Synergetics and 'Insight' Strategy for Speech Processing

YU. KOSAREV
Institute for Informatics and Automation of the Russian Academy of Sciences

R. PIOTROWSKI
Hertzen State Pedagogic University of Russia, St Petersburg, Russia

Abstract

The methodology and developmental technology were created for a speech understanding system based on synergetics and semantic–pragmatics ideas. A real speech processing system GOLOS-2 was then designed as a modular assembly whose modules correspond to similar levels of man's verbal and mental processes. Its behaviour is regulated by a through understanding 'insight' procedure that functions as a computer analogue for a pragmatic–communicative operator of human verbal/mental activity.

1. Introduction

Modern civilization is stimulating language engineering to consider the problems of speech processing. One might even talk in terms of a 'conversational computing explosion', which is a direct result of a universal informational explosion and of multilingual communication demands. In these conditions, business is booming in the language engineering technology market, with the most commercially viable speech recognition systems being DragonDictate for Windows 2.0, Macintosh PowerSecretary, IBM Voice-Type Dictation etc. (Language Industry Monitor, 1995). Therefore, such systems deal with acoustic–phonetic and partially linguostatistical analyses as well as with lexical analysis and synthesis of speech. However, text entropy measurements and psycholinguistic research in man–machine interaction have shown the bulk of the text information to be provided by lexical units and by contextual (semantic–syntactic and pragmatic) relationships in a sentence or in an entire text. Potential (statistic) and syntactic information, which imposes combinatorial and quantitative constraints on the text frequency and combinability of letters, phonemes, and distinctive features, contributes moderately to the recognition of utterance content (Piotrowski, 1984, pp. 219–245, 255–265; 1986, pp. 36–40; Kosarev, 1989, 1994).

On the other hand, unlike normalized printed text, our speech is full of grammatical errors, incomplete sentences, and words, and it is constantly disrupted by repetitions and hesitations. However, disregarding the defects, we gather the utterance meaning using knowledge based on our language and life realities.

Hence, it follows that the progress made in developing computer speech analysis and in going from recognition to speech understanding (SU) has been due to high-level, i.e. semantic–syntactic and pragmatic, speech processing based on the use of advanced language engineering technologies (Zue *et al.*, 1990; Seneff, 1992; Prieto *et al.*, 1994; Smith *et al.*, 1995; Waibel 1996).

The elaboration of high-level speech processing strategies and techniques involves a choice between the following two philosophies of languages. One of them, dating back to Plato, Leibnitz, Hjelmslev, and Chomsky, holds that language is a closed logical system (ἔργον) and speech utterance is generated and understood on the basis of a propositional calculus. According to the second approach, language and speech are considered a fuzzy dynamic mechanism—an ἐνέργεια (von Humboldt, 1907, pp. 44–48; Baudouin de Courtenay, 1903; Zadeh, 1976) rather than a static logical subject.

The entire experience gained in the domain of language engineering suggests that NLP models developed on the basis of the strict rules of the first philosophy are unable to explain or indeed resolve many of the paradoxes of human mental/verbal activity and man–machine interaction (Piotrowski 1984, pp. 47–53; Kosarev, 1995, pp. 1211–1214). Neither can they account for antinomies involved in SU and NLP such as:

- how to get reliable sentence from fuzzy, less reliable units;
- how to get a correct decision from inexact or incorrect hypotheses set;
- how to get the best decision from a correct set of equivalent hypotheses.

However, using some mysterious 'demons' of self-regulation and self-organization (Koehler, 1990, 1992; Hřeßiček, 1995, pp. 12–14) hidden in our subconsciousness, a man, in the course of his everyday mental/verbal activity and communication, solves these problems advantageously. The challenge now is to reveal these mysterious synergetic mechanisms and to simulate them on the computer.

2. Conception

As a consequence of the use of inadequate basic postulates based on the "ἔργον" approach, researchers have been unable to generate a single stably working NLP-MT and, particularly not an SU-MT system.

All things considered, we should abandon the traditional logical approach to spoken language understanding and use some psycholinguistic synergetic and

Correspondence: R. Piotrowski, apt. 6, 9 ave. M. Thorez, St Petersburg, 194021 Russia. E-mail: rgp@lel.emissia.spb.su

Literary and Linguistic Computing, Vol. 12, No. 2, 1997

stochastic ideas of the ἐνέργεια philosophy. First and foremost, we should proceed from Saussure's assumption that every linguistic unit is a multidimensional system entering in a number of paradigms and associative series simultaneously (de Saussure, 1983, Part II, Chapter V, p.3; Chapter VI, p.2).

Therefore, a working SU system should be designed as a modular assembly, where every acoustic, semantic–syntactic, or pragmatic program module corresponds to a similar level of human speech and text processing. However, the construction of such a modular system involves some complications. The well-known language engineering principle of level-by-level text processing generates an enormous number of incorrect solutions, increasing the uncertainty within the SU system. Hence, a new paradox crops up: the more sources of information we consider, the worse SU quality we get. This difficulty may be obviated with the aid of a through understanding procedure which may serve as a synergetics look-alike for the communicative pragmatic operator of human mental/verbal activity (Piotrowski and Tambovtsev 1994, pp. 291–293; Piotrowski 1996, pp. 86–89). This procedure consists of choosing an optimal hypothesis in accordance with integral estimation worked out by the SU system on the basis of the sender's/receiver's world model with its own thesaurus vocabulary pre-loaded into the system database (see below).

3. SU System Architecture

Each SU system is a complex which should be described through a multiaspect representation. Two description strategies are presented below: a structural/functional and a synergetic (more precisely, a management/decision) one.

3.1 *Structural/functional representation of the SU system*

This type of representation disregards the physical substrate of the SU system and presents it as a hierarchy of three levels.

3.1.1 *Acoustic–lexical level.* At this basic level, special modifications of traditional algorithms of autocorrelation, vector quantitization, dynamic programming, and word selection are used (Kosarev, 1989, pp. 37–59, 1994, pp. 1281–1284). Using the symbol S^* for signal, these procedures give

$$S^* = S^*_1, S^*_2, \ldots, S^*_i, \ldots, S^*_L,$$

where L is number of words in the utterance.

By applying the dynamic programming (or HMM method), each S^*_i can be mapped one-to-one into the subset W^*_i which includes the most probable hypothetical wordforms (w/f) from the vocabulary V, each hypothesis being a pretender to embodying s^*_i. Thus, the S^* is transformed into a w/f set sequence

$$W^* = W^*_1, W^*_2, \ldots, W^*_i, \ldots, W^*_L.$$

In the long run, the incoming signal is mapped onto a input sentence hypothesis set:

$$F^* = \prod_{i=1}^{L} W^*_i = \{f_n = W_{n_1}, W_{n_2}, \ldots, W_{n_L} \mid W_{n_i} \in W^*_i, i = \overline{1, L}\},$$

and
$$|F^*| = \prod_{i=1}^{L} |W^*_i|.$$

Each hypothesis has its own acoustic estimate:

$$E_{ak} = \frac{1}{L} \sum_{i=1}^{L} C(S^*_i, e_i), i = \overline{1, L},$$

where C is DP distance between the signal S^*_i and a word sample from an acoustic–lexical database. For a more detailed discussion, see Kosarev (1994, pp. 1281–1284) and Biermann *et al.* (1992).

3.1.2 *Syntactic–semantic associative level.* Associative analysis of phrases–hypotheses is based on the following pre-requisites.

1. Semantic–syntactic knowledge is realizing in human consciousness and subconsciousness by association mechanisms (Lyons 1972, 2.2.1, 2.3.3, 9.4.4; cf. Oaksford and Chater, 1991),
2. The association connection between two wordforms can be evaluated through the use of binominal word-combination statistics (Gorodeckij *et al.*, 1971; Danejko *et al.*, 1973; Kravez, 1973; Gustaffson, 1975) or by using expert estimates (Howes, 1957; Deese, 1962, pp. 161–175; Leontiev, 1977).
3. The connection degree within a binomial word-combination can be used as a quantitative assessment of its comprehensiveness and sensibleness (Kosarev and Jarov 1995).

Now, let

$$W = \{W_1, W_2, \ldots, W_g, \ldots, W_N\}, \quad g = \overline{1, N}.$$

Be a vocabulary, where for each ordered pair of wordforms (W_g, W_h) we set, statistically or expertly, a cost factor a_{gh} evaluates the semantic–syntactic connection between W_g and W_h. The cost factor values are adjusted so that the ultimate result will be equal to zero in the ideal case and will increase with the decay of the W_g and W_h connection. Indeed, such a connection in the word pair *request clearance* from Airspeak language is stronger than that in the binomial combination *Riga route.* Note that $a \geqslant 0$, $a_{gh} \in [a_{min}, a_{max}]$ and generally $a_{gh} \neq a_{hg}$. As a result, we obtain a vocabulary connectivity matrix

$$A_{\{N, N\}} = \|a_{gh}\|.$$

The matrix is built up statistically or by experts in accordance with a four-point scale: low (3), medium (2), high (1), very high (0) (cf. Zadeh, 1976, pp. 254–258).

Then, we take an arbitrary wordform sequence (utterance) with length L

$$f_n = W_{n_1}, W_{n_2}, \ldots, W_{n_i}, \ldots, W_{n_L}.$$

Next we extract from A a subset A^* that would involve all (W_g, W_h) pair factors ordered according to g:

$$A^* = \{a_{n_1, n_2}, a_{n_1, n_3}, \ldots, a_{n_2, n_3}, a_{n_2, n_4}, \ldots, a_{n_{L-1}}, n_L\},$$
$$|A^*| = C^2_L.$$

As a result, the syntactic–semantic associative estimate for the utterance f_n may be expressed as a normalized sum

$$E_{ass}(n) = \frac{1}{C^2_L} \sum_{n_k=1}^{L-1} \sum_{n_s=n_k+1}^{L} a_{n_k,n_s}, \quad k < s$$

3.1.3 *Pragmatics level.* It is known that complete speech understanding is possible through consideration of a sufficiently wide situation context in concert with sender's/receiver's life knowledge or their professional pragmatics. In our SU system, the pragmatics component is to be realized by means of 'soft' quantitative comparison of input utterance hypotheses with canonic templet sentences, each of which corresponds one-to-one to a certain act within the framework of a situation of man's activity (Bekwith *et al.*, 1992). As a result of this comparison, each hypothesis takes a quantitative pragmatic estimate E_{pr} (see below) that is essential for integral estimation E (see below) of the input sentence semantic interpretation. From the synergetic standpoint, the pragmatic level is the most important SU stratum. It is at this level where the majority of decision/management and control operations are performed. Therefore, it makes sence to describe pragmatic procedures and technique under the synergetic representation.

3.2 *Synergetic (management/decision) representation of the SU system*

An SU system involves recognition operations which are performed under uncertainty presented in the input signal, as well as in algorithmic blocks of higher levels by a set of versions, from among which the SU system selects purposefully the optimal decision. That is the reason why the SU model should be described in synergetic terms. As was shown, the SU process is to be represented as a mapping of input signal sequence S^* (source sentence) first into a space of syntactic–semantic hypotheses A and next into a set of their pragmatic estimates.

Taking into consideration that the number of situations in any domain is finite and that there are no infinitely long sentences describing them, the pragmatics processing procedure is achieved on a limited domain model represented in the form of the known state diagram as an oriented graph. Its arches are transitions from situation to situation, each arch connected with a subset of equivalent sentences signifying a definite speech intention or a concrete command. For instance, in tower–aircraft talks regarding the situation 'Request emergency landing', two alternative acts are allowable: 'Cleared to land' or 'Do not land'.

As may be seen from Fig. 1, we obtain estimates of correspondence between the input hypothesis and a concrete act in the framework of a current situation in the domain under consideration, along with a command corresponding to this act. In other words some of man's activity model contain a limited set of situations:

$$\mathrm{St} = \{\mathrm{St}_1, \mathrm{St}_2, \ldots, \mathrm{St}_b, \ldots, \mathrm{St}_B\},$$

where B is a number of situations.

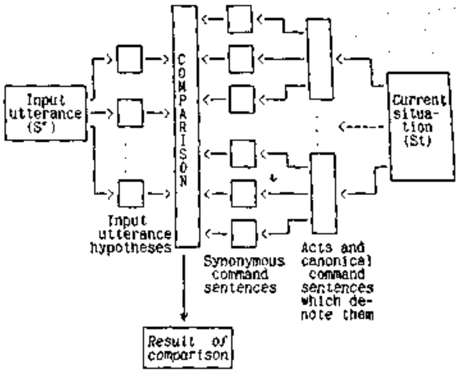

Fig. 1 Pragmatics estimation of input utterance hypotheses.

For each St_b there exists a subset of acts; each act j is represented by a canonical command K_{bj}:

$$\mathrm{St}_b \rightarrow K_b = \{K_{b1}, K_{b2}, \ldots, K_{bj}, \ldots, K_{BJ}\}.$$

Similarly, for each K_{bj}, a certain subset of synonymous sentences is specified or generated

$$K_{bj} \rightarrow \{K_{bj1}, K_{bj2}, \ldots K_{bjr}, \ldots K_{bjR}\}.$$

The challenge now is to estimate the semantic distance between input sentence hypothesis $f_n = H$ and a canonical variant $K_{bjr} = K$. Let us suppose that this estimate will be based on the following assumptions.

- H and K sentences can be treated as subsets of wordforms but not as tuples (i.e. sequences of wordforms); the problem is that workshop slang (Kosarev and Kulakov, 1994) makes wide use of syntactic variation for the short command sentences, cf. in Airspeak language: *request clearance to taxi*, or *request taxi clearance*, or *taxi clearance*, similarly rus. прошу разрешения на руление, прошу разрешения рулить, разрешения рулить прошу, разрешите рулить, рулить разрешите, etc.
- Each wordform W_i from the command sentence has its own semantic weight V_i which may be evaluated using expert estimates.
- the sum of the all wordform weights for each phrase is constant:

$$\sum_{i=1}^{L} V_i = \mathrm{const.}$$

It is then possible to present a command sentence K as a non-regulated set of pairs <wordform, its weight>:

$$K \rightarrow \{<W_1, V_1>, <W_2, V_2>, \ldots <W_i, V_i>, \ldots <W_L, V_L>\}, \quad L = |K|.$$

It is practically impossible to evaluate a weight for each wordform of an unexpected input utterance. Therefore, we write, for simplicity, the hypothesis about input utterance as

$$H = \{W_1, W_2, \ldots W_1, \ldots W_M\}, \quad M = |H|.$$

Now consider semantic discrepancy between the sets K and H, which depends on concrete lexical contents of K and H, as well as on L, M, and V_i. Let us assess quantitatively this distance using set differences $A_1 = H \backslash K$ and $A_2 = K \backslash H$, and the intersection $A_3 = K \cap H$. (see Fig. 2). It is easy to verify that the absolute value of the above discrepancy increases with growth of $|A_1|$, $|A_2|$ and the weights sum of wordforms from the command sentence. If we normalize our function on phrases lengths with the factor

$$\frac{1}{L+M}$$

and introduce expert weight coefficients p_1 and p_2, then the formula for the estimation of pragmatic discrepancy between an input utterance and some variant of canonical sentence becomes

$$D(H, K) = \frac{p_1 |A_1| + p_2 |A_2|}{L+M} \Big(\sum_{i \in A_3} V_i + 1 \Big).$$

Hence, it follows that the minimal value of $D(H, K)$ shows an optimal meaning correspondence of the source signal S^* with a canonical sentence K_{ijr} from the set of commands K. These commands are considered as models for input utterances within the limits of the domain in question. As a result, we obtain a pragmatics estimate

$$E_{pr} = \min_{n,j,r} D\{H_n, K_{bjr}\}$$

for input signal S^*

An integrative evaluation of hypotheses about the meaning of input utterance is as a weighed sum of partial estimates

$$E = [\alpha_1 E_{ac}^2 + \alpha_2 E_{ass}^2 + \alpha_3 E_{pr}^2(j)]^{1/2},$$

where α_1, α_2, and α_3 are weight coefficients and j denotes a particular act in the framework of the situation in hand. E shows an integral deviation measure for each hypothesis. A hypothesis with the minimal value of E is defined as a final 'insight' decision in the process recognition and understanding of the input signal. Thus the filtering estimate procedure is an analogue of a communicative–pragmatic operator controlling human mental/verbal activity. The reciprocal of the averaged data for all signal estimates $[1/(E+1)]$ may be used as a quantitative performance for the synergetics organization of the SU system (see Haken, 1978).

4. Experimental Procedures. Speech Understanding System GOLOS-2

4.1 *Speech material and talkers, hardware, and software*

In order to evaluate our SU algorithm implemented by GOLOS-2 hardware, about 2,000 Russian command sentences were used. These utterances were essentially non-expanded simple sentences taken from tower–aircraft and nuclear power station talks. Notice that not only canonical commands were suggested to the system but also utterances with different deviations, such as word sequence disorder, the replacement of one of the wordforms by a synonym unknown to the system, the omission of a non-key wordform, and the addition of some parasitical lexical units like пожалуйста 'please', TAK 'well', etc. In most cases, the system neutralized such violations successfully, giving the impression of real understanding of an input text by the GOLOS-2 system.

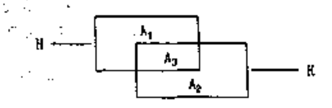

Fig. 2 Semantic discrepancy between a hypothesis about the input utterance and a synonymous canonical command sentence.

The speech material was read by two male and two female speakers. These subjects suffer from no speech or hearing disorders. Their dialect is typical of St. Petersburg Russian. They had received no training in phonetics and were completely unaware of the possible hypotheses that the experiment had been designed to test.

The hardware platform is based on a 486-processor, is PC-compatible, and includes a special Sound Card which incorporates a spectrographic device (ten-channel spectrograph), based on TMS-320 processors and synthesizer. The SU system requires about 1.5 Mb disk space and works under MS-DOS. The software is implemented on Turbo C (Borland Inc.).

4.2 *Experimental results*

As can be seen from Table 1, the accuracy of sentences–commands understanding by application of integral meaning interpretation is much better than isolated word and utterance recognition executed without recourse to these procedures.

5. Some Applications

The above-discussed architecture exhibits a number of behavioural advantages for efficient text processing and human–machine dialog. First and foremost these behaviours can be used

- in designing systems for the detection and correction of spelling errors in scientific and scholarly text (see Fink and Biermann, 1986);
- for developing automatic speech translation systems.

5.1 *Semantic–syntactic and pragmatics help for automatic text correction*

By replacing the analysing and synthesizing modules of the Sound Card with a scanner and an orthographic

Table 1 The relationship between recognition and understanding of input signal (word or utterance) by the GOLOS-2 system

Speakers	Recognition accuracy of isolated words (P_1)	Utterance accuracy recognition without semantic–syntactic and pragmatic interpretation (P_2)	Understanding of input utterances (sentence–command) (P_3)
1	0.90	0.730	0.946
2	0.92	0.770	0.969
3	0.95	0.860	0.990
4	0.98	0.994	0.999

Note that the values P_1 and P_3 were obtained using machine experimental data; the value P_2 results as $P_2 = P_1^L$, where L is a mean input utterance length (P_2 was also tested experimentally).

synthesizer respectively, the GOLOS-2 SU system may be converted into a semantic-syntactic and pragmatic speller whose scheme is represented graphically in Fig. 3.

It is easy to see that the initial processing is analogous to the spectrographic analysis module in our SU system. It consists of scanning the input text and transforming it into a chain of ASCII codes which reflect only graphical processing results. Such a text contains lots of errors of spelling and grammar and uncertainties like erroneous insertion, transposition, and erasing of characters. Therefore, the module for spelling analysis generates some wordform hypotheses for each grapheme chain in accordance with its spelling resemblance to vocabulary words and in conformity with morphological rules.

At semantic–syntactical and pragmatical levels, some sentence hypotheses from available wordform hypotheses sets are constructed. Each hypothesis should be evaluated in the manner described above for oral utterance (see Sections 3.1.2, 3.1.3, and 3.2).

In the decision-making block, all spelling, semantic–syntactic, and pragmatic analysis results are summed over their evaluations (see Section 3.2). In accordance with this integral estimate, a final decision about the optimum input sentence hypothesis is made. In this way, the system performs the necessary spelling and grammar text corrections.

An experimental model of the above-discussed system was tested on a sample of about 200 English sentences from the domain 'A man in the city' composed from wordforms taken from a limited vocabulary of about fifty words. Between 10 and 20% wordforms of the sample were distorted by random substitution, insertion, extraction, and transposition of the characters. Once the text sample had been put into the system, it corrected 99% of distorted wordforms. As an example, we refer to the distorted utterance *Hof can I get tthe senter?* The system has processed and transformed it into a correct sentence *How can I get to the center?*

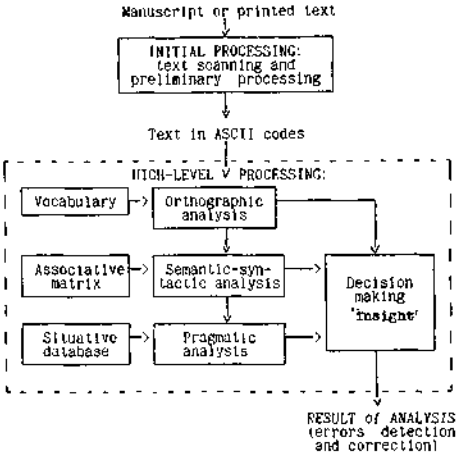

Fig. 3 The scheme for a system for correction of errors in hand- or typewritten text.

Using semantic–syntactic and pragmatic information, the model is sometimes able to eliminate some mistakes from previous processing and to substitute an erroneously perceived word by another correct lexeme, cf. *He is table . . . He is able.*

5.2 *Machine speech translation*

The GOLOS-2 system can also be used as an input module of a multilingual linguistic automation (LINGTON) applied to machine speech translation that will become increasingly important with advances in communicative technology and with the necessity of overcoming language barriers (see Stentiford and Steer, 1988, Rayner *et al.*, 1993, Kitano, 1994).

An indispensable condition for correct input utterance translation and its conversion to faithful oral output is the adequacy of the sender's and receiver's world models. That is to say that all the situations and acts from the first world model must have semantic analogues in the second one. All the acts must also possess adequate linguistic descriptions in both models. The latter can be present in two forms, as a sufficiently representative subset of equivalent sentences or by application of a generative procedure.

Being a translating block for the whole system, the LINGTON should comply with the following requirements.

1. To be multifunctional, i.e. to be able to achieve a variety of behaviours such as machine translation, indexing, annotation, abstracting of a source text, and man–machine dialogue,
2. To allow for further development and improvement, by adopting the LINGTON to the communication informational evolution of society and to the changing pragmatic outlook of the actual users of information,
3. To process an unbuilt ability to preserve its most essential properties in case of failure, caused by viruses, RAM breakdowns, distortion of words, etc.

6. Concluding Remarks

To summarize, two points can be made about the works that have been analysed in this paper.

First, we have tried to show how to integrate knowledge of different nature to design a robust working speech dialogue system which can form the input and output parts of a linguistic automaton simulating mental/verbal human behaviour. The next long-term aim of our work is to construct a model of communicative interaction that will be able to support the negotiation of meaning tracking a dialogue topic in a task-oriented domain and its knowledge database.

Second, oral text translating or abstracting, even when it is carried out by man, cannot always be perfect, and it would be unfair to expect a computer to do better. However, as language engineering technologies improve and more is understood about how the oral or written texts are recognized and synthesized, so more sophisticated methods can be used in machine translation and abstracting to convey the speaker's intention.

References

Baudouin de Courtenay, J. (1903). Opsychicznych podstawach zjawisk językowych. Postawienie kwestii. *Przegląd Filozoficzny*, Vol. VI, Warszawa.

Bekwith, R., Fellbaum, C., Gross, D., and Miller, G. A. (1992). WordNet: a lexical database organised on psycholinguistic principles. In *Using On-line Resources to Build a Lexicon*. Zernic, U. (ed.), Lawrence Erlbaum, Hillsdale, NJ.

Biermann, A. W., Fineman, L., and Heidlage, J. F. (1992). A voice- and touch-driven natural editor and its performances. *International Journal of Man–Machine Studies*. 37: 1–21.

Danejko, M. E., Maschkina, L. E., Nechaj, O. A., Sorkina, W. A., and Scharanda, A. N. (1973). Statistische Untersuchung der lexikalischen Distribution der Wortformen. In *Sprachstatistik. Mit zahlreichen Skizzen, Tabellen und Schemata im Text. Uebersetzt von einem Kollektiv unter Leitung von Lothar Hoffman*. Wilhelm Fink, Muenchen/Salzburg, pp. 239–251.

Deese, J. (1962). On the structure of associative meaning. *Psychological Review*. 69: 161–175.

de Saussure, F. (1983). *Course in general linguistics*. Translated and annoted by R. Harris. Duckworth, London.

Fink, P. E. and Biermann, A. W. (1986). The correction of ill-formed input using history-based expectation with application to speech understanding. *Computational Linguistics*. 12: 13–36.

Gorodeckij, B. Yu., Kibrik, A. E., Logahina, L. C., Maksimova, G. V., Prytkov, E. S., and Raskin, V. V. (1971). *Slovari slovosočetanij i častotnyje slovari slov ograničennogo voennogo podjazyka* (Word-combination dictionaries and word frequency dictionaries of a limited military sublanguage). University Press, Moscow.

Gustaffson, M. (1975). Binominal expressions in present English. A syntactic and semantic study. Turku.

Haken, H. (1978). *Synergetics. An Introduction*. Springer-Verlag, Berlin.

Howes, D. (1957). On the relation between the probability of a word as an association and in general verbal usage. *Journal of Abnormal and Social Psychology*. 54: 75–85.

Hřeßíček, L. (1995). *Text levels. Language constructs, constituents and the Menzerath-Altman law*. Wissenschaftlicher Verlag Trier, Trier.

Kitano, H. (1994). *Speech-to-speech Translation: A Massively Parallel Memory-based Approach*. Kluwer Academic Publisher, Boston, MA.

Koehler, R. (1990). Elemente der synergetischen Linguistik. In Hammer, R. (ed.), *Glottometrika 12*. N. Brockmeyer, Bochum, pp. 179–187.

Koehler, R. (1992). Self-regulation and self-organisation in language. In *What is Language Synergetics?* Saukkonen, P. (ed.), University Printing Center, Oulu, pp. 17–16.

Kosarev, Yu. A. (1989). *Estestvennaja forma dialoga s EVM* (A natural form of dialog with computer). Masinostroenie, Leningrad.

Kosarev, Yu. A. (1994). The model of oral speech semantic interpretation: quantitative pocessing and integration of acoustic, syntactic, semantic and pragmatic data. In *Proceedings of German Acoustics Conference, DAGA-94*. Dresden, pp. 1281–1284.

Kosarev, Yu. A. (1995). Communicative language model: structure and functioning. In *Proceedings of the 4th European Conference on Speech Communication and Technology, EUROSPEECH'95*. Madrid, pp. 1211–1214.

Kosarev, Yu. A. and Jarov, P. A. (1995). Associations help to recognize words. In *Proceedings of German Acoustics Conference, DAGA-95*. Saarbruecken, pp. 979–982.

Kosarev, Yu. A. and Kulakov, F. M. (1994). The model of robot speech operation: integration of signal processing and application area modelling. In *The 3rd International Conference on Automation, Robotics and Computer Vision, ICARCV'94*. Singapore, pp. 1389–1391.

Kravez, L. G. (1973). Quantitative Merkmale englischer Nominalverbindungen. In *Sprachstatistik. Mit zahlreichen Skizzen, Tabellen und Schemata im Text. Uebersetzt von einem Kollektiv unter Leitung von Lothar Hoffman*. Wilhelm Fink, Muenchen/Salzburg, 252–264.

Language Industry Monitor. (1995). The world of natural language computing. Issue no. 26.

Leontiev, A. A., ed. (1977). *Slovar' Associativnych Norm Russkogo Jazyka* (Associations Dictionary of Russian). Moscow University press.

Lyons, J. (1972). *Introduction to Theoretical Linguistics*. Cambridge University Press.

Oaksford, M. and Chater, N. (1991). Against logicist cognitive science. *Mind and Language*. 6: 2–37.

Piotrowski, R. (1984). *Text—Computer—Mensch*. N. Brockmeyer, Bochum.

Piotrowski, R. (1986). Text processing in the Leningrad research group 'speech statistics'—theory, results, outlook. *Literary and Linguistic Computing*. 1: 36–40.

Piotrowski, R. (1996). Machine translation in new Russia. In Tufiş, D. (ed.), *Limbaj şi Tehnologie*. Editura Academiei Române, Bucureşti, pp. 85–92.

Piotrowski, R. and Tambovtsev Yu. (1994). Development of a linguistic automaton on the basis of statistics of speech. *Literary and Linguistic Computing*. 9: 291–294.

Prieto, N., Sanchis, E., and Palmero, L. (1994). Continuous speech understanding based on automatic learning of acoustic and semantic models. In *Proceedings of the International Conference on Speech and Language Processing*. Vol. 4, pp. 2175-2178.

Rayner, M., Bretan, I., Carter, D., *et al.* (1993). Spoken language translation with mid-90's technology. In *Third European Conference on Speech Communication and Technology (Eurospeech '90)*. Berlin, pp. 1299–1302.

Seneff, S. (1992). TINA: a natural language system for spoken language application. *Computational Linguistics*, 18: 61–86.

Smith, R. W., Hipp, D. R., and Biermann, A. W. (1995) An architecture for voice dialogue systems based on prolog-style theorem proving. *Computational Linguistics*, 21: 281–320.

Stentiford, F. W. M. and Steer, M. G. (1988) Machine translation of speech. *British Telecom Journal*, 6:

Von Humboldt, W. (1907). Ueber die Verschiedenheit des menschlichen Sprachßoaues und ihren Einfluss auf geistige Entwicklung des Menschengeschlecht. In Leitzmann, A., Bebhardt, B., and Richter, W. (eds), Gesammelte Schriften. Bd. VII. Koenigl. Preussische Akademie der Wissenchaften, Berlin.

Waibel, A. (1996). Interactive translation of conversational speech. *Computer. Innovative Technology for Computer Professionals*, 29: 41–48.

Zadeh, L. (1976). A fuzzy-algorithmic approach to the definition of complex or imprecise concepts. *International Journal of Man–Machine Studies*, 8: 249–291.

Zue, V., Glass, J., Goodine, D., Leung, H., Phillips, M., Polifroni, J., and Seneff, S. (1990). The VOYAGER speech understanding system: preliminary development and evaluation. *IEEE International Conference on Acoustics, Speech and Signal Processing*. Albuquerque, NM, pp. 73–78.