Eighth Int. Workshop on GNs, Sofia, 26 June 2007, 32-37

GENERALIZED NET MODEL OF THE AUTOMATIC NATURAL LANGUAGE TRANSLATION

Krassimir T. Atanassov[1], Georgi Gluhchev[2] and Anthony G. Shannon[3]

[1] Centre for Biomedical Engineering – Bulgarian Academy of Sciences,
Acad. G. Bonchev Str., Bl. 105, Sofia-1113, BULGARIA
e-mail: krat@bas.bg
[2] Institute of Information Technologies – Bulgarian Academy of Sciences
Acad. G. Bonchev Str., Bl. 29A, Sofia-1113, BULGARIA
e-mail: gluhchev@iinf.bas.bg
[3] KvB Institute of Technology, North Sydney, 2060, AUSTRALIA
and
Warrane College, University of New South Wales, Kensington, 1465, AUSTRALIA,
emails: *tonySHANNON@raffles.kvb.edu.au* and *tony@warrane.unsw.edu.au*

1 Introduction

Up to now we have described by Generalized Nets (GNs; see [1]) some processes of pattern recognition – particularly, speach, handwriting, face, etc (see [2]).

The text recognition and translation is another direction of research that will be illustrated with the next example. The process will be modelled in an abstract form. We will suppose that we have one or more expressions that will be interpreted as ordered sets of words - elements of a fixed language. The aim is to recognize the separate words, to translate them on another language and construct corresponding expressions in the second language. Here we will not discuss some specific aspects of the linguistic translation, related to specific language constructions like idioms, proverbs, slang and others.

2 A GN-model

Here we shall describe a GN-model (see Fig. 1) that translates a given sentence and recognize its sense. Let us have:

• a DB of words of one language and their corresponding words in another language that we shall represent by token β permanently staying in place l_5,

• a DB of the context in which separate words are used, that we shall represent by token γ permanently staying in place l_6,

• a DB of the contexts of separate expressions, that we shall represent by token δ permanently staying in place l_{16}.

Let us have an expression of the first language. We shall represent its words by a set of α-tokens that will enter sequentially place l_1 with initial characteristic (for the i-th token):

$$x_0^{\alpha_i} = \text{“}i\text{-th word of the expression”}.$$

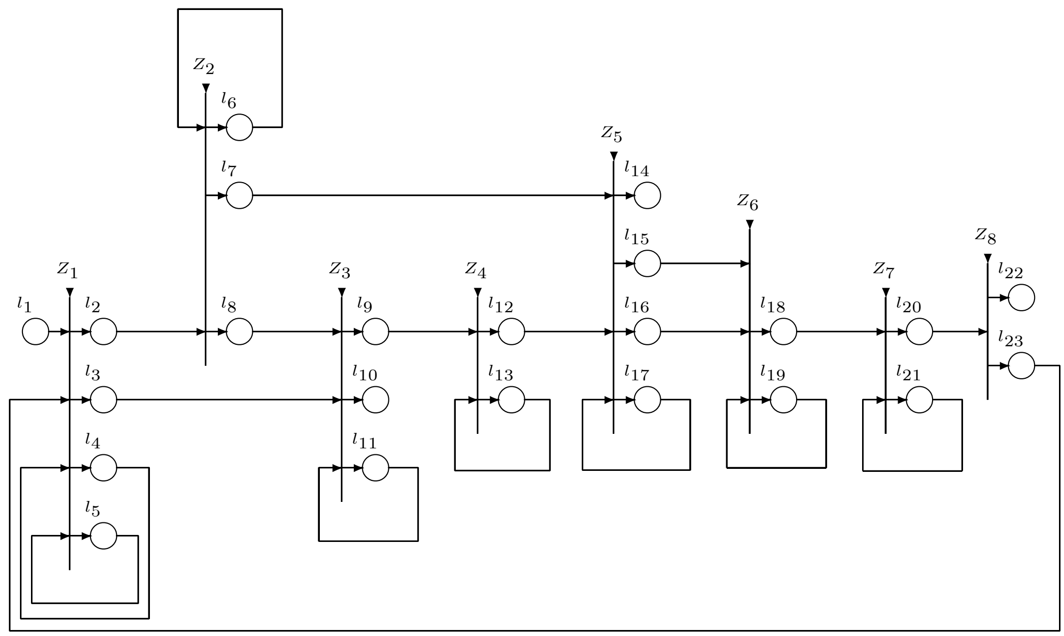

Fig. 1: GN-model

$$Z_1 =< \{l_1, l_4, l_5, l_{23}\}, \{l_2, l_3, l_4, l_5\}, \begin{array}{c|cccc} & l_2 & l_3 & l_4 & l_5 \\ \hline l_1 & W_{1,2} & W_{1,3} & W_{1,4} & false \\ l_4 & W_{4,2} & false & W_{4,4} & false \\ l_5 & false & false & false & true \\ l_{23} & false & false & false & true \end{array} >,$$

where
$W_{1,2} = W_{4,2} =$ “the current word is recognized”,
$W_{1,3} =$ “the current word is not recognized (not included in DB β”,
$W_{1,4} = W_{4,4} =$ “the current word is recognized, but it has more senses than the already used”,
$W_{4,4} = \neg W_{4,3}$.

When token α_i enters place l_2 it obtains a characteristic

“the list of the words from the second language which corresponds to the current word”,

while in places l_3 and l_4 it does not obtain any characteristic. If predicate $W_{1,4}$ is not true, the token keeps its identifier; if predicate $W_{1,4}$ is true, it changes its identifier to $\alpha_{i,1}$.

When predicate $W_{4,4}$ is true, then the current α_i-token split to two tokens - the same token, that already will be marked by $\alpha_{i,j}$ and that will stays in place l_4 and a new token – $\alpha_{i,j+1}$, where $j - 1$ is the number of last time-steps in which there is a token in place l_4.

For simplicity, we shall use α-token instead of $\alpha_{i,j}$-token where possible.

When α-token from place l_{23} enters place l_5 it unites with token β that obtains characteristic

$$x_{cu}^{\beta} = x_{cu-1}^{\beta} \cup x_{cu}^{\alpha}.$$

$$Z_2 =< \{l_2, l_6\}, \{l_6, l_7, l_8\}, \begin{array}{c|ccc} & l_6 & l_7 & l_8 \\ \hline l_2 & false & W_{2,7} & W_{2,8} \\ l_6 & true & false & false \end{array} >,$$

where

$W_{2,7} =$ "there is not a suitable context for the current word",

$W_{2,8} = \neg W_{2,7}$.

When the α-token enters place l_8 it obtains a characteristic

"list of the suitable contexts for the current word".

while in place l_7 it does not obtain any characteristic.

Token γ in place l_6 does not change its (initial) characteristic.

$$Z_3 =< \{l_3, l_8, l_{11}\}, \{l_9, l_{10}, l_{11}\}, \begin{array}{c|ccc} & l_9 & l_{10} & l_{11} \\ \hline l_3 & W_{3,9} & W_{3,10} & W_{3,11} \\ l_8 & true & false & false \\ l_{11} & W_{11,9} & W_{11,10} & W_{11,11} \end{array} >,$$

where

$W_{3,9} = W_{11,9} =$ "the current unrecognized at the moment word is recognized in the context of the already recognized words",

$W_{3,10} = W_{11,10} = \neg W_{3,9}$.

When an α-token enters place l_9 it obtains a characteristic

"the current word from the second language which corresponds contextually

(in the sense of DB γ) to the current word",

while in place l_{10} it does not obtain any characteristic.

$$Z_4 =< \{l_9, l_{13}\}, \{l_{12}, l_{13}\}, \begin{array}{c|cc} & l_{12} & l_{13} \\ \hline l_9 & false & true \\ l_{13} & W_{13,12} & W_{13,13} \end{array} >,$$

where

$W_{13,12} =$ "a token from place l_7 enters place l_{15} and its current characteristic is obtained on the basis of the characteristics of the present token",

$W_{13,13} = \neg W_{13,12}$.

When the first α-token enters place l_{13} it stays there till the moment when all other α-tokens enters this place and unites with the first one. Otherwise, it splits to new α-tokens, that obtain consequtively the characterisrtics:

"$\langle w_{1,1}, w_{2,1}, ..., w_{s,1} \rangle$",

$$“\langle w_{1,2}, w_{2,1}, ..., w_{s,1}\rangle”,$$

$$. \quad . \quad .$$

$$“\langle w_{1,t_1}, w_{2,1}, ..., w_{s,1}\rangle”,$$

$$“\langle w_{1,1}, w_{2,2}, ..., w_{s,1}\rangle”,$$

$$“\langle w_{1,2}, w_{2,2}, ..., w_{s,1}\rangle”,$$

$$. \quad . \quad .$$

$$“\langle w_{1,t_1}, w_{2,2}, ..., w_{s,1}\rangle”,$$

$$. \quad . \quad .$$

$$“\langle w_{1,t_1}, w_{2,t_2}, ..., w_{s,1}\rangle”,$$

$$“\langle w_{1,t_1}, w_{2,t_2}, ..., w_{s,2}\rangle”,$$

$$. \quad . \quad .$$

$$“\langle w_{1,t_1}, w_{2,t_2}, ..., w_{s,t_s}\rangle”,$$

where $w_{i,j}$ is the j-th sense of the i-th word in the second language, or $w_{i,j} = *$, if this word of the first language is not recognized. The α-tokens do not obtain any characteristic in place l_{12}.

$$Z_5 =< \{l_7, l_{12}, l_{17}\}, \{l_{14}, l_{15}, l_{16}, l_{17}\}, \begin{array}{c|cccc} & l_{14} & l_{15} & l_{16} & l_{17} \\ \hline l_7 & W_{7,14} & W_{7,15} & false & false \\ l_{12} & false & false & true & false \\ l_{17} & false & false & false & true \end{array} >,$$

where
$W_{7,14} =$ “a suitable context for the current unrecognized word is not found on the basis of DB δ and the tokens from places l_{12}, l_{13}, l_{16} and l_{19}”,
$W_{7,14} = \neg W_{7,16}$.

When the α-token enters place l_{15} it obtains a characteristic

“the word from the second language which corresponds contextually (in the sense of DB δ)

to already recognized words”,

while in place l_{14} it does not obtain any characteristic.

The α-tokens from place l_{12} enter place l_{16} without any new characteristics.

Token δ in place l_{17} does not change its (initial) characteristic.

$$Z_6 =< \{l_{15}, l_{16}, l_{19}\}, \{l_{18}, l_{19}\}, \begin{array}{c|cc} & l_{18} & l_{19} \\ \hline l_{15} & false & true \\ l_{16} & false & true \\ l_{19} & W_{19,18} & W_{19,19} \end{array} >,$$

where
$W_{19,18} =$ “all entered α-tokens are collected in place l_{19},
$W_{19,19} = \neg W_{19,18}$.

When the α-tokens from place l_{16} enter place l_{19} they stay there till the moment when all α-tokens from place l_{15} enter place l_{19} and unite with the respective first ones, that consequtive obtain the characterisrtic

"a sequence of all recognized words (the places of unrecognized words are marked

by symbol $*$)",

while in place l_{18} they do not obtain any characteristic.

$$Z_7 =< \{l_{18}, l_{21}\}, \{l_{20}, l_{21}\}, \begin{array}{c|cc} & l_{20} & l_{21} \\ \hline l_{18} & true & false \\ l_{21} & false & true \end{array} > .$$

Token ε staying permanently in place l_{20} symbolizes a checker of the translated sentence sense. This token can be represented by a subnet by the hierarchical operator $\mathcal{H}_6$ (see [3]). A similar result can be obtained if we use hierarchical operator $\mathcal{H}_1$ (see [1]) that will replace l_{20} by a subnet. Token ε does not obtain any characteristic, while each token α in place l_{20} obtains the characterisrtic

"the final sequence of all recognized words (the places of unrecognized words are marked

by symbol $*$) and an estimation of its correctness".

$$Z_8 =< \{l_{20}\}, \{l_{22}, l_{23}\}, \begin{array}{c|cc} & l_{22} & l_{23} \\ \hline l_{20} & true & W_{20,23} \end{array} >,$$

where
$W_{20,23} =$ "there are recognized words that are not included in DB β".

When each one of the α-tokens has in its current characteristic words that are not included in DB β, then it splits into two tokens: one of them leaves the net through place l_{22} without a new characteristic and the other enters place l_{23} with a characteristic

"list of the words that are not included in DB β".

3 Conclusion

The above constructed GN-model is the first one, related to the GN-approach in natural language processing. Up to now there is only one GN-model, representing the functioning and the results of the work of a semantic network [4].

In the future this model will be extended in some directions. For example, the GN for semantic networks can be included as a subnet. Also, the processed related to recognition and translation of idioms will be described as a separate subnet of the above one.

References

[1] Atanassov, K. Generalized Nets, World Scientific. Singapore, New Jersey, London, 1991.

[2] Atanassov K., G. Gluhchev, S. Hadjitodorov, A. Shannon, V. Vassilev. Generalized Nets and Pattern Recognition. KvB Visual Concepts Pty Ltd, Monograph No. 6, Sydney, 2003.

[3] Atanassov, K. On a new hierarchical operator over the generalized nets. Issues in Intuitionistic Fuzzy Sets and Generalized Nets, Vol. 3, 2006, 29-34.

[4] Nikolov N., Generalized nets and semantic networks, Advances in Modelling & Analysis, AMSE Press, Vol. 27, 1995, No 1, 19-25.