INFORMATION STRUCTURING IN NATURAL LANGUAGE COMMUNICATION: SYNTACTICAL APPROACH

Andrzej Bargiela, Wladyslaw Homenda***

* Department of Computing and Mathematics, The Nottingham Trent University, Nottingham NG1 4BU, England, (e-mail: andre@doc.ntu.ac.uk)

**Faculty of Mathematics and Information Science Warsaw University of Technology, 00-661 Warsaw, Poland, (e-mail: homenda@mini.pw.edu.pl)

ABSTRACT

This paper introduces a new framework for processing Natural Language statements. The parallel is drawn between the Natural Language processing and the Data Mining technology of information granulation. The formalism affords consistent representation of a well-known phenomenon of ‘approximate’ grammatical correctness of Natural Language statements. The approach is validated on some simple Natural Language statements and the directions for the future development of the system are outlined.

1. INTRODUCTION

Fundamentals of fuzzy sets were formulated on the basis of max and min operators applied to membership functions [14]. These operators were, then, generalized to triangular norms. In both theory and applications the concept of triangular norms borrowed from [10] play important role. They are widely utilized in many areas, e.g. logic, set theory, reasoning, data aggregation, etc. To satisfy practical requirements, besides triangular norms, new operators were proposed and developed including new kinds of operators.

One reason for which IE is of significant research interest is that it provides a basic reference for comparing different natural language processing technologies. However, a more fundamental reason is that IE focuses on the essence of intelligent information processing, that of formation of abstractions. In this sense IE parallels the endeavors of Data Mining that is primarily motivated by ‘making sense of data’.

The rich track record of data mining research provides a valuable insight into the methodologies that lead to comprehensive and interpretable results and that ensure the transparency of final findings. In one way or another there arises an issue of casting the results as information granules – conceptual entities that capture the essence of the overall data set in a compact manner. It is worth stressing that information granules not only support conversion of clouds of detailed data into more tangible information granules but, very importantly, afford a vehicle of abstraction that allows to think of granules as different conceptual entities; see [3, 17, 18, 19, 25, 26] and the references therein.

Clearly the task of information granulation is not a trivial one and it is dependent to a large extent on the application domain. Zadeh [25, 26] promoted a notion of information granulation in the framework of fuzzy sets. Other formal and commonly exploited environments of information granulation deal with rough sets [16] and set theory (interval analysis) [3, 15, 18]. In the context of granular computing the analysis of the Natural Language statements can be represented as operations on fuzzy sets. To make this point clearer we formalize the definition of a Natural Language. If a set of all words in a given Natural Language is denoted by X, then the set of all possible utterances in this language is represented by a set of all subsets of X, i.e. the power set of X, denoted as P(X).

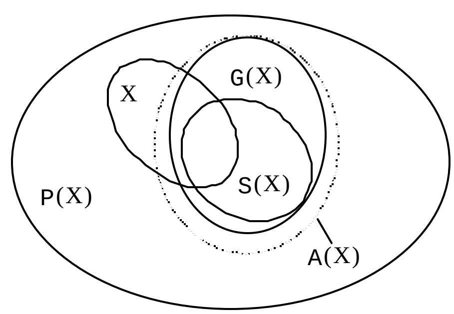

Figure 1. Set representation of a Natural Language

Of course only a small proportion of elements of P(X) represent statements which conform to the rules of grammar of this language. These represent a subset of P(X), referred to as G(X). And the grammatically correct statements that are meaningful are represented by $S(X)$ that is a subset of $G(X)$. The words (X) themselves have a dual nature; on one hand they have grammatical meaning (parts of sentence) and on the other hand they have semantics defined by the concepts they represent. The latter is expressed in a thesaurus-style explanation through the related words. In this sense every word represents a fuzzy set defined over the whole domain X.

The essence of syntactical analysis is to discriminate whether a given sentence belongs to $G(X)$ or $P(X)-G(X)$. Syntactical analysis dissects the sentence into noun- and verb- phrases and checks for conformity with the rules of grammar until the bottom-most level of individual words is reached. A standard approach to syntactical analysis involves application of parsing techniques that use binary logic in checking the conformity of a given sentence with the rules of grammar. However, in real life many Natural Language statements can be considered meaningful even if they do not conform fully to the rules of NL grammar. So, in terms of the above set formalism, we suggest that the syntactical analysis should adopt a wider scope $A(X)$ (of 'acceptable' statements) which is a superset of $G(X)$, i.e. $G(X) \subset A(X)$. The degree of relaxation of the NL grammar rules is represented in terms of fuzzy logic operations where we no longer deal with binary predicates for a specific rule of grammar, but have a full spectrum of the 'degrees of truth'. One thesis of this paper is therefore that a fuzzy sets formalism applied in the context of parsing NL statements captures the natural tolerance to grammatical errors that one encounters in real life.

Complementary to syntactical analysis is the semantic analysis that can be seen as a bottom-up process of distilling meaning from words and the relationships into which they enter. As the words are combined into phrases and sentences they evolve a more specific meaning then the component words taken in isolation. Ultimately the semantics of a sentence or a group of sentences is conveying the most specific message/knowledge that arises from the purposeful combination of words into sentences. Discussion of the semantic analysis in the context of fuzzy set formalism is a subject of a separate publication.

2. NATURAL LANGUAGE PROCESSING

Natural language, as opposed to artificial or formal one, is something that already exists and fulfills various functions in inter-personal contacts. We primarily use natural language to exchange information. Natural language has been evolving (for thousands of years) to enable us to express all sorts of things in all sorts of situations. Words, phrases and sentences expressed in Natural language are often ambiguous, their meaning depend on context or situation of their appearance. The structures of natural language are often complex, incomplete, interlaced. These could be considered mechanisms of adaptation of the Natural language enabling it to fulfill new tasks in an evolving world. What is amazing is that humanity created numerous of natural languages and can communicate inside given natural language and between natural languages.

Natural language is a tool supporting information representation and exchange in the process of human communication. On the other hand, natural language is rule driven, what is obvious in the context of its fundamental property of information exchange. Information encoded in a natural language construction (phrase, sentence, text) by a human being is addressed to other human being(s) with intention to be decoded and properly understood in their meaning. So, obviously, both subjects of information exchange supported by natural language constructions have to use the same rules in order to encode and decode respective information.

The tasks and process of information representation and exchange are fundamental subjects of science, research and technology in the computerized information era. Thus, these objectives raised temptation and forced attempts to formalization of natural languages and automation of their processing as well as processing of information supported by natural languages.

The goal of the natural language processing task is to design and build a computer system that will analyze, understand, and generate languages that humans use naturally. This goal is not easy to reach. "Understanding" language means, among other things, knowing what concepts a word or phrase stands for and knowing how to link those concepts together in a meaningful way. It is ironic that natural language, the symbol system that is easiest for humans to learn and use, is hardest for a computer to master. Long after machines have proven capable of inverting large matrices with speed and grace not achievable by human beings, they still fail to master the basics of human spoken and written languages cf. [4].

After decades of fruitful development of methods of natural language processing, it is clear that formalization of a full natural language and automation of its processing is far from completeness. Natural language formalization and processing is often being restricted by different factors, for instance restricted to areas limited with regard to syntax, semantics, knowledge, style, etc., and even in these local areas of meaning, automation of its processing is still defective.

The challenges we face stem from the high flexibility and ambiguity nature of natural language. Having English as his mother language, one effortlessly understand the sentence "Visiting aunts can be fun" assuming that you have some context knowledge about this sentence. Yet this sentence presents some difficulties to a software program that lacks both your knowledge of the world and human experience with linguistic structures. Is the more plausible interpretation that aunts are fun, or that rather visit is fun? Should the word "can" be analyzed as a verb or as a noun? Obviously, human being easily solves all these doubts with information recovered from contextual knowledge. However, information recovery that is subconscious for humans, raises challenge for automation, cf [12].

2.1. Social fundamentals

There is a very long-standing debate about the relative importance of nature and nurture in the development of the human intellect. Are we creatures endowed from birth with rich structures of knowledge and understanding, which require the stimulus of experience only in order to be jolted into conscious awareness? Or do we begin life essentially as blank sheets of paper on which the outside world writes what it may, and which begin with no predisposition towards one eventual set of contents rather than another?

In recent decades, it has been argued that scientific study of language gives us new evidence favouring a strikingly nativist account of human cognition. According to this view, genetics fixes the contents of our minds just as it fixes the detailed structure of our bodies.

The argument was first constructed in the 1960s by Noam Chomsky, in books such as *Cartesian Linguistics* [7]. In this and a series of subsequent writings Chomsky identified a large range of considerations (for instance, all human languages share certain universal structural features, and young children acquire their first language surprisingly fast), each of which, he urged, forces us to accept that `we do not really learn language; rather, grammar grows in the mind' [9]. Also cf. [23].

This discussion justify effort put on syntactical analysis of natural language constructions as the one of the most important and efficient analysis method, cf. [23, 24].

2.2. Lexical Acquisition

Many tasks in natural language processing need detailed, accurate information about the grammatical behavior of individual words of the subjective language construction. The primitive solution of this demand is to provide a natural language algorithm with a kind of lexicon of respective words and their features However, manual coding of lexicons for every task of language construction analysis is expensive, error-prone, and needs to be carried out afresh for new domains. It also cannot reliably capture statistical tendencies. To enable computers to process human language, we need databases (corpora) of words and language samples annotated to show their structural features, as a source of information and statistics to guide the development of language-processing algorithms. The important aspect of the annotated structural features of the words of such a lexicon is their ambiguity and uncertainty. The phenomenon of natural language is that phrases, sentences and texts bring contextual knowledge allowing for resolving ambiguity and uncertainty. In [12] the case of restricted subsets of natural language was discussed. It was shown that in such cases contextual information could be approximated by crisp relations defined on the set of noun phrases. It appears that the analysis of the natural language contextual information corresponds more closely to the operations on fuzzy sets that are essentially defined on the universe of the set of individual words and are labeled by given words. The value of membership function expresses the grade to which the labeling word can be related to the universe element. In this paper it is assumed that the fuzzy information is an immanent element of the lexicon and, as it is outlined in next sections, it can be dynamically modified during language constructions analysis.

2.3. Syntactical Analysis

Syntactical approach to natural language processing is the study of how words fit together to form structures up to the level of a sentence. Syntactical approach is a crucial stage and a crucial problem in natural language processing and in particular in extracting and representing information supported by natural language constructions.

The syntactical approach to natural language processing was extensively explored in the past Until fairly recently, almost all work on automatic parsing has treated the task as essentially similar to `compiling' programs in a formal computer language. Parsing was based on rules defining `all and only' the valid grammatical structures in a language; faced with a particular input string, the task was to find the structural analysis by virtue of which it is a valid string.

The trouble with this approach is that human language is quite messy and anarchic, by comparison with languages like Pascal or C. Programming languages are designed to conform to rigid (and fairly simple) rules. Programs that break the rules are rejected by the computer. But it isn't clear that a language like English is rule-governed in the same rigid way. If there is a definite set of rules specifying all and only the valid sentences of English, the rules are certainly much more complicated than those of C. But, to many of us, it seems that complexity is not the whole point - it is questionable whether English is fully definable by rules at all, and such rules as there are will often be broken with no bad consequences for communication. Thus, parsing natural languages must be intensely flexible and deeply tolerant to natural anarchy of its subjects. With these notes in mind, the proposed approach to parsing will relay on sensible application of proposed context free grammars, i.e. that it will not be applied maliciously to generate incorrect examples.

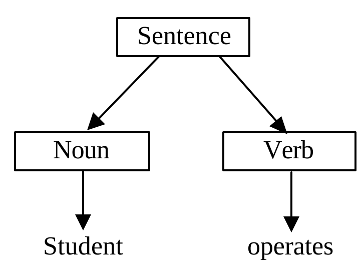

Fig. 1 Syntactical graph of the sentence " Student operates". The part of this graph with lower row dropped creates a pattern for other sentences of the same format.

2.4. Context Free Grammars

We focus attention on syntactical approach to natural language processing based on context free grammars CFG and transformations of CFG grammars. Of course, a grammar that is powerful enough to be able to analyse all English sentences is an impossibly large and complex, so we even will not try do construct it here. Rather we will try to develop a grammar that will meet the following three criteria:

- it will allow for analysis of all language phrases and sentences discussed in the paper,
- it could be developed to a more complete grammar that will extend the set of language phrases and sentences and will restrict generation of ungrammatical constructions
- it will use phrases and rules that are generally applicable in English, even if in some cases they involve a gross simplification of the way English works.

It is worth underlining that the grammar we develop generates a language that neither is included in English (the natural language), nor includes English. The language generated by the grammar intersects the English natural language and just their common part is a subject of our discussion.

A formal definition of context free grammar is available in most introductory computer science texts:

Definition: A tuplet G = (V, T, P, S) is a context free grammar assuming that

- V - is a finite set of variables
- T – is a finite set of terminal symbols
- S – is a variable being beginning symbol of the grammar
- P – is a finite set of production, every production is a pair with variable as its first element and a finite sequence of variables and terminal symbols.

Any sequence of terminal symbols derivable from the beginning symbol of the grammar belongs to the language generated by the grammar. Cf. [13] or any readings in mathematical linguistic on detailed description of context free grammars and context free languages.

For instance the following grammar:

G = ({“Sentence”, “Noun”, “Verb”},
{student, operate},
{“Sentence” -> “Noun” “Verb”,
“Noun” -> student
“Verb” -> operate },
“Sentence”)

generates the language consisting of the sentence “Student operate”.

Instead of a formal description of a grammar, we will be presenting only respective set of production or even only derivation of a language construction in the form of syntactical graphs. For extended description of context free grammar of English natural language cf. e.g. [4], pp. 51-53. Note: it would be better to say – for context free grammar of a language that includes a subset of English natural language as its part. In light of previous comments, creating a grammar that fully describes English or other natural language is more sophisticated than using context free mechanism and more complicated than a few pages description. Anyway, this grammar will be referred as the grammar of English natural language, for simplicity.

The syntactical graph in Figure 1 outlines derivation of the sentence “Student operates” in the above trivially simple grammar.

This simple example outlines the method of construction of simple sentences: noun *student* can be replaced by any other noun, e.g. *bird*, *wind* while verb *operate* by other verbs e.g. *fly*, *blow*, etc. Thus, the syntactical graph – equivalent to the production “Sentence” -> “Noun” “Verb” - creates a pattern of simple sentences, cf. Figure 1, while the production from variables noun and verb into terminal words – *student* and *operate* creates an instance of this pattern – the sentence *“Student operates”*.

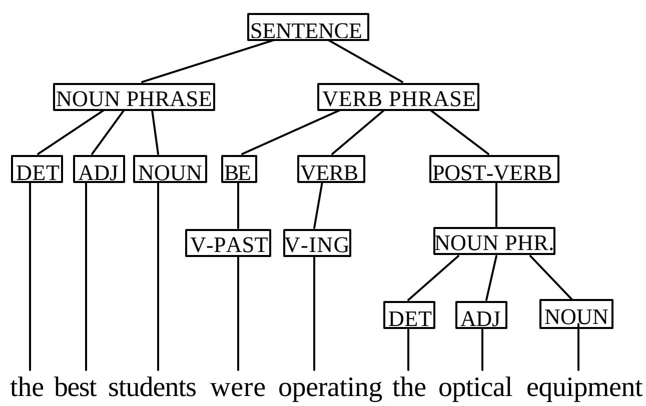

Fig. 2. Syntactical graph of the sentence “The best students were operating the optical equipment”.

3. PARSING NATURAL LANGUAGE

Having a lexicon of words and a grammar describing a language, it is possible to formulate an algorithm to determine whether or not any given text is constructed according to the rules of the grammar. If a sentence is grammatical, the algorithm should be able to describe its structure. If a sentence is ambiguous, then the algorithm should be able to describe all its possible structures. An algorithm performing such a task is called a parser. As it was noticed, the lexicon and the grammar are essential elements of a parser and they decide about the quality of natural language processing.

3.1. Syntactical structuring

Syntactical structuring as a main task of parsing process has also its powerful contribution to information extraction from natural language and – subsequently – to the process of information structuring and formation of a space of granular information. As it will be outlined below, syntactical structuring of language constructions strictly corresponds to granular structuring of information space. The following example gives intuition of the structure correspondence.

Let us consider the real sentence “The best students were operating the optical equipment”. The basis of this sentence “Student operates” creates the simple sentence with relation between both words: noun *students* and verb *operate*. Then, the sentence is developed to build the more complex relation between noun phrase and verb phrase, each of them having complex structure. The central element of noun phrase – the noun *students* – is described by an adjective and a determinant. The central element of the verb phrase – the verb *operate* – is transformed to past –ing form and is supplemented by post-verb phrase. Despite of the complex structure of the sentence, the main relation is built on both central elements: noun *students* and verb *operate*. The additions wrap these central elements in extra information that define more specific meaning of – still the same – relation between noun and verb. And finally, the sentence has unique derivation in the grammar. Note: such features as adjective comparisons, noun plurality, numeral ordinality, etc. are omitted since they do not raise any novelty in the discussion.

3.2. Ambiguity

The trivial example of Figure 1 can developed in order to specify included data. For instance, The sentence “The first student saw the man” develops embryonic sentence describing more exactly the noun *student*. And then the sentence: “The first student on the list saw the man with the camera” extends specificity of other parts of initial and subsequent sentences, cf. [4]. Armed with the grammar of English natural language, as e.g. in [4], and

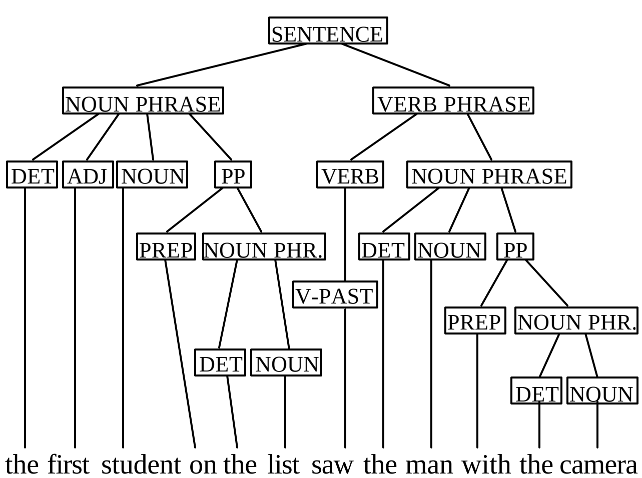

Fig. 3. Syntactical graph of the sentence “ The first student on the list saw the man with the camera”. The prepositional phase “with the camera” is assumed to be a part of noun phrase “the man ...”

with a lexicon of English words and idioms, one can start parsing English natural language constructions. Parsing the last sentence leads to the syntactical graphs presented in Figure 2 and 3:

Interpretation of this sentence is ambiguous: it is not clear whether the student was using camera when saw the men or rather the student saw the man and the man had camera. This ambiguity brings to two different syntactical graphs or – remembering that syntactical graphs are equivalent to derivations - to two derivations in context free grammar outlining the ambiguity.

The ambiguity could be resolved on the basis of contextual information. Considering both sentences “ The best students were operating the optical equipment. The first student on the list saw the man with the camera.” As a cohesive text, it would be easily deducted that the interpretation outlined in Figure 3 is correct.

4. INFORMATION GRANULATION

The naive example Figure 1 can be considered from the perspective of information supported by the sentences. On one hand, the pattern of language construction outlined in the form of syntactical graph defines relation between basic data of the noun type and basic data of the verb type. On the other hand, the sentence “*Student operates*” defines a relation between two simple pieces of data: “*student*” and “*operate*”. Both pieces of data would be seen as elementary granules of data emerging from a plain surface of single words. In the first case, we have a pattern that defines a set of relations between words that can fill in the pattern. In the second case, the relation is defined on strictly defined words. We will focus our attention on the relation defined on words, phrases and sentences rather than on patterns of language constructions.

4.1. Granular space formation

It would be observed that natural language constructions provide the family of relations between words themselves creating tuplets of related words, between words and a tuplets of related constructions and between tuplets of constructions. It is clear that language constructions may have recursive structure, i.e. one phrase may include another phrase of the same type (noun phrase includes another noun phrase) or of different type (verb phrase includes noun phrase). And so, the structuring relations have a character of a tangled up net rather than simple hierarchical tree structure. However, a kind of a hierarchical structuring of these relations could be defined as elements of different language constructions supporting contextual knowledge. Considering the example of simple text of two sentences presented in Figures 2 and 3, this hierarchy could be defined as:

- the set X of simple words, i.e. the set X = { be, best, camera, equipment, first, list, man, on, operate, optical, see, student, the, with},

- the set X with simple features added to form grammatical forms that can be seen as a syntactical graph defining proper grammatical form. In the presented example, these features can be illustrated by the following set of constructions: {operating, saw, the students, the equipment, the student, the list, the man, the camera, were},
- basic noun and verb phrases that do not include other noun or verb phrases as their parts. Again, they can be seen as syntactical graphs defining given phrases. The following phrases could be distinguished in the example: { on the list, the best students, the first student, the optical equipment, were operating, with the camera }
- compound phrases, i.e. noun and verb phrases that have other phrases as their parts. The example gives the following two compound phrases { the first student on the list, the man with the camera }
- sentences as pairs of noun and verb phrases. The example gives two sentences {The best students were operating optical equipment. The first student on the list saw the men with the camera.}

The granular space formed by language constructions is a kind of superstructure of the set of simple words and, from that point of view, can be considered as a dynamic extension of the static lexicon. This superstructure is built on the static fundament of lexicon every time new language construction is analysed.

4.2. Resolving ambiguity

The ambiguity of the second sentence outlined in Chapter III can be easily resolved while both sentences are considered as a cohesive text. We operate with a kind of dynamical environment called “sentence neighborhood” that moves information between consecutive sentences. We assume that some piece of information, a granule, defined in a sentence, is valid in next sentences as long as it is not redefined. And then, if the first sentence defines the granule *the optical equipment*, this granule is still valid in the second sentence. Unlikely, the granule *the best students* is moved to the second sentence, but it is redefined to the granule *the first student on the list*. However, the redefined granule, as being more specific than its origin, inherits properties of its predecessor, so it still remains in relation with the granule *the optical equipment*. The lexicon dependency between *camera* and *optical equipment* allows for binding the prepositional phrase *with the camera* to *the student* granule rather than to *the man* granule.

It is worth underlining that newly defined granules create dynamic extension of the lexicon, as it was flagged in Section II. And, of course, the newly added granules are bound with other granules of both static and dynamic parts.

These relations could be numerically described in the form of fuzzy sets having lexicon elements as their domains and labeled by given lexicon element. For instance, the granule *camera* interpreted as a fuzzy set may get the following membership values

fs(camera) = { ... , 0.9/camera, 0.7/equipment, 0.7/optical, 0.5/student, 0.5/man, ...}

at the at the basic level of static lexicon. This fuzzy set can then be developed to

fs(tcamera) = { ... , 0.9/camera, 0.7/equipment, 0.7/optical, 0.5/student, 0.5/man, ... , 0.9/the camera, 0.9/the best students, 0.9/the optical equipment, 0.5/the man, ... }

when the first sentence is analysed.. The fuzzy sets represented the granule the camera will have similar numerical values of membership function. The second sentence includes its granules to the lexicon and, finally, will bind the granule *the camera* with other granules in the fuzzy sets possibly having the following membership functions:

fs(tcamera) = { ... , 0.9/camera, 0.7/equipment, 0.7/optical, 0.5/student, 0.5/man, ... , 0.9/the camera, 0.9/the best students, 0.9/the optical equipment, 0.5/the man, ... 0.9/the first student on the list, ...}

Note: the numerical value expressing the relation between granules *the camera* and *the man* remains unchanged during all the process of text analysis and the dynamic extension of the lexicon. On the other hand, the sentences presented to text analyser increase the numerical value of the relation between wrapping granule based on the noun *student*. Thus, the ambiguity could be solved by simple comparison of membership values of certain fuzzy sets. The same solution could be reached when two fuzzy sets *the man* and *the first student on the list* are utilized. Membership values in the point *the camera* would also express tighter link between granules *the camera* and *the first student on the list* rather than between *the camera* and *the man*.

5. CONCLUSIONS

The paper casts the natural language processing problem in a novel framework of fuzzy sets and fuzzy logic based information processing. Simple examples considered in the paper indicate the feasibility of the task but a further research is needed to investigate the methodology for building the lexicon of fuzzy sets. Also an investigation into fuzzy parsing, that will capture the varying degree of tolerance to grammatical inconsistencies, will need to be undertaken.

6. ACKNOWLEDGMENT

The support for Visiting Professorships from the University of Alberta and the support from the Engineering and Physical Sciences Research Council (EPSRC, UK) are gratefully acknowledged.

7. REFERENCES

[1] J. Allen, *Natural language understanding*, Benjamin / Cummings, Menlo Park, California, 1988.

[2] M.R. Anderberg, *Cluster Analysis for Applications*, Academic Press, N. York, 1973

[3] A. Bargiela, 2001, *Interval and Ellipsoidal Uncertainty Models*, in W. Pedrycz (ed.) Granular Computing, Springer Verlag.

[4] C. Beardon, D. Lumsden, G.Ho;mes, *Natural Language and Computational Linguistics, an introduction*, New York, 1991.

[5] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, N. York, 1981

[6] Bikel, D., Schwartz, R., Weischedel R., 1999, An algorithm that learns what's in a name, *Machine learning-Special Issue on NL Learning, 34,1-3*.

[7] N. Chomsky, *Aspects of a theory of syntax*, MIT Press, Cambridge, Massachusetts, 1965.

[8] N. Chomsky, Cartesian linguistics : a chapter in the history of rationalist thought, Harper and Row, New York, 1965.

[9] N. Chomsky, *Rules and representations*, Columbia University Press, New Yourk, 1980.

[10] N. Chomsky, M. Halle, *The sound patterns of English*, Harper and Row, New York, 1968.

[11] M.A.K. Halliday, *An Introduction to Functional Grammar*, Edward Arnold Press: Baltimore, 1985.

[12] W. Homenda, Databases with Alternative Information, *IEEE Transactions on Knowledge and Data Engineering*, vol. 3, no. 3, September 1991.

[13] . J. E. Hopcroft, R. Motwani, J. D. Ullman: *Introduction to automata theory, languages and computation*. Addison-Wesley, Boston : c2001

[14] T. Kohonen, Self-organized formation of topologically correct feature maps, *Biological Cybernetics*, 43, 59-69 (1982)

[15] R. Moore, *Interval Analysis*, Prentice Hall, Englewood Cllifs, NJ, 1966

[16] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic, Dordrecht, 1991

[17] W. Pedrycz, *Computational Intelligence: An Introduction*, CRC Press, Boca Raton, FL, 1997.

[18] W. Pedrycz, Bargiela, A., 2002, *Granular clustering: A granular signature of data*, IEEE Trans. Syst. Man And Cybern. – B, vol. 32, 2, pp. 212-224.

[19] W. Pedrycz, M. H. Smith, A. Bargiela, A granular signature of data, *Proc. 19^{th} Int. (IEEE) Conf. NAFIPS'2000*, Atlanta, July 2000, pp. 69-73

[20] J. Pinkham, M. Corston-Oliver. Adding Domain Specificity to an MT system, In *Proceedings of the Workshop on Data-driven Machine Translation at 39th Annual Meeting of the Association for Computational Linguistics*, Toulouse, France, 2001, pp. 103-110

[21] S., Richardson et al., Overcoming the customization bottleneck using example-based MT In *Proceedings, Workshop on Data-driven Machine Translation, 39th Annual Meeting and 10th Conference of the European Chapter, Association for Computational Linguistics* 2001, Toulouse, France, pp. 9-16.

[22] G. Sampson, *English for the Computer*, Oxford University Press, 1995.

[23] G. Sampson, *Educating Eve*, Continuum International, London, New York, 1999.

[24] T. Winograd, *Language as a cognitive process*, Addison Wesley, Readings, Massachusetts, 1983

[25] L. A. Zadeh, Fuzzy logic = Computing with words, *IEEE Trans. on Fuzzy Systems*, vol. 4, 2, 1996, 103-111

[26] L. A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems*, 90, 1997, pp. 111-117.