Knowledge representation of passages type TOEFL

Meliza Contreras Gonzalez, Mireya Tovar Vidal, and Guillermo De Ita Luna

Benemerita Universidad Autonóma de Puebla, Facultad de Computación
mcontreras@cs.buap.mx, mtovar@cs.buap.mx,deita@cs.buap.mx

Abstract. Reading comprehension in the English language is a process that has been studied from different disciplines, due to the need for certification of the English language as a second language; this process is required in the qualification processes of students enrolled in a postgraduate course. Therefore, in this paper, a new model is proposed, based on the versatility of first order logic, situation calculus and semantic relationships to model the passages of this type of certification. The model is represented by a knowledge base, which considers the cognitive model of Van Dijk and Kinstch.

Keywords: inference · situation calculus · semantic model · FOL

1 Introduction

Reading requires the development of a complex cognitive system that supports the processing of information at different levels, whether conscious or unconscious. A good reader is one who is able to construct an integrated mental representation of the text, which is also coherent and accurate [1].

Readers of texts in languages other than their native one have two challenges: first to translate to their native language, and then to map the structure from the vocabulary that they know of the foreign language [2].

In the TOEFL (Test of English as a Foreign Language), in particular the reading comprehension section, in order to answer the questions, the reader builds a model of knowledge representation, which requires applying inferential processes to understand the meaning of the text [2].

For this reason, it is important to pose models of knowledge representation of the passages of the reading comprehension section, with their corresponding questions and answers. The purpose is to establish the meaning of the text according to the context in which it is found.

Although first order logic allows us to model assertions or predicates, it is also important to establish a model of the contexts in the passages, so situation calculus is a useful tool to do so. Another important aspect to favor this representation of knowledge is the identification of the semantic relationships present in the passages. So, depending on the type of relationship, inferences can be produced that help establish strategies to respond appropriately to the questions of the reading comprehension section of the TOEFL.

The content of this paper is divided as follows: section 2 shows a theoretical framework of cognitive models of reading comprehension and inferential processes. Section

3 shows the semantic relationships, the calculation of situations and their relation to the modeling of predicates considering the context. In section 4, an example is presented of how a passage is converted to a knowledge base according to semantic relationships and situation calculus. Finally the conclusions and future work are presented in section 5.

2 Cognitive Models

Cognitive models are characterized by studying how human beings know, think and remember. It explores the capacity of human minds to modify and control the way in which stimuli affect behavior and sustains learning as a process where meanings are modified internally.

This is achieved by integrating the mechanisms of short and long term memory with those of inference and although they are performed automatically, not all human beings perform it at the same time or in the same way. As Johnson[4] mentions in 2006, to understand instructions in reasoning experiments, students need to understand the concepts of premise, conclusion and implication to make a correct deduction.

Reading comprehension is the process of simultaneously extracting and constructing meaning with the following objectives: to decipher how letters represent words, to accurately and efficienly translate letters to sounds (extract meaning from text), to formulate a representation of the information that is being presented, which inevitably requires the elaboration of new meanings; and to integrate the new information with the old (construction of meaning) [7]. This last objective is the one that has proven most interesting both to psychologists through the generation of cognitive models, and to computer scientists who have sought mechanisms to explain or emulate the thought processes with the help of artificial intelligence and natural language processing. From the viewpoint of cognitive psychology, interested in the understanding of discourse, for a couple of decades, endless theoretical models have been developed that have tried to explain how comprehension occurs, specifying as key factors the role of the reader's prior knowledge, the making of inferences or the construction of different levels of mental representation that interact with the characteristics of the text.

The precursors of these mental models were Van Dijk and Kintsch[1] in 1978 with their article in the journal Psychological Review, which explained in detail the cognitive processing of a university text of social psychology. In this work, they sought to understand how the text read is remembered. Also, it is postulated that when reading a text, one works with three levels of mental representation: the surface code, the base text and the situational model.

Two key concepts in this recall process were the 'macrostructure' and the 'superstructure', which were proposed in that investigation. This theory assumed that textual processing is done in cycles, due to the limited capacity of short-term memory after decoding the code, and that a representation of the text (base text) in the memory was gradually built up in this way. This base text not only consists of a connected sequence of 'propositions', but also establishes a hierarchical structure of 'macro propositions', which correspond to the most important and least important themes of the text deduced (inferred) by the reader[1]

The base text, then, results from sequences of propositions that are made coherent by the 'repetition of arguments'. The macro structures, on the other hand, can be defined as higher order propositions that include underlying propositions. In other words, macropropositions are constructed with the micropropositions of a text, and are a summary or other abstract structure underlying a text, so they must be inferred from the text. Thus the micro and macropropositions form a 'macrostructure' of the text, that is, a semantic structure that defines the overall meaning of a text. However, these structures must associate with a context associated with the reader's experience. This forms situational model, which is a cognitive model of the situation reflected in the text that contains inferred material[1]..

Another proposed model is that of Miller[6] and Kintsch (1980) which consists of two components: a block program and a microstructure coherence program; so they made use of a coherence graph of working memory, but the making of inferences was not considered.

Also in 1995, the 3CAPS model was proposed by Goldman, Varma and Cote[3], which provides interactions between text processing, a priori knowledge and strategic reading processes.

Later Kintsch[8] in 1998, proposed the Construction-Integration model considering the networks of nodes and links between them, mapping these relationships to a coherence matrix.

Considering the recovery of memory to support the inference process, in 2001, Singer[9] and Kintsch proposed a platform based on the Construction-Integration model with the memory link model of Gillund and Shiffrin, in an effort to model the complex patterns in the inferential memory data.

On the other hand, Lin and Patel [11] in 2001 use the Minipar model to extract inference rules of the type if "X solves Y" then "X deals with Y".

Even though several cognitive models have been proposed, the Kinstch and Van Dijk model has interrelated elements that fuse cognitive psychology and predicate logic for support in the process of reading comprehension, which is interesting to address from the point of view of nonmonotonic reasoning.

2.1 Reading Comprehension

García García[12] in 1993 differentiates the following processes in reading comprehension: decoding, literal comprehension, inferential comprehension and meta understanding.

In the case of the decoding process, this consists of deciphering a code; in this case it is about giving a meaning to the printed letters. Two decoding processes are allowed: associating the written word with the available meaning in the subject's memory and recoding what it means to transform the printed letters into syllables and into sounds in order to activate the meaning.

Literal comprehension consists of combining the meaning of several words in an appropriate way to form propositions and adheres to the information explicitly reflected in the text.

Inferential comprehension provides a deeper understanding of the text and goes beyond what is explained in it. The reader, through inferences, elaborates a more integrated and schematic mental representation from the information expressed in the text and from his previous knowledge.

Metacomprehension is the consciousness and control that the reader has of his process of understanding. It consists of establishing some goals for reading, checking if they are being reached and rectifying quickly if necessary.

As can be seen, the process of reading comprehension is complex and although there are theoretical models and computational approaches, it is necessary to identify the elements that hinder the construction of meaning that in this case are the inferences, since they are applied according to the level of experience of the reader and the strategies used to build new meaning from the existing one.

2.2 Inferential Processes

An inference is considered a conclusion or opinion that is reached taking into account evidence or known facts. From the point of view of logic, an inference is defined as the process of deriving logical consequences from assumptions [13].

They fulfill diverse functions in the speech processing, allow the identification of referents, disambiguate and / or complete semantic meanings, as well as emptiness substitute when the information is not explicitly available. Linguistic inference has been defined as: ”any conclusion that a reader reasonably derives from a sentence” (Hurford and Heasley 1983)[13]. Another categorization divides them into propositional and pragmatic inferences. The first ones are based on linguistic knowledge and are derived from the semantic content of the explicit propositions of the text. Some authors call them logical inferences and are considered as necessarily true [14].

Within linguistic inferences, Kempson[14] defines entailment as a logical-semantic inference established between propositions. Generally, entailment is presented as a relation of interdependence, in that the truth value of one proposition is derived from the value of another and can be inferred independently of the context of enunciation. Therefore, ”any sentence p will imply q if when p is true, q must also be true”.

There are several types of inference, as mentioned in the works of León and Pérez [15] and Gutiérrez-Calvo[16] and are shown in Table 1.

3 Preliminary Concepts

3.1 Semantic Relationships

Semantics is the part of linguistics that studies the meaning of words, sentences and expressions of the language. All the words that maintain a relationship of meaning between them are part of the same semantic field. For example: carnation and rose belong to the semantic field of flowers [18].

Among the words that form a semantic field can exist relations of hyponymy and hyperonymy, synonymy and antonymy. [18]

A word is a hyponym of another if its meaning is included in it. For example, a rose is a hyponym of flower. A word is a hyperonym of another if its meaning includes the meaning of it. For example, flower is a hyperonym of rose.

Table 1. Inferences Classification.

Criterial	Name	Description
Degree of probability vs.Certain [15]	logic pragmatics	it is generated in formal reasoning systems it is based on the knowledge of the people
Temporality [15]	on-line off-line	It happens after reading It happens during a reading
Cognitive resources [17]	automatic strategics	It is done consciously they are produced in greater time by processing
Direction [17]	Forwards backward	It looks for precedents It predicts derivative information
Need of understands [17]	Obligatory Not necessary	They support the formation of the mental model generated additionally to the mental model
Coherence [9]	local global	in phrases Connection between distant parts of the text
Types of content [15]	Inference information Question	It sticks to the content It is triggered according to the question
Information sources [15]	textual based on knowledge	explicit content Constructed on the basis of his experience
Level of representation [17]	propositional structure Situational model	based on local coherence Contextualize the information

Synonymy is a semantic phenomenon by which the same concept or idea can be expressed with two or more different words. The synonymous words have, therefore, an equal or very similar meaning within the same context[18].

Antonymy is a semantic phenomenon that occurs when two words have an opposite meaning, e.g., bad and good.

These semantic relationships are present in the texts and in most cases their identification supports the inferential strategies to answer the questions of reading comprehension. However, to map the meaning of the texts to a knowledge base requires modeling these relationships from two aspects that depend on the context[18].

When a context is not required, but have a vocabulary of terms is available that allows the reader to determine hyponyms and hyperonyms, in this case, the entailment can be used to express these relationships, as shown in Zenteno[14]. But if the text presents synonymous relations, then the calculation of situations can support the modeling of these relationships by adapting their elements to the context.

3.2 Entailment

According to Zenteno, entailment is also identified as 'inference', as proposed by Kempson (1977) [14]. In terms of the inferential process, the entailment is widely used by linguists, both to explain the relations of meaning on the lexical level as in the case of: hyponymy, hyperonymy, synonymy at the level of sentence.

Because of its ambivalent nature, entailment can be defined, logically, in terms of valid rules of inference or, semantically, in terms of the assignment of truth or falsehood of related propositions: "(a proposition) p semantically entails (a proposition) q if and only if in every situation where p is true q is also true (or in all the worlds where p is true, q is true) "(Levinson 1983: 174) [14].

If entailment is handled as a logical relationship between propositions expressed by sentences, this idea has made it possible to relate systematically (with reference to predicate calculus and predicate relations, such as symmetry, transitivity, and reflexivity) notions such as hyponymy, synonymy, antonymy, related opposition and contradiction. Thus, for example, Palmer (1981) points out that hyponyms, predicates that establish a relationship of meaning, such that the meaning of one is included in that of another, involve entailment: for example, rose implies flower. The lexemes that are associated through a hyponymy relationship can also establish transitivity: rose implies flower, and flower implies being alive [14].

There are other types of entailment between propositions, for example, in the following propositions: David killed Golitat, Goliath died. The relationship is valid considering killing(David, Goliath) can be sure that if this proposition is true implies that dying(Goliath) is also true even if there are not hyponymy relations, but rather a cause with an effect.

While semantic relationships can be modeled with entailment, the context that is fundamentally required in synonymy has not yet been considered, so modeling the context will be addressed in the next subsection.

3.3 Situation Calculus

To address the calculation of situations, first the preliminary elements of first-order logic are defined below:

A first order language with equality is specified by two disjoint sets of symbols called the vocabulary of the language (Reiter)[5]:

Logical symbols: the interpretation of these is fixed by the rules of first-order logic.

- Parentheses of all shapes and sizes.
- Logical connectives: $\supset, \sim$.
- Variables:x,y,z, $\ldots$.
- Equality $=$.

Parameters these vary with interpretation

- Quantifier symbol : $\forall$.
- Predicate symbols: For each $n \geq 0$, a set (posibly empty) of symbols, called n-place or n-ary predicate symbols.
- Function symbols: for each $n \geq 0$, a set (posibly empty) of symbols, called n-place or n-ary function symbols.

Terms, atomic formulas, literals, well formed formulas are defined as usual, as are the concepts of free and bound ocurrences of a variable in a formula. A sentence is a formula with no free variables. The symbols $\vee, \wedge, \exists$ are defined to be suitable abbreviations ocurrences of a variable in a formula.

Assume given a nonempty set I, whose members are called sorts, in this case those terms are defined:

Logical symbols: As before, except that for each sort i, there are infinitely many variables $x_1^i, x_2^i, \ldots$ of sort i. Each term is aasigned a unique sort, as follows:

1. Any variable of sort i is a term of sort i
2. If $t_1, \ldots, t_n$ are terms of sort $i_1, \ldots, i_n$ respectively and f is a function symbol of sort $< i_1, \ldots, i_n >$, then $f(t_1, \ldots, t_n)$ is a term of sort i_{n+1}

Atomic formulas are defined as follows:

1. When t and t are terms of the same sort, $t = t$ is an atomic formula
2. When P is an n-ary predicate symbol of sort $< i_1, \ldots, i_n >$ and $t_1, \ldots, t_n$ are terms of sort $i_1, \ldots, i_n$ respectively, then $P(t_1, \ldots, t_n)$ is an atomic formula.

The calculation of situations is a logical formalism designed for the representation and reasoning about dynamic domains. It was proposed by John McCarthy in 1963.

Baker mentions that it allows to represent changing scenarios as a set of formulas of first order logic. Reiter[5] defines the basic elements of the calculation in the following way: The actions are considered all the changes in the world. A possible history of the world, formed by a sequence of actions is represented by a first order term called situation. A fluent is a property that may or may not sustain a given situation.

It is also defined the function $do(\alpha, s)$ that denotes a successor situation for s to execute the action α. For example, the predicate $cause(virus, disease)$ indicates the action that a virus causes a disease on the object x to the object y. If the function do (put (A, B), s) is applied, it means the situation resulting from putting A in B when the situation s occurs. In the calculation of situations, actions are denoted as function symbols and situations are first order terms.

With this scheme, the logic of first order can be used to formalize the effects of various actions.

- two function symbols of sort situation:
  - A constant symbol S_0, denoting the intial situation
  - A binary function symbol $do : action \star situation \rightarrow situation$
- A binary predicate symbol $\sqsubset$: $situation \star situation$, defining an ordering relation on situations
- For each $n \geq 0$, countably infinitely many predicate symbols with arity n, and sorts $(action \cup object)^n$. These are used to denote situation independent relations like $virus(Rotavirus)$,

Reiter[5] thinks that this representation allows to shape schemes of type answer-query, considering to raise a sequence of terms of action and a formula G, this one will be true in agreement to the execution of the actions that it contains. Nevertheless, there exists the so called problem of projection that allows to analyze this approach and Reiter to define it of this form:

The Projection Problem

Suppose D is a basic action theory, $\alpha_1, ..., \alpha_n$ is a sequence of ground action terms, and $G(s)$ is a formula with one free variable s, whose only situation term is s. Determine whether: $D \models G(do([\alpha_1, ..., \alpha_n], S0))$.

For example, a projection query for the sequence of actions

Closest_meaning(proven, showed)

might be: The word "proven" is closest meaning of showed over virus theories

$$D \models is(do(Closest\_meaning(proven, showed), virus\_theories))).$$

To return results of this projection, Reiter[5] defined the theorem of regression to solve the projection problem, for finding a representation for the return of the projection by means of a sentence regresable defined as $R[W]$.

In case of the evaluation of querys in the database, the response to a query W answers to the projection query in the resulting situation.

These definitions of the calculation of situations, it is possible to extend to formalize the context. McCarthy[10] proposes the term assertion of the form $assert(c, p)$, in this case the assertion indicates that the proposition p under the context c can be evaluated or executed. On the other hand, examining conversations with the query-answer block in this model raises two types of questions: the propositions that are used to determine if a proposition is false or true, so they require a Yes or No answer, and qualitative questions are used to find objects that hold a formula. To model the discourse, the query and reply functions can be proposed, which are the central representation in these discourses. Thus the query function establishes a context in which the answer to the question will be interpreted. For example, if you have the proposition p it is possible that it has true

Table 2. TOEFL type passage

Passage	Relations
” The term ’virus’ is derived from the Latin word for poison or slime. ”	$derived(virus,poison)$. $derived(virus,slime)$.
It was originally applied to the noxious stench emanating from swamps that was thought to cause a variety of diseases in the centuries before microbes were discovered and specifically linked to illness.	$emanate(noxiousStench,swamps)$. $caused(diseases,virus)$ $caused(diseases,$ $emanate(noxiousStench,swamps))$.
But it was not until almost almost the end of the nineteenth century that a true virus was proven to be the cause of a disease.	 proved(caused(diseases,virus))
The nature of viruses made them impossible to detect for many years even after bacteria had been discovered and studied.	detected(virus,hard). detected(bacteria,easy).
Not only are viruses too small to be seen with a light microscope, they also cannot be detected through their biological activity, except as it occurs in conjunction with other organisms.	 biggerThan(bacterias,virus) detected(virus,conjunction(virus,organisms)).
In fact, viruses show no traces of biological activity by themselves.	no(traces(virus,biologicalActivity) not(hyponim(virus,LivingAgents))
Unlike bacteria, they are not living agents in the strictest sense Viruses are very simple pieces of organic material composed only of nucleic acid, either DNA or RNA, enclosed in a coat of protein made up of simple structural units. (Some viruses also contain carbohydrates and lipids.)	hyponim(bacterias,LivingAgents) hyponim(virus,materialOrganic). composed(organicMaterial,nucleicAcid) hyponim(DNA,nucleicAcid) hyponim(RNA,nucleicAcid) enclosed(nucleicAcid,coatProtein) contain(virus, carbohydrates) contain(virus, lipids)
They are parasites, requiring human, animal, or plant cells to live.	hyponim(virus, parasites) live(parasites,require(livingBeings)) hyponim(human, livingBeings) hyponim(animal, livingBeings) hyponim(plant, livingBeings)
The virus replicates by attaching to a cell and injecting its nucleic acid.	replicate(virus, attack(virus,inject(nucleicAcid,cells)))
’ once inside the cell, the DNA or RNA that contains the virus’ genetic information takes over the cell’s biological machinery, and the cell begins to manufacture viral proteins rather than its own	 manufacture(controlled(cells,virus),viralProteins)

value according to the context that is interpreted. Thus the reply function will update the information, that is, it will only change the epistemic state of the discourse context. This derives a series of axioms in this regard:

- Interpretation Axiom (propositional): if ϕ is an closed formula, then $assert(query(K,\phi),\phi \equiv yes)$
- Frame Axiom (propositional): if ϕ is an closed formula, and yes does not ocurr in the context Ψ then $assert(K,\Psi) \supset assert(query(K,\phi),\Psi)$
- Interpretation Axiom (Qualitative) if x is free unique variable in ϕ then $assert(query(K,\phi(X)),\phi(X) \equiv answer(X))$
- Frame Axiom (Qualitative) if x is free unique variables in ϕ and *answer* does not ocurr in the context Ψ, then $assert(K,\Psi) \supset assert(query(K,\phi(X)),\Psi)$
- Answer Axiom : $assert(reply(K,\phi),\Psi) \equiv assert(K,\phi \supset \Psi)$

If this approach is possible to build applications that improve the training in the section of reading comprehension and thus the rates of upgrading of use of the TOEFL.Considering this support theory, a knowledge base is modeled below by means of a TOEFL type passage.

4 Example of TOEFL pasage

Three main reading skills are tested in TOEFL Reading[21]:

- First, TOEFL Reading tests what specific facts are mentioned in the passage, as well as what is not mentioned. The typical format is the good old ”One Best Answer”. An effective strategy is to make a ”road map” of the passage right away so that you can find the answers more efficiently. Certain skills such as skimming and scanning will help you more efficiently establish this map.
- Second, they test about certain pronouns, like ”its” or ”their”, refer to in specific parts of the text.
- Finally, they'll ask what inferences can be made form certain information.

In the reading passages [20], questions often ask what a word could be replaced by or what a word means. The context of the word in the sentence and in the whole passage will provide clues to its meaning. In this section there are five or six passages that have 400-500 words. Each passage is followed by eight to twelve questions. In some TOEFL questions, however, the context is not reliable for figuring out the meaning of the words. In this case your knowledge of synonyms, word forms, Latin and Greek roots, prefixes, and suffixes, will help to answer the questions about word meanings.

The following example of passage is available at the University of Chuvanan[19] in this case the main topic is about the viruses.

In the process of reading comprehension of these passages there are implicit inferences to be made according to the situational model of each reader, as Van Dijk commented in 1978. It is therefore necessary to identify semantic relatios as: synonyms, hyponyms, hyponyms, antonyms, cause-effect relations and entailment. In Table [**?**] the knowledge base about passage is represented.

The content of the passage allows to establish predicates and hyponims and hyperonims can be identified, in the case of these structures it is not required to establish a context to define them, but the entailment is present in this form:

- hyponim(human, livingBeings) means human entail livinBeings.
- hyponim(virus, parasites) means virus entail parasites.

In the case of synonyms, in this case, they are not present in the content of the passage.

Once the passage is represented, it is necessary to process the questions and identify the context to answer them considering the answer options as shown in Table 3:

Table 3. TOEFL type passage questions

Passage questions	assertions
The word proven in line 4 is closest meaning to which of the following (A) Shown (B) Feared (C)Imagined (D) Considered	assert(synonim(proven, showed),context) where context=disease
Before microbes were discovered It was believed that some diseases were caused by (A) germ-carrying insects (B) certain strains of bacteria (C) foul odors released from swamps (D) slimy creatures living near swamps	assert(sinonym(emanate, release),context) where context=swamps

So in this case, it is required to answer the question, according to the context of disease, these querys can be modeled by calculating situations, for example, the modeled assertion: $assert(synonim(proven, showed), disease)$ imply that a answer will be found, in this case it is necessary to prove that proven is a synonim of showed, applying the axiom of qualitative interpretation, this result is $assert(query(K, synonim(proven, showed), disease), disease \equiv answer(X))$ where $\phi(X)$ would represent the predicate of synonymy and the answer X is showed, discarding the others.

In the same case, in the assertion $assert(sinonym(emanate, release), swamps)$, the terms $emanate$ and $release$ are not similar unless they are related in the context of $swamps$.

5 Conclusions

Although reading comprehension is a complex process, designing representation models that allow the identification of terms, semantic relationships, entailments and context-related assertions will favor the generation of inferences to design query-answer systems to improve the achievement in the reading comprehension sections of TOEFL exams.

The assertions generated from the calculation of situations provide intuitive expressiveness to associate the semantic relations to a context, so this representation will favor to identify properties and enrich inferential processes.

Kinstch and Van Dijk's model, emphasize the situational model as an element dependent on the reader's experience, with the description of the contexts generated from the calculation of situations, it is possible to generate a representation closer to the reader's experience,thus strategies can be elaborated to improve the process of reading comprehension.

Acknowledgment

This work was carried is supported by the Sectoral Research Fund for Education with the CONACyT project 257357 and Thematic Network in Language Technologies(RedTTL) with its program Research stays for students.

References

1. Kintsch, W., Van Dijk, T.A.: Toward a model of text comprehension and production. Psychological Review, **85**, 363-394. (1978)
2. MacMillan, F.: Lexical patterns in the reading comprehension section of the TOEFL test. Revista do GEL, **3**, 143-172. (2006)
3. Goldman, S. R., Varma, S.: CAPping the construction-integration model of discourse comprehension. In C. A. Weaver, S. Mannes, C. R. Fletcher (Eds.), Discourse comprehension: Essays in honor of Walter Kintsch (pp.337358). (1995)
4. Johnson-Laird, P. Como razonamos. Oxford University Press. (2006)
5. Reither, R.: Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems. Oxford MIT Press. (2001)
6. Miller, J. R., Kintsch, W.: Readability and recall of short prose passages: A theoretical analysis. Journal of Experimental Psychology: Human Learning and Memory **6**(4), 335-354 (1980)
7. Snow, C. E.: Reading for understanding. Santa Mónica, CA: RAND Education the Science and Technology Police Institute, (2001)
8. Kintsch, W.: The use of knowledge in discourse processing: A construction-integration model. Psychological Review,**95** 95 , 163182.(1998).
9. Singer, M., Kintsch, W.: Text retrieval: A theoretical exploration. Discourse Processes, **31** , 2759.(2001)
10. McCarthy,J., Buvac, S.: Formalizing Context (Expanded Notes),Stanford University, USA,(1994)
11. Lin, D., Pantel, P. Discovery of inference rules for question answering. Natural Language Engineering, **7** (4), 343360. (2001)
12. García, E.: La comprensión de textos.Modelo de procesamiento y estrategias de memoria. Revista Didáctica **5**(1), 87-113 (1993)
13. Bruner, J.S.: Going beyond the information given. In H.E. Gruber, K.R. Hammond y R. Jessor (Eds.), Contemporary approaches to cognition (pp. 4169). Cambridge, MA: Harvard University Press. (1957).
14. Lin, D., Pantel, P. Discovery of inference rules for question answering. Natural Language Engineering, **7** (4), 343360. (2001) Zenteno, C.: El entrañamiento. Inferencia proposicional o léxica?. Lenguas Modernas,**26**,227-244,(2000)

15. León, J.A., Pérez, O. :Taxonomías y tipos de inferencias. En J.A. León (Coord.), Conocimiento y Discurso: Claves para inferir y comprender.Madrid: Pirámide. (2003)
16. Gutiérrez-Calvo, M.: Inferencias en la comprensión del lenguaje. En M. de Vega y F. Cuetos (Coords.), Psicoling´uistica del español. Madrid: Trotta. (1999)
17. Escudero, I., León, J.A.: Procesos inferenciales en la comprensión del discurso escrito. Influencia de la estructura del texto en los procesos de comprensión. Revista Signos, **40**, pp 311-336, (2007)
18. Campo y Garrido, N.: Relaciones semánticas entre las palabras: hiponimia, sinonimia, polisemia, homonimia y antonimia. Los cambios de sentido, en Contribuciones a las Ciencias Sociales, (2010)
19. Chuvanan University, `http://http://www.cvauni.edu.vn/`. Last accessed 4 sept 2018
20. Sungatullina, D., Zalyaeva, E., Gorelova, Y.: Metacognitive awareness of TOEFL reading comprehension strategies. SHS Web of Conferences **40** (2016)
21. Mahnke, M.K., Duffy, C.B.: TOEFL Preparation Course. Heinemann. (1993)