Linguistic Constraints as a Sub-component of a Framework that has Multi-linguistic Applications

Salwa Hamada

Researcher at Electronics Research Institute
Cairo, Egypt
hesalwa@yahoo.com

Adil Al Kufaishi

Lecturer at the University of Copenhagen
adil@hum.ku.dk

Abstract

The paper is an attempt to promote Arabic language technologies and resources. It proposes a framework of linguistic constraints LCs that could be utilized to account for Arabic textual ambiguity and to hinder the generation of Arabic ungrammatical structures. The proposed constraints operate at all linguistic levels: lexical, morphological, sentential and supra-sentential. The paper contributes to solving a number of textual analysis problems which both translators and text analysts encounter in their deconstruction process. It has made an extensive use of the linguistic componential analysis theory, the distinctive feature matrices and the conceptual frames. The framework has a built-in statistical sub-component which, in essence, determines the frequency of each linguistic unit LU and ranks its order in terms of its frequency. The lexical items are arranged in the order of their frequency of occurrence; this is referred to as the Frequency Rate (FR). The framework is applied for disambiguating various types of texts: narrative, descriptive, expository conversation or argumentative at different linguistic levels and has proved to be capable of resolving different types of textual ambiguities.

1. Introduction

Ambiguity is a linguistic phenomenon that is not restricted to a particular language. Disambiguating texts is a process that is essential for multilingual comprehension and information retrieval. The difficulty in accounting for the mental processes of encoding, decoding, storing, retrieving, analysis and synthesis becomes more evident when one tries to communicate with machines as is the case with machine translation. The computer programmers have also encountered several problems in constructing their natural language tools. At the lexical level, for instance, a single item may have more than one semantic interpretation corresponding to its semantic domain. The context is indispensable for accurate interpretation since it disambiguates meaning in certain contexts and adds meaning in others. A computer programmer collaborating with a linguist has to define all the LCs for a certain lexical item, i. e., to ensure that his framework can function in a manner similar to that of a human mind but this task is not easily manageable as we will clarify in the paper.

The objective of this paper is to create a framework of LCs for the Arabic language that has multi-purposes and applications. A sample of the findings of the application is attached to Appendix 1. Hamada (2006) specified one hundred examples of ambiguous instances for the Arabic texts and classified them into seven main types. The techniques she developed are applied to Arabic as well as to English to disambiguate texts. It is pertinent to point out that the collaboration among linguists and computer experts in different countries is a prerequisite for any successful and fruitful attempt for constructing Arabic Language Resources ALS. The paper falls into nine sections. Section one presents the introduction together with the objective; section two identifies the materials and methods applied to collect the linguistic data for analysis purposes; section three highlights the distinctive features of the lexical items; section four clarifies the linguistic ambiguity from a computational perspective, specifies the LC types, resolves the conflict between the LC components and elucidates steps for disambiguating texts; section five explicates the frame concept; section six presents and assesses the traditional disambiguating methods; section seven displays the proposed framework; section eight presents, discusses and assesses the application of the proposed framework and section nine rounds up the discussion with a conclusion.

2. The Materials and Methods:

The linguistic data that constitute the core of the paper are gathered from: a corpus of secondary school English language texts and a corpus of articles and press releases published by the Egyptian English newspaper Al-Ahram. Types of ambiguities at the different linguistic levels are pinpointed and analyzed.

3. The Distinctive Features Matrix:

Each lexical item has its idiosyncratic syntactic, semantic and phonological distinctive features. The syntactic features of a verb, for instance, specify whether the verb is transitive, intransitive, bi-transitive, etc., what form a verb takes (e. g. present, past, etc.). Furthermore, they identify additional semantic characteristics related to whether a verb, for instance, “die” has got something to do with other verbs that relate to the concept of “death” such as “kill”, “shoot”, “hang”; other added features may signify

the manner of killing such as “breaking the neck” or “hanging”; the tool that brought about death such as the gun, the pistol or the organ that is affected such as the neck, the leg and so on. In specifying the distinctive features matrices, one needs to consider:

3.A. The semantic change: as it is evident, some lexical items change their meaning in the course of time, others acquire an additional meaning. For instance, "leaves" in old English refers to a writing material but in modern English, it refers to the green parts of the tree that cover the branches, the "leaves of a tree".

3.B. The co-text: the near and distant sentences that highlight meaning. For instance, a sentence such as “A falcon can not hear the falconer”, literally means, in the “hunting domain”, that no falcon approaches his falconer no matter how long or hard s/he shouts at it”. But in the political domain, it means “the one who is in power does not care for the one who is not”.

3.C. The semantic field: lexical items share a semantically-related domain. For instance, the item “electricity” is a super-ordinate term that embraces all items related to electricity such as the “wire”, “switch”, “lamb”, etc.

4. The Proposed Framework of Constraints

4.1. Seven LC types are identified:

4.1.A. The grammatical LCs: these are the constraints that impede the generation of ungrammatical sentences. They do not allow one part of speech to be treated as another, for instance a preposition can not be treated as a noun. They also consider problems related to the idiomatic expressions and terminologies.

4.1.B. The morphological LCs: the available inflectional and derivational rules of a language.

4.1.C. The syntactic LCs: the syntactic constrains of a language that, for instance, do not permit the use of the past form of the verb after “to” as in “to ate”. They also include the agreement rules that do not allow a plural form of the noun to be followed by a singular verb or vice verse.

4.1.D. The semantic LCs: they embrace the componential analysis features of the noun such as: animate, inanimate, human, nonhuman, living, nonliving, etc. They restrain the generation of a sentence such as: “The trees keep thinking all the time” because only living human beings are capable of thinking. They also help the readers make deductions from the available data, e.g., it is common sense that spilling a colored liquid on cloths dyes the cloths with the liquid color. One, thus, infers that spilling ink on a bed sheet makes it blue though this item of information is not stated explicitly in the text.

4.1.E. The logical LCs: these constrains are determined by the culture, the physical and the social surroundings of the speakers of a certain language. They provide certain information items that disambiguate sentences. The information is related to logical parameters pertinent to the size, the time whether it is distant or near, the height whether it is too high or medium high, the capacity whether it is too wide or a bit wide and the heat degree whether it is boiling or warm.

4.1.F. The contextual LCs: these help the readers determine the appropriate contextual meaning. There are three types of contexts**:**

1. The sentential context which determines the relationship of the sentence elements within the sentence boundary

2. The supra-sentential context which determines the relationship of a given sentence to the neighboring sentences in the extended text: the pretext and post-text

3. The extra-linguistic context which refers to the external parameters that affect our understanding of the text such as our knowledge about the subject matter, the text producer, our background knowledge and experience, etc.

1. The sentential context**:** the stored and the retrieved data may help resolving ambiguity as in**:**

قال متحدث باسم الجيش الأمريكي في بغداد إنَّ حمود اعتقل في العراق.

"A spokesman for the American army in Baghdad said that Humud was arrested in Iraq". The stored data tell us that the American army in Baghdad is an occupying force, that Iraq is an occupied country, that Humud is freedom fighter from the national perspective and a terrorist from the American perspective and that "arrested" means that Humud is in detention now. Such contextual information help us interpret the post text**:** the sentences that follow which may either state that Humud still insists that he has done nothing wrong, that the American spokesman refrained from providing extra information, that Iraq is still far from regaining sovereignty, etc.

The concepts	The words that collocate with them
speaker	Provides information, announced, declared, made a statement
The one in detention	Committed a crime, suicide attempt, fight, activist, operative, terrorist, freedom fighter,
An occupied country	Under occupation, sovereignty, resistance, army, liberation,
The occupying enemy	Occupation, suppression, oppression

Table (1) clarifies the concepts and the words that collocate with them

1. The supra-sentential context**:** the lexical and syntactic constraints play an important role in disambiguating sentences as in the following example from Arabic **ذهب أيمن إلى المدرسة** Ayman went to school". The Arabic form of the verb "went" could either be a noun meaning "gold" or a "verb", the past form of "go". If it is a noun followed by Ayman, it would be a genitive construction which could not be followed by a prepositional phrase. Thus, **ذهب** could not be interpreted as a noun; the lexical frequency component will also reject such interpretation. It is thus classified as a verb but in the sentence, **ذهب أيمن براق.** "Ayman`s gold is glittering", the context will indicate that "gold" is a noun since it is followed by the noun modifier "glittering" which is considered in Arabic as a "nominal predicate". The lexical constraints accept "glittering" as a noun modifier modifying "gold" rather than Ayman. If it qualifies Ayman, the grammatical rule will accept "go" as a verb as in ذهب أيمن ضاحكًا "Ayman went laughing" whereby the participle qualifies the verb. One of the identifying characteristics of nouns is their collocational patterns. In the sentence: **شرب البدويّ من ماء العين.** "The man drank from the well water", the Arabic equivalent of the English "well" is ambiguous; it could either mean "a human eye" or a "well" among other meanings. The words that collocate with "well" in the sentence resolve the ambiguity, namely "Bedouin" and "drank".

The structure of a sentence may also contribute to resolving ambiguity whether or not a modification or a conjunction structure is used as in

English sentence	Function	Arabic sentence
I saw an eye full of tears.	Description	رأيت عينًا دامعة.
I saw the eye, the mouth and the lips.	Conjunction	رأيت العين والفم والشفاه.
3. I saw the well and the beautiful oasis that surrounds it.	Conjunction	رأيت العين والواحة الجميلة حولها

3. The extra-linguistic context**:** to resolve an ambiguity in a sentence, one may also consider the extended context. The Arabic sentence below is ambiguous and the ambiguity could not be resolved by considering the extra-linguistic parameters.

English sentence	Arabic sentence
"I was introduced to the owner of the dog that was terrible"**.**	**تعرفت على صاحب كلب مرعب.**

The Arabic equivalent for "terrible" could either modify "the man" or "the dog". One may either resort to one's background knowledge, common sense, and cosmopolitan knowledge or to the post text. If the following sentence is "it was barking furiously", then it becomes clear that reference is made to the dog but if the following sentence is "he was wearing black glasses" or "he was frowning", then reference is made to the owner himself. The text domain may also help resolve ambiguity. In the sentence below, if the text is medical, the Arabic equivalent of "the operation" means "a medical surgery" but if the text deals with engineering, it then means "an engineering or a mathematical process" and if the text is legal, then it means a" criminal act".

English sentence	Arabic sentence
"Ali performed the operation".	**أجرى أيمن العملية.**

4.1.G. The collocation constraints: collocations are the concomitant lexical items that always collocate with each other. The collocation constrains can help disambiguating meaning. The word “bank” for instance is ambiguous but if the text under consideration deals with commercial transactions, i. e., if the word "commercial" collocates with "bank", or if the text discourses on checks, invoices, bills, financial statements, then the reference is definitely to the commercial institution. If, on the other hand, the text deals with farming, i. e., with trees, water, plants, then the reference is to the river bank.

4.2. The statistical sub-component: the language data are analyzed to determine the frequency of each linguistic item. The specified items are then ordered in terms of their frequency of recurrence; this is referred to as the frequency rate FR. For instance, an ambiguous word such as “bank” could either mean “the river bank”, “bank1”, or “the financial institution”, “bank2”. The statistical sub-component assigns the first meaning “bank1” first because it has the highest FR. If the word can not be processed, the next alternative will be assigned: the one that is lower in frequency, “bank2”.

4.3. Resolving the LCs conflict:

The LCs may conflict with each other since they assign different interpretations but this conflict is soon resolved by resorting to other constraints. For instance, in the ambiguous noun phrase, "the pregnant women and children”, the adjective "pregnant" could either modify "women" alone or both "women and children" but it is common sense that only women can be pregnant. The semantic constrains then negotiate with the logical constrains to resolve the ambiguity: to provide one interpretation: that "pregnant" modifies only "women" rather than both "women and children”.

4.4. Steps that guide a programmer in the disambiguation process: a programmer and/or linguist should:

4.4.A. Collect all the possible sources of ambiguity from a corpus that covers types of texts written by almost all famous authors. This is the linguist task guided by a computer expert.

4.4.B. Identify the source of each ambiguity and the LCs that are used to account for it. This task could be shared by both the linguist and the computer expert.

4.4.C. Count the frequency of the collected lexical items. This is the computer expert task.

4.4.D. Manipulate the LCs in a manner that is suitable for computerization in terms of rules and frames. This is the core of the present paper. This is also

the computer expert task.

4.4.E. Utilize the specified contextual types. This is the task of the computer expert.

4.5 The Frame Concept:

A language expresses the different types of meaning in terms of distinctive features. A concept frame is just a matrix representing these features. It incorporates all the features a concept represents. The frame may be linked to another super-ordinate or subsumed one (Hamada, 1998). So any sentence of the language can't be linked to any feature of the concept frame as in the following**:**

Concept	Link	POS	Def.	Uses	Volume	Time
Sun		**noun**		• **Gives heat.** • **Helps in photosynthesis**		**morning**

Fig (1) a concept frame for "the sun" concept

A Sentence	**Distinctive Features**
The sun is bigger than the earth.	**Volume** feature
The Sun rises in the morning.	**Time** feature
The sun gives us warmth, light.	**Function** feature
The sun is shining up there.	**Function** feature

4. The Shortcomings of the Traditional Approaches:

The traditional methods define features that are not exhaustive in their nature and that are primarily lexical, e.g., they specify the verb tense (past, present or future), the verb type (intransitive, transitive, bi-transitive).

5.1.A. Approaches to anaphora resolution include:

The activation concept (Kantor) is introduced by Graeme (1981) who examines the factors affecting the pronominalization of a concept. The task oriented dialogue (Grosz) is advocated by Yang (2000) *et al*, who utilize the priori knowledge of the discourse structure to resolve the pronominal reference. Frames such as focus (Sidner) are recommended by Graeme (1981) who applies the discourse clues to select a frame from the knowledge structure and to act as a focal point. Graeme (1981) also opts for predicate calculus as a representation to handle problems related to quantification for reference resolution. This is referred to as the logical formalism (Webber). Furthermore, he (Ibid, 1981) adopted a discourse approach focusing on pronominal reference and examining the cohesive ties in a text. For resolving comprehension ambiguities in Arabic sentences, Othman, *et al* (2002) specify a number of semantic features for the lexical entry of verbs, nouns, and participles. These are:

5.1.B. The verb has a stem with the syntactic and lexical features: the syntactic features specify the voice, tense, transitivity and the subject-gender, subject-number, object-gender and object-number agreements. The lexical features determine subject-rationality, object-rationality, etc.

5.1.C.A noun has a stem with the syntactic and lexical features: the syntactic features provide information concerning the definition, gender, number, end-case, irregular-plural forms, etc; and the lexical features specify information concerning the divisibility, category and rationality.

5.1.D. A particle has a stem with certain specific features

5. The Lexical Constraint Guide and its Applications:

This proposed framework tackles most of the drawbacks of the traditional methods since it specifies linguistic constraints at different levels. The context directs the text analyst; if the lexical constraint guide LCG comes across a word that can be considered a noun, the syntactic LCs then specify the next lexical item which may be a noun or a verb. They, thus, guide the analyst to choose a certain verb assisted by the FR. The system will then determine that the next word is a verb. If it is not so, the LCG suggests a noun. The system then re-consults the syntactic LCs to find out the construction that consists of a noun + noun.

7. The Research Findings, Discussion and Assessment:

7.1.The LCs can be used to account for the following ambiguities:

7. 1. A. The part of speech POS ambiguity:

The verbs: the verb "receives", for instance, may be followed by "a guest" or a "gift". The syntactic LCs indicate that the verb may be followed by "someone" or "something" as follows:

(receives) + someone --> means "welcome".
(receives) + something --> means "get".
The adjectives and their modified nouns: the adjective "old" may refer to "age" or "antiquity"
Adj (old) + Noun (living creature) --> refers to "age".
Adj (old) + Noun (a non living creature) --> means "ancient".
The uninfected words: the word "so" may have more than one meaning.
So + Adj …. --> means very. Hope + so --> that thing. So + sentence + …. --> Consequence

7.1.B. The modification ambiguity:

An adverb that qualifies a verb in a sentence that has more than one verb: if the two verbs have the same tense form as in:

Ex:- We met the man you told us about yesterday.

"Yesterday" may refer to the "meeting" or the "telling". The extra-linguistic parameters of the communicative event and the contextual LCs are supposed to resolve this ambiguity but if one verb has a different tense from the other as in:

We will meet the man you told us about tomorrow, the syntactic LCs of the verb form "will meet" will resolve the ambiguity since they agree with the adverb "tomorrow" in reference, i. e., the "meeting" will be tomorrow.

7.1.C. The referential ambiguity:

a- A pronoun in a sentence may have more than one reference.

Ex: a-1 The monkey eats the banana because it was hungry.

a- 2 The monkey eats the banana because it was delicious.

The syntactic structure does not specify whether the pronoun "it" refers to the subject "monkey" or to the object "banana". In a-2, the syntactic LCs which provide information related to the character of the noun and its modifiers, restrict the reference of "it" to "banana".

1- Adj (a living creature, […, "hungry", …]).

2- Adj (food, […,"delicious", …..]).

Underlying the complex sentence a-1 which consists of a main clause and a subordinating one, there are two simple sentences; the first one is: "the monkey eats the banana" but the second one has two semantic interpretations: The monkey was hungry. The banana was hungry. As evident from the database, only the first matches reality.

Another example would be: a- The monkey touches the wire and it electrified. Underlying this compound sentence, there are two simple sentences conjoined by "and": "The monkey touches the wire." and "The wire electrified the monkey." In fact, the second sentence has two semantic interpretations:

1- The monkey electrified the wire. 2- The wire electrified the monkey.

Concept	Subject	Object-	Subject-adj.	Object-adj.	Results-	Secondary-results
electrified	electricity's-family	a living creature-		burning	light, fire	

Fig (2) The Frame of the Verb "electrified"

7.1.D. The prepositional phrase reference ambiguity:

John bought books for the children.

The prepositional phrase "for the children" could either mean "the books are written about children" or "the books are written for the children to read". This ambiguity can be resolved by recalling the extra-linguistic parameters. The sentence: "Ali buys books for money" has also two semantic interpretations. The statistical constrains will exclude the second one, (Jiri, *et al*, 1999).

7. 1. E. The lexical ambiguity:

For the sentence "Ahmed draws the table and fills it with arguments", the lexicon provides two semantic representations for "table":

table table1 = furniture and table2 = chart

The construct distinguishes meaning relying on the "entire word collective features". It will collocate words such as "chair, house, eat,.." with "table1" and words such as "draw, numbers, print," with "table2".

7.1.F. The syntactic ambiguity:

In the sentence "The mother of the bride with the white dress saw us", the prepositional phrase "with the white dress" could either qualify "the mother" or "the bride". By referring to the "entire word collective features", it will become apparent that the bride wears the white dress. These features can also resolve ambiguity by considering the relation between the adjective and its modified noun as in the case of the noun phrase "a metallic rose vase", that has two noun heads: the "rose" and the "vase". The syntactic LCs delimit the reference of adjective "metallic" to the "vase".

7. 2. The detection of illogical constructions:

The LCs and the statistical sub-component do not only resolve ambiguity but they also state whether or not two concepts are related to each other. For a sentence such as the one in the box, the system responds:

User: Mohammed ate the bed.
System: This is wrong; Mohammed can not eat the bed.

Fig (2) The interface screen

The system analyzes a sentence such as 'Mohammed ate the bed' as follows:

A) The system looks for the object of the verb 'eat' which is 'food' from the relation *obj*. Obj ("eat- human", "food").

B) The system tries to revoke 'the bed' from the 'food' chain. It will find that "the bed" is not in the chain so it responds as above in.

7.3. Helping text recipients infer new information not stated explicitly in the text:

This is realized by invoking the "verb-result" restriction as in "Huda ate rice". The semantic analyzer will analyse "Huda ate rice" giving the following output:

- Rice is a carbohydrate --- from the "kind" feature.
- Rice is a part of food ---- from the "link" feature.
- Positive results of eating rice:- ---- from the Positive result features.

1- Feels good, 2- Being healthy, 3- Having energy.

The negative results of eating rice: ----- from the Negative result features: Being fat.

From the sentence, "The station bell is ringing" one may infer the source of the noise**:** that the train is speeding up and that no one can ride it.

Concept	Subject	Results	Secondary results	…
ring	station bell	ringing noise	no one can ride the train	

Fig (3) The Concept Frame of the Verb "ring"

7.4. Functioning as a grammatical checker:

An incorrect sentence such as "He came to ate" is automatically corrected to: He came to eat.

8. The Application of the Proposed Framework:

By applying the Prolog Language on the PC computer to:

1- The monkey electrified the wire. 2- The wire electrified the monkey,

the mapping of the frame 'electrified' to sentence (1) does not verify the frame whereas the mapping to sentence (2) does, so the system will accept S2 and reject S1.

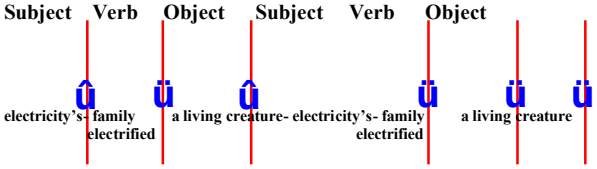

Fig (4) The Frame Mapping of the Sentences

Sentence 1 "unification in UBG will fail". Sentence 2 "unification in UBG will succeed"

1] Example: If the system input is: "Mohamed eats the bread.", the system can infer a new information item from the given facts:

is-a ("human", "Mohamed")

is-a ("bread", "carbohydrate") is-a ("carbohydrate", "food").

After mapping the frame of "eat" with a "human" subject and "carbohydrate" as an object; the result ("eat", "carbohydrate", ["became fat", "feel energy",…]). The result is the list of [become fat, feel energy, …].

Results/ System output: Bread is a carbohydrate food.

Mohammed will become fat or Mohammed will feel energetic.

2] Given the following input

Example: *System input* The tea is spilled on the bed sheet.

Example: *System input* The ink is spilled on the shirt.

Programming Steps:

1- Search types of ink, tea, shirt and bed sheet.

From fact; type-of ("tea", X), type-of ("ink", X). X will be a "colored liquid". Type-of ("bed sheet", Y), Type-of ("shirt", Y) and Y will be a "cloth".

2- Search verb "spill" for the results with "colored liquid" as a subject and "cloth" as an object. Result (spill, X (colored liquid), Y (cloth), ["The", Y "will have the " Z (colored liquid) color]).

Color ("ink" (Z), [all (W)]). Color ((Z)"tea", [red(W)]).

Results /System output: the system will infer the information:

The bed sheet (Y) will have the red (W) color.

8. 1. The framework assessment:

By applying the Prolog Language on the PC computer, the system produces the following results:

Example (1); the system input: "Mohamed eats the bread." The system can infer a new information item from the given facts:

is-a ("human", "Mohamed")

is-a ("bread", "carbohydrate") is-a ("carbohydrate", "food").

After mapping the frame of "eat" with a "human" subject and "carbohydrate" as an object; result ("eat", "carbohydrate", ["became fat", "feel energy",…]). The result is the list [become fat, feel energy, …].

Results/ System output: Bread is a carbohydrate food.

Mohammed will become fat or Mohammed will feel energetic.

Example (2): *System input* The tea is spilled on the bed sheet.

Example (3): *System input* The ink is spilled on the shirt.

Programming Steps:

1- Search types of ink, tea, shirt and bed sheet.

From fact; type-of ("tea", X), type-of ("ink", X). X will be a "colored liquid". Type-of ("bed sheet", Y), Type-of ("shirt", Y) and Y will be a "cloth".

2- Search verb "spill" for the results with "colored liquid" as a subject and "cloth" as an object.

Result (spill, X (colored liquid), Y (cloth), ["The", Y "will have the " Z (colored liquid) color]).

Color ("ink" (Z), [all (W)]). Color ((Z)"tea", [red(W)]).

Results /System output: The system will infer the information item:

The bed sheet (Y) will have the red (W) color.

When more examples are given, the system responds as follows:

1- Input: Give an analysis for the following sentence:

The monkey touches the wire and it electrified.

Output: The monkey touches the wire. The wire electrified the monkey.

2- Input: The monkey eats the banana because it was delicious.

Output: The monkey eats the banana. The banana was delicious.

3- Input: The monkey eats the banana because it is hungry.

The monkey eats the banana. The monkey is hungry

8.2. Applying the statistical sub-component:

In appendix 1, an output for the construct that depends on the statistical constrains of the collocation words is presented. Although it is just one constrain among many we have presented, it verifies the presented results for our corpus. It is not possible to specify all the language constrains of the lexical items but it is our conviction that we have demonstrated the validity of our framework.

9. Conclusion

This paper is an endeavor to promote Arabic language technology. It introduces a linguistic framework that helps in solving many textual analysis problems. The constraints that constitute the core of the framework are based on the componential analysis theory, the distinctive feature matrices and the frame concept represented by Nida among others, (2003). The paper specifies some additional distinctive features that could enrich the linguistic resources. The proposed construct is applied for disambiguating various types of texts: narrative, descriptive, expository conversation or argumentative at the different linguistic levels: lexical, morphological, sentential and supra sentential. In certain contexts, it invokes some extra linguistic parameters that are necessary for the accurate interpretation of utterances. The proposed framework has been tested and has proved to be capable of resolving different types of textual ambiguities.

Appendix 1 :The Disambiguation System

An Application for the Possible Interpretations of the Verb "play"

References

Graeme, Hirst, 1981. "Discourse-Oriented Anaphora Resolution in Natural Language Understanding: A Review", the American Journal of Computational Linguistics, Volume 7, Number 2, April-June

http://ucrel.lancs.ac.uk/acl/J/J81/J81-2001.pdf

Hamada, Salwa, 1998. "Natural Language Processing as applied to Arabic Texts", Faculty of Engineering, Ain Shams, Egypt

Hamada, Salwa, 2006. "Ambiguity Types in Human Languages", the fifth conference of the Language Engineering Society, Cairo, Egypt, 14-15 September

John, Justeson and Suny, Albany, 1993. "Principled Disambiguation: Discriminating Adjective Senses with Modified Nouns", http://ucrel.lancs.ac.uk/acl/J/J95/J95-1001.pdf,.

Nida, E.A, 2003. "Componential Analysis of Meaning", Mouton Paris, http://www2.tltc.ttu.edu/schneider/pub/cluster_qual_qant.pdf,

Othman, Eman; Shaalan, Khaled and Rafea, Ahmed, 2002. "Towards Resolving Ambiguity in Understanding Arabic Sentences" http://www.claes.sci.eg/publication/document_view asp?doc_id=200&File_Name=ambiguity_resol_nemlar.pdf4.

Papadopoulou, Despina and Clahsen, Harald, 1999. "The Role of Lexical and Contextual Information in Parsing Ambiguous Sentences in Greek" ,http://citeseer.ist.psu.edu/cache/papers/cs/27457/http:zSzzSzprivatewww.essex.ac.ukzSz~haraldzSzgreek-L1.pdf/the-role-of-lexical.pdf.

Stetina, Jiri and Nagao, Makoto, 1999 "Corpus Based PP Attachment Ambiguity Resolution with a Semantic Dictionary", http://acl.ldc.upenn.edu/W/W97/W97-0109.pdf.

Yang, Fan; Heeman, A. and Hollingshead, Kristy, 2000. "Towards Understanding Mixed-Initiative in Task-Oriented Dialogues", http://speech.bme.ogi.edu/publications/ps/yang04.pdf