle de l'analyse et de la synthèse de la déclinaison des substantifs que l'on a déjà expérimenté sur une calculatrice électronique. Le modèle présenté se réfère aux substantifs appartenant à une langue réelle et par conséquent le but poursuivi concerne la solution de certains problèmes qui se rapportent à la déclinaison des substantifs ayant des radicaux qui contiennent des lettres d'une valeur variable.

Tout d'abord, introduisons quelques notions.

On entend par « alphabet » un ensemble fini de symboles, et on appelle « lettres » les éléments de cet ensemble.

Une succession limitée de lettres sera nommée « mot » dans l'alphabet considéré. On entend par « mot vide » celui qui ne contient aucune lettre.

Soit $\mathfrak{A}$ un alphabet quelconque. On prend en considération deux ensembles finis et disjoints R et T dont les éléments sont des suites finies de mots dans l'alphabet $\mathfrak{A}$. R et T étant des ensembles finis, on peut les

* Facultatea de Filologie, 31, Horia, CLUJ (R. P. de Roumanie).

écrire sous la forme de deux suites en utilisant des indices. Soit :

$R^1, R^2, \ldots, R^n$ et $T^1, T^2, \ldots, T^m$

les suites respectives. Chaque élément $R^i \in R$ ($i = 1, 2, \ldots, n$) est à son tour une suite de mots. On notera les termes de cette suite en utilisant les indices inférieurs. Ainsi l'élément R^i représente la suite $R^i_1, R^i_2, \ldots, R^i_{ni}$. R^i_j représente ici une écriture abrégée d'un certain mot dans l'alphabet et non pas une lettre de cet alphabet. On suppose également que les termes R^i_j sont différents du mot vide.

Les éléments R^i seront nommés des « suites de radicaux » et les termes R^i_j d'une telle suite seront nommés « radicaux ». D'une manière analogue on emploie la notion de « suite de terminaisons » pour un élément $T^i \in T$ et les termes d'une telle suite seront nommés « terminaisons ».

On attachera à chaque radical son indice j, de telle manière qu'à chaque suite de radicaux R^i correspondent une suite numérique qui sera nommée la « suite des indices des radicaux respectifs ».

D'une façon similaire on introduira la notion de « suite des indices des terminaisons T^i ».

On suppose qu'entre R et T est donnée une correspondance qui satisfait les propriétés suivantes :

a. A chaque élément $R^i \in R$ ($i = 1, 2, \ldots, n$), correspond un élément et seulement un $T^j \in T$. Les éléments R^i et T^j rappelés plus haut s'appelent correspondants. On note avec D^j le sous-ensemble des éléments de R auxquels correspondent un même élément $T^j \in T$.

En ce qui concerne D^j on a les propriétés suivantes :

b. Les sous-ensembles D^j ($j = 1, 2, \ldots, m$) ne sont pas vides ;

c. Aux éléments du sous-ensemble D^j correspondent une même suite d'indices ;

d. Si R^i et T^j sont deux éléments correspondants, entre les suites des indices de ces éléments existe une correspondance pour laquelle, à chaque terme de la suite des indices des terminaisons T^j, correspond uniquement un terme de la suite des indices des radicaux R^i. Cette correspondance est unique pour n'importe quelle suite de radicaux de D^j.

En se fondant sur ce qu'on a démontré plus haut on peut définir la notion de substantif dans l'alphabet $\mathcal{A}$ et ensuite on peut formuler le problème de l'analyse et de la synthèse de la déclinaison des substantifs.

On dit du mot A de l'alphabet $\mathcal{A}$ qu'il est un substantif relatif à la paire des ensembles R et T, s'il existe au moins une paire des mots $R^i_j \in R^i \in D^s$ et $T^s_k \in T^s$, tel que

$$A = R^i_j \quad T^s_k \tag{1}$$

Lorsqu'il existe exactement une paire de mots R^i_j, T^s_k pour laquelle a lieu la relation (1) on dit que le substantif A est non-dégénéré relatif à la paire R, T.

Supposons que le symbole * n'appartient pas à l'alphabet $\mathcal{A}$. Soit $\mathcal{B} = \mathcal{A} \cup \{*\}$ l'extension de l'alphabet obtenu en ajoutant la lettre *.

On appelle analyse de la déclinaison du substantif A, le mot B de l'alphabet $\mathcal{B}$, pour lequel on a la relation

$$B = R^i_j * T^s_k \tag{2}$$

où $R^i_j \in R^i \in D^s$, $T^s_k \in T^s$ et $A = R^i_j \; T^s_k$. Dans l'hypothèse que A est un substantif non-dégénéré relatif à la paire R, T, le mot B est unique. Dans ce cas on peut aisément construire un algorithme qui appliqué à A peut donner pour résultat B. Notons avec Γ cet algorithme qu'on appelera par la suite l'algorithme de l'analyse de la déclinaison des substantifs. On a la relation

$$B = \Gamma\,(A) \tag{3}$$

On appelle synthèse de la déclinaison le substantif A de l'alphabet $\mathcal{A}$ par rapport au mot analyse $B = R^i_j * T^s_k$, le mot C de l'alphabet $\mathcal{B}$, qui satisfait la relation

$$C = R^i_{j1} * T^s_1 * R^i_{j2} * T^s_2 * \ldots * R^i_{js} * T^s_{ms} \tag{4}$$

où j_r est le terme de la suite des indices de radicaux R^i qui correspond conformément à la propriété *d*, au terme r de la suite des indices des terminaisons T^s, m_s étant le nombre des termes de cette dernière suite et $1 \leqslant K \leqslant m_s$ et $j_k = j$.

On construit la synthèse C tout de suite étant donné l'analyse B d'un substantif et ce, uniquement conformément aux propriétés *a, b, c, d*. Soit Ω l'algorithme qui détermine le mot synthèse C par son application au mot analyse B. Donc

$$C = \Omega\,(B) = \Omega\,(\Gamma\,(A)) = \Delta\,(A) \tag{5}$$

On a noté par Δ la composition des algorithmes Γ et Ω. L'algorithme Δ est applicable à tout substantif non-dégénéré et réalise l'analyse et la synthèse du substantif respectif.

On peut donner maintenant le schéma abrégé de l'algorithme Δ dans le langage des algorithmes normaux. On en suppose que $\alpha_1, \alpha_2, \ldots, \alpha_m$, $\square$, $O \notin \mathcal{A}$. Soit $\mathcal{C} = \mathcal{B} \cup \{\alpha_1, \alpha_2, \ldots, \alpha_m, \square, O\}$. Le schéma abrégé de l'algorithme normal Δ dans l'alphabet $\mathcal{C}$ sera dans ce cas le suivant :

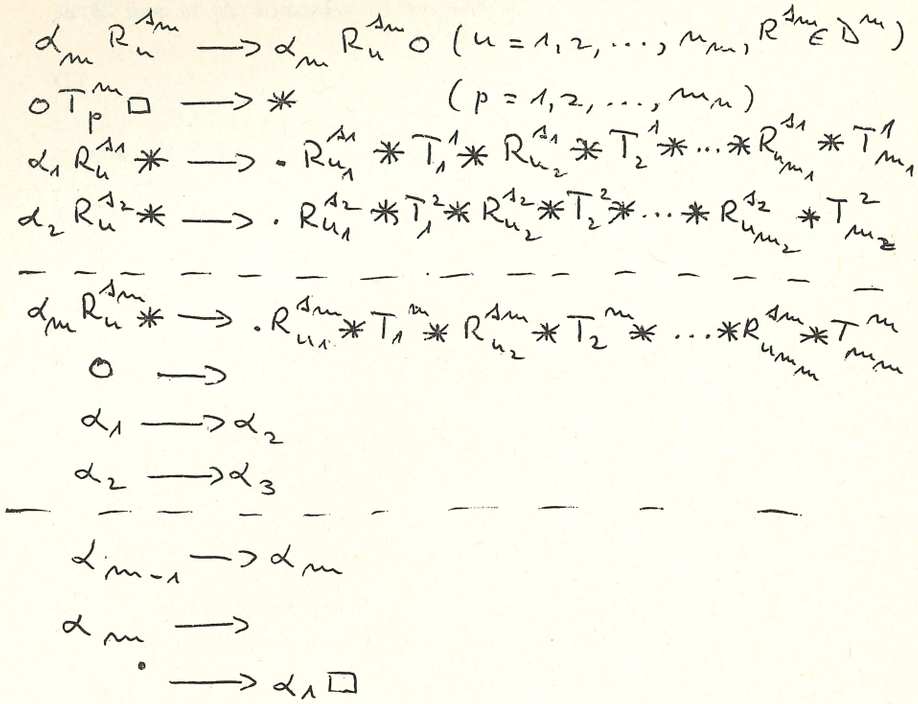
$\square a \longrightarrow a \square \quad (a \in \mathcal{A})$

$\alpha_1 R^{A_1}_u \longrightarrow \alpha_1 R^{A_1}_u O \quad (u = 1, 2, \ldots, m_1, R^{A_1} \in D^1)$

$O T^1_p \square \longrightarrow * \qquad (p = 1, 2, \ldots, m_1)$

$\alpha_2 R^{A_2}_u \longrightarrow \alpha_2 R^{A_2}_u O \quad (u = 1, 2, \ldots, m_2 \; R^{A_2} \in D^2)$

$O T^2_p \square \longrightarrow * \qquad (p = 1, 2, \ldots, m_2)$



On a noté par m_i et n_i ($i = 1, 2, \ldots, m$) le nombre des termes de la suite T^i respectivement R^i.

L'algorithme ci-dessus a été écrit également dans le langage de la calculatrice électronique DACICC-1 et a été expérimenté à propos des substantifs russes. Le programme respectif, contient 320 instructions et a été employé sur un nombre d'à peu près 300 radicaux groupés en 28 types de déclinaisons. Le nombre des radicaux peut être encore agrandi de beaucoup, celui-ci étant limité seulement par la capacité de mémorisation de la calculatrice, le programme étant général.

De plus il peut être adapté aussi à d'autres langues naturelles.

La vitesse relativement grande de la calculatrice assure la réalisation de la déclinaison dans un intervalle de temps assez court de n'importe quel substantif dont le radical se trouve parmi les 300 radicaux mentionnés plus haut. Lorsqu'on donne au programme à décliner un mot erroné il le signalera par l'arrêt de la calculatrice. Tenant compte de ce que l'on a rappelé plus haut, le programme est en état de décliner aussi des substantifs dont les radicaux contiennent des lettres d'une valeur variable.

Imprimé en France
IMPRIMERIE-PHOTOGRAVURE GRAMA - NEVERS
Dépôt légal : 3e trimestre 1968 - N° 9748

ÉDITIONS DU CENTRE NATIONAL
DE LA RECHERCHE SCIENTIFIQUE

15, quai Anatole-France - PARIS 7e
C.C.P. PARIS 9061-11 Tél. : 705 93-39

LE BULLETIN SIGNALÉTIQUE

Revue bibliographique
annonçant tous les travaux scientifiques,
techniques et philosophiques
publiés dans le monde entier

Format in-4° coquille

Section 24 - SCIENCES DU LANGAGE
(parution trimestrielle)

Prix de l'abonnement annuel | France 60,00
Etranger 65,00

T.A. informations est publié par l'A.T.A.L.A. Association pour le développement de la Traduction Automatique et de la Linguistique Appliquée avec le concours des ÉDITIONS KLINCKSIECK qui en assurent l'impression et la diffusion

LA. T.A. DANS LE MONDE / LA T.A. DANS LE MONDE
LA. L.A. DANS LE MONDE / LA L.A. DANS LE MONDE