ISSN 2007-9737

Knowledge Representation in TOEFL Expository Texts

Meliza Contreras González[1], Mireya Tovar Vidal[1], Guillermo De Ita Luna[1], Aurelio López López[2]

[1] Benemérita Universidad Autonóma de Puebla,
Faculty of Computer Science,
Mexico

[2] Instituto Nacional de Astrofísica, Óptica y Electrónica,
Computer Science Department,
Mexico

{mcontreras, mtovar, deita}@cs.buap.mx, allopez@inaoep.mx

Abstract. Reading comprehension in the English language is a process that has been studied from different disciplines. Many postgraduate programs require certification in another language, hence the importance of seeking semantic patterns that allow the creation of intelligent tools to train students in these tasks. The objective of this work is to characterize expository passages semantic structures through FOL and situation calculus with the question-answer block.

Keywords. Inference, situation calculus, semantic patterns, 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 can construct an integrated mental representation of the text, which is also coherent and accurate [7].

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 [8]. In the Test of English as a Foreign Language(TOEFL), in particular, the reading comprehension section, to answer the questions, the reader builds a model of knowledge representation, which requires applying inferential processes to understand the meaning of the text [8].

For this reason, it is crucial 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. Although first-order logic allows us to model assertions or predicates, it is also essential to establish a passages context model, so situation calculus is a useful tool to do so. Another crucial aspect of favoring this knowledge representation is the identification of the semantic relations present in the passages.

So, depending on the type of relationship, the inferences can be produced that help establishes strategies to respond appropriately to the questions of the reading comprehension section of the TOEFL.

Efforts have been made to improve understanding through tips, strategies, identification of rules, and practices. But in most cases does not take into account the context that is a fundamental element in the mental representation used by the brain to generate inferences and build a network of related concepts. In this work is intended to show the knowledge representation of expository texts considering to model the context by situation calculus as well as identify the semantic relations present in the explanatory texts of the TOEFL to facilitate the selection of answers in questions.

The content of this paper is divided as follows: Section 2shows a theoretical framework

of cognitive models of reading comprehension. Section 3 the semantic relations, the situation calculus, and their relation to the modeling of predicates considering the context are presented. In section 4, an example is given of how a passage is converted to a knowledge base according to semantic relationships and situation calculus. A knowledge representation of the passage is presented in Section 5, an algorithm for the generation of responses is proposed in section 6. Finally, the conclusions and future work are presented in section 7.

2 Related Work

Cognitive models study how human beings learn, think, and remember. It explores the capacity of human minds to modify and control how stimuli affect behavior and sustain learning as a process where meanings are changed internally. 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[5] 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.

This process include to decipher how letters create words, to accurately and efficiently translate them to sounds (extract meaning from text), to formulate a representation of the information that is being presented, which inevitably requires the elaboration of new definitions; and to integrate the previous knowledge with the old (construction of meaning) [12]. 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 [7], for decades, endless theoretical models have been developed that have tried to explain how comprehension occurs. Key factors are considered how 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[7] 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 critical and least essential themes of the text deduced (inferred) by the reader[7].

The base text, then, results from sequences of propositions that are made coherent by the 'repetition of arguments'. The macrostructures, on the other hand, can be defined as higher-order propositions that include underlying propositions. In other words, macropropositions are constructed with the micropro positions of a document and are a summary or different abstract structure underlying a text, so they must be inferred from it. Thus the micro and macropropositions form a 'macrostructure', 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. Thus, a situational model is formed, which is a cognitive model of the situation reflected in the text that contains inferred material [7]. Also in 1995, the 3CAPS model was proposed by Goldman, Varma and Cote[4], which provides interactions between

text processing, a priori knowledge, and strategic reading processes. Later Kintsch [6], proposed the Construction-Integration Model considering the networks of nodes and links between them, mapping these relationships to a coherence matrix.

Even though several cognitive models have been proposed, the Kinstch and Van Dijk model [6] has interrelated elements that fuse cognitive psychology and predicate logic for support in the process of reading comprehension, which is interesting from nonmonotonic reasoning point-of-view.

This work proposes a knowledge representation based on the importance of the situational model, taking advantage of the benefits of semantic structures and the situation calculus to approach the construction of inferences closer to the creation of mental representations.

3 Preliminaries

In this section, the theory that will support the expository passages representationis mentioned.

Initially, the benefits of the calculation of situations to model contexts are described, which in the case of TOEFL texts are required, in addition to generating queries on predicates, which favors the modeling of inferences. Subsequently, the entailment is described as an inference tool that allows establishing rules to associate the correct answers to specific questions, these in the case of the expository texts depend mainly on the semantic relations of these, so they are mentioned in the last subsection.

3.1 Situation Calculus

To address the situation calculus, first, the elements of first-order logic are defined below:

Two disjoint sets of symbols specify a first-order language with equality, called the vocabulary of the language [11]:

Logical symbols: the rules of first-order logic fix the interpretation of these:

— Parentheses of all shapes and sizes.

— Logical connectives: $\supset, \sim$.

— Variables:x,y,z, ...

— Equality $=$.

Parameters these vary with interpretation.

— Quantifier symbol : $\forall$.

— Predicate symbols: For each $n \geq 0$, a set (possibly empty) of symbols, called n-ary predicate symbols.

— Function symbols: for each $n \geq 0$, a set (possibly 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 occurrences 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 occurrences of a variable in a formula.

Assume given a nonempty set I, whose members are called sorts, in this case; those terms are defined as
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 assigned 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.

Situation calculus is a logical formalism designed for the representation of dynamic domains. John McCarthy proposed it in 1963.

Baker in [11] mentions that it allows for the representation of changing scenarios as a set of formulas of first-order logic. Reiter[11] defines the essential elements of the calculus in the following way: All the changes in the world are considered actions. 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.

The function $do(action(A, B), s)$ is also defined, it means the situation resulting from executing the action(predicate) A over B when the situation s occurs. For example, the sentence "Virus causes diseases in humans ", could be modeled how $disease = do(cause(Virus, human), health)$; that is, the term $x = Virus$ generate to the object $y = human$ a situation change of health to disease (next situation).

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

— two function symbols of sort situation:
  – A constant symbol S_0, denoting the initial situation.
  – A binary function symbol $do : action \star situation \rightarrow situation$.

— A binary predicate denoted by $\sqsubset$: $situation \star situation$, defining an ordering relation on situations.

— For each $n \geq 0$, any number of predicate symbols with arity n, and sorts $(action \cup object)^n$. These are used to denote situation independent relations.

It is possible to extend these definitions of situation calculus to formalize the context. McCarthy[10] proposes the assertion term 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: propositional questions are used to determine if a proposition is false or true, so they require a Yes or No answer; qualitative questions are used to find objects that fulfill a formula. The discourse could be modeled between the query and reply functions. 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 truth value that depends on the context in which it is is interpreted. Thus the reply function will update the information; that is, it will only change the epistemic state of the discourse context. A series of axioms are therefore derived. Let K a knowledge base, Ψ a context term and a ϕ a resulting answer of query.

— Interpretation Axiom (propositional): if ϕ is a closed formula, then:
  $assert(query(K, \phi), \phi \equiv yes)$

— Frame Axiom (propositional): if ϕ is a closed formula, and yes does not occur in the context Ψ then: $assert(K, \Psi) \supset assert(query(K, \phi), \Psi)$

— Interpretation Axiom (Qualitative) if x is a free unique variable in ϕ then
  $assert(query(K, \phi(X)), \phi(X) \equiv answer(X))$

— Frame Axiom (Qualitative) if x is a free unique variables in ϕ and $answer$ does not occur in the context Ψ, then
  $assert(K, \Psi) \supset assert(query(K, \phi(X)), \Psi)$

— Answer Axiom : $assert(reply(K, \phi), \Psi) \equiv assert(K, \phi \supset \Psi)$

3.2 Entailment

According to Zenteno, entailment is also identified as 'inference', as proposed by Kempson (1977) [13]. 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 the sentence.

Because of its ambivalent nature, entailment can be defined, logically, in terms of inference rules 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) [13].

If entailment is handled as a logical relationship between propositions expressed by sentences, this idea has made it possible to relate systematically (regarding 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 [13].

There are other types of entailment over propositions, for example, in the following propositions: David killed Goliath, 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-effect relation.

While semantic relations 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 Semantic Relations Representation in Expository Text

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 [3].

Among the words that form a semantic field can exist relations of hyponymy, hyperonymy, synonymy and antonymy [3].

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.

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[3].

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 depend on the context [3].

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[13]. However, 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.

In the following section, considering this support theory, a knowledge base of TOEFL type passage is modeled.

4 TOEFL Passage Knowledge Representation

Three main reading skills are tested in TOEFL reading comprehension section [9]:

— First, this section evaluates the ability to detect explicit facts and infer implicit facts in the passage. 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.

Topic: Virus

The term 'virus is derived from the Latin word for poison. or 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. But it was not until almost the end of the nineteenth century that a true virus was proven to be the cause of a disease.

The nature of viruses made them impossible to detect for many years even after bacteria had been discovered and studied. 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. In fact, viruses show no traces of biological activity by themselves. 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.) They are parasites, requiring human, animal, or plant cells to live. The virus replicates by attaching to a cell and injecting its nucleic acid.' 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.

1. Which of the following is the best title for the passage.

(A) New Developments in Viral Research (B) Exploring the Causes of Disease (C) DNA: Nature's Building Block (D) Understanding Viruses

2. 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

3. The word "proven" in line 4 is closest meaning to which of the following.

(A) Shown (B) Feared (C) Imagined (D) Considered

4. The word nature" in line 6 is closest in meaning to which of the following?

(A) Self-sufficiency (B) Shapes (C) Characteristics (D) Speed

5. All of the following may be components of a virus EXCEPT

(A) RNA (B) plant cells (C) carbohydrates (D) a coat of protein

6. The author implies that bacteria were investigated earlier than viruses because

(A) bacteria are easier to detect (B) bacteria are harder to eradicate (C) viruses are extremely poisonous (D) viruses are found only in hot climates

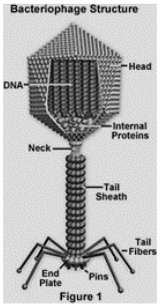

Figure 1

Fig. 1. Example of Expository Passage, source text: Chuvanan University source [1], figure source: MicroMagnet Dictionary [2]

— Second, It measures the coreference about certain pronouns, like ”its” or ”their”, refers to in specific parts of the text.

— Finally, this proof generate the capacity of create inference from certain information.

In the reading passages questions often ask what a word could be replaced by or what a word means. The context of the word in the sentence and the whole text 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. A typical document of TOEFL is the expository text, that is non-fiction text meant to inform, analyze, explain, or give additional detail about a topic. These texts show several semantic relations with patterns both in the description of the passage and in its question-answer block.

4.1 Expository Passage Example

Figure 1 shows an example of an expository text on the subject of viruses, available on the website of University of Chuvanan [1]. The structure of the passage is as follows: In the first paragraph, the main topic is introduced, while in the subsequent sections, the components of viruses and their way of attacking are described. Regarding the semantic relations, hyperonyms and hyponyms are found, as in the sentences: "Viruses are very simple pieces of organic material"; "They are parasites", meronymy in the sentences: "Viruses are 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)".

Finally, after the text of the passage, the questions are asked about it, which have different levels of complexity from the identification of the main topic (question 1), synonymy (2, 3, and 4), meronymy (5), cause-effect relations (6).

Therefore, from this analysis, a knowledge representation can be made using the theoretical elements of semantic relations, situation calculus, and entailment, whose proposal is shown in the following subsection.

5 Knowledge Representation with Situation Calculus

It is, therefore, necessary to model the semantic relations such as synonyms, hyperonyms, hyponyms, cause-effect relations, and entailment to build inference construction rules. In Table 1, the knowledge base is represented regarding predicates about situation calculus(FOL) for the passage.

The knowledge representation was generated regarding the predicates formulation, the verbs of the sentences were identified how predicates, for example, "derive(virus, poison)", concerning the arguments were placed subjects, objects,events, "emanate(noxious stench,swamps)".

In some cases, the predicates needed to be modeled with nested predicates since the action was required as an argument, as in the case of "cause(diseases, emanate(noxious stench, swamps))".

Also, the content of the passage allows us to determine predicates to model hyponyms and hyperonyms; it is not required to establish a context to define these structures, but the entailment is present in this form:

— $hyponym(livingBeings, human)$ means human entails livingBeings.

— $hyponym(parasites, virus)$ means virus entails parasites.

In the case of synonyms, they are not present in the content of the passage, but they appear in the question section.

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 2. So in this case, it is required to answer the questions, according to the context, these queries can be modeled by situation calculus, for example, the modeled assertion: $assert(synonym(proven, showed), disease)$ imply that an answer will be found, in this case, it is necessary to determine that established is a synonym of showed, applying the axiom of qualitative interpretation, this result is $assert(query(K, synonym(proven, showed), disease), disease \equiv answer(X))$ 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 $assert(synonym(emanate, release), swamps)$, the terms $emanate$ and $release$ are not similar unless they are related in the context of $swamps$.

Table 1. Knowledge Base Representation regarding situation calculus and semantic relations

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

Table 2. TOEFL type passage queries and assertions

Passage questions	assertions
1. Which of the following is the best title for the passage.	
(A) New Developments in Viral Research	assert(describe(virus, innovation))
(B) Exploring the Causes of Disease	assert(describe(virus, diseases))
(C) DNA: Natures Building Block	assert(describe(virus, composition))
(D) Understanding Viruses	**assert(describe(virus, behavior))** correct
2. Before microbes were discovered	
It was believed that some diseases were caused by	**assert(synonym(emanate,release),context)**
(A) germ-carrying insects	where context=swamps
(B) certain strains of bacteria	
(C) foul odors released from swamps	
(D) slimy creatures living near swamps	
3. The word proven in line 4	
is closest meaning to which of the following	**assert(synonym(proven,showed),context)**
(A) Shown	where context=disease
(B) Feared	
(C)Imagined	
(D)Considered	
4. The word nature" in line 6 is closest in meaning	
to which of the following?	
(A) Self-sufficiency	**assert(synonym(nature,characteristics),context)**
(B) Shapes	where context=biology
(C) Characteristics	
(D) Speed	
5. All of the following may be components of a virus EXCEPT	
(A)RNA	assert(meronym(virus,RNA))
(B) plant cells	**assert(not(meronym(virus,plant cells))) correct**
(C) carbohydrates	assert(meronym(virus,carbohydrates))
(D) a coat of protein	assert(meronym(virus,a coat of protein))

6 Algorithm Proposal over Knowledge Representation

So in this work, the Algorithm 1 is proposed to simplify the detection of responses using semantic relationships and generating inferences by situation calculus, from the knowledge representation raised.

The first step of the algorithm is to generate all the predicates of the passage. This action needs an interpretation of the verbs, the objects, the events how appears in Table 1 as exemplification. Later there will be predicates that are not sufficiently modeled with action and require using the hyperonyms, hyponyms, synonyms, meronyms and the cause-effect relations, so they have to be separated from the rest of the predicates. Depending on the type of relationship found, different inference processes have to be applied, which are shown in the representation of Table 1.

In the case of hyponyms, hyponyms can be worked through a hierarchical structure; therefore, it is suggested to build a tree of terms to establish the generalization of the specification of concepts. This structure supports Van Dijk's situational model, considering that it will begin construction through some system, the inference would only be to do the verification of what would be the parent node towards the children to determine if they correspond to the same category. But more specific, for example, viruses are an organic material, thus, this material is the parent concept, and viruses and bacteria would be child concepts. In the case of the passage questions do not correspond to relations of hyponymy, hyperonymy, but in table 1 are shown the assertions.

Algorithm 1 Algorithm for answer detection over expository texts

```
Get the predicates from the passage
Classify the predicates according to relation semantic found
if relation = hyponym then
    Build a terms tree that contains the hyponymic and hyperonymic relations
end if
if relation= meronymic then
    Construct a graph called part-all where the leading term is in the center and its connected parts around of it.
end if
Get the predicates from the questions and generate the associated context for each answer.
if relation question =synonym then
    Generate a predicate for each answer choice
    Applicate the Frame axiom qualitative and link the answer to the nearest context
else if relation question =hyponym then
    Generate a predicate for each related hyponym
    Applicate the Frame axiom qualitative and link the answer to the hyperonym in the answer
else if relation question =meronym then
    Generate a predicate for each related member
    Apply the Frame axiom qualitative and link the answer to the member that appears in the passage and adjusts to the answer
else if relation question =effect-cause then
    Generate predicates with verbs and agents associated with the cause-effect relation
    Apply the Frame axiom qualitative and link the answer to the verb closest to the meaning.
else if question = factual then
    Generate predicates associated with question words
    Applicate the Frame axiom qualitative and link the answer to the type of response: event, place, character, related action
end if
```

On the other hand, if a relation meronym is found, it is required to construct a graph so that conceptually the central concept can be modeled, and around it, the parts can be generated. So that of the term virus has several elements that compose it, like RNA, proteins, carbohydrates, etc., the items would form the essence with the central node, and this would facilitate the understanding of the relation part-whole. Once with the representations of the relations present in the expository passages, other predicates are constructed over questions that allow us to generate the inferences and thus detect the correct answers.

The passage contains three questions of synonymy relation (2, 3, and 4), in table 2 shows the approach of the solution. In the first instance it is required to find the term Therefore, it is applied to the axiom frame(situation calculus) to indicate that an answer can be found. In question three, when you place the closest meaning of testing within the other four actions, really one that can be generated according to a context of disease is closer. Then, the synonymy relationship is determinated that the other solutions would not have felt. In the same way in what would be question four, there is a synonymy question with a context about biology, nature, and characteristics are compared, according to use a lexicon.

In the case of the questions of the example passage, there is no hyponymy in the issues, but in this case, it would have to be established based on the structure of the tree of concepts what would be the query from top to bottom. The answer is generated in some branch to determine that there is a hyponymy relationship. The question five has a meronymy case, with the graph with the central structure and the elements can generate assertions considering the frame axiom so that some of the parts that are being proposed do not appear as a connection with the central concept is more natural to infer which would be. The answer indicates that given the passage and the structure, it is constructed that point B effectively does not correspond to an element of what would be viruses.

And in case of the cause-effect relations as commented on as question six and question one are more involved. In the case of issue 1, is a hypothesis based on several proposed topics, a predicate would have to be elaborated for each one of the answers, and one would have to search for the terms of a frequency analysis which would be the most common term.

On the other hand, question six is more complicated because it requires the generation of

predicates from Table 1 and assembled so that you can determine the issues to answer what you are working. In some paragraphs mention that bacteria in an implicit sense when doing research, it was easier to detect their study because they are bigger than viruses.

Several steps are required to be able to generate inference. Some answers can not be created. For example, the answer to this question is no one circumstance speaks of the hot climates. On the other hand, the answer c describes that the viruses were a kind of poison, but that is not anything related to the question of the bacteria, and as an alternative, the answer B (eradicate) is not mentioned the use of antibiotics or some other mechanism. Therefore, it is out of context.

Finally, in the case of the factual issues there would have more specific patterns, that it would be in this case to identify if an event, a place, a character or an action, and from there to search the predicates that one has that type of relation according to the order in which the predicate was generated.

Thus depending on the type of questions a different inference has to be generated, in the proposals indicates that goes a verb would be the name of the predicate at least a subject and an object and from there functional the corresponding evaluation can be made.

7 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 create a representation closer to the reader's experience. Thus strategies can be elaborated to improve the process of reading comprehension.

Acknowledgements

This work was carried is supported by the Sectoral Research Fund for Education with the CONACyT project 257357 and VIEP Project 2019.

References

1. **Chuvanan University (2018).** Collection of TOEFL passages.
2. **Davidson, M. (2019).** *Virus*.
3. **Garrido, N. C. (2010).** Relaciones semanticas entre las palabras: hiponimia, sinonimia, polisemia, homonimia y antonimia. los cambios de sentido. *Contribuciones a las Ciencias Sociales*, Vol. mayo 2010.
4. **Goldman, S. & Varma, S. (1995).** *Discourse comprehension: Essays in honor of Walter Kintsch*, chapter CAPping the construction-integration model of discourse comprehension. pp. 337–358.
5. **Johnson-Laird, P. (2006).** *Como razonamos*. Oxford University Press.
6. **Kintsch, W. (1998).** The use of knowledge in discourse processing: A construction-integration model. *Psychological Review*, Vol. 95, pp. 163–182.
7. **Kintsch, W. & Dijk, T. (1978).** Toward a model of text comprehension and production. *Psychological Review*, Vol. 85, No. 1, pp. 363–394.
8. **MacMillan, F. (2006).** Lexical patterns in the reading comprehension section of the TOEFL test. *Revista do GEL*, Vol. 3, No. 1, pp. 143–172.
9. **Mahnke, M. & C.B.Duffy (1993).** *TOEFL Preparation Course*. Heinemann.
10. **McCarthy, J. & Buvac, S. (1994).** Formalizing context (expanded notes). Technical report, Stanford University.
11. **Reither, R. (2001).** *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. Oxford MIT Press.

12. Snow, C. (2001). *Reading for understanding*. CA: RAND Education the Science and Technology Police Institute.

13. Zenteno, C. (2000). El entrañamiento. ¿inferencia proposicional o léxica? *Lenguas modernas*, Vol. 26, pp. 227–244.

Article received on 29/10/2019; accepted on 04/03/2020.
Corresponding author is Meliza Contreras González.