ISTANBUL UNIVERSITY –
JOURNAL OF ELECTRICAL & ELECTRONICS ENGINEERING

YEAR : 2005 (1341-1352)
VOLUME : 5
NUMBER : 2

DETERMINING EFFECTIVE FEATURES FOR WORD SENSE DISAMBIGUATION IN TURKISH

Zeynep ORHAN[1] **Zeynep ALTAN**[2]

Istanbul University, Engineering Faculty, Computer Engineering Department
34320, Avcilar, Istanbul, Turkey

[1]E-mail: zeyneporhan@yahoo.com [2]E-mail: zaltan@istanbul.edu.tr

ABSTRACT

Word sense disambiguation is necessary or at least helpful for many natural language processing applications. This paper deals with the feature selection strategies for word sense disambiguation task in general for all types of words in Turkish language. There are many different features that can contribute to the meaning of a word. These features can vary according to the metaphorical usages, POS of the word, etc. The observations indicated that detecting the critical features can contribute much thanthe contribution of using various current learning methodologies.

Keywords: *Word sense disambiguation, feature selection*

1. INTRODUCTION

The tasks, either physical or mental, that are carried out by humans are attracting the attention of scientists from the early times of history, since exploring the details of these tasks may lead to many improvements in various applications. One of these interesting tasks is the language processing and it has been the topic of many philosophical discussions for centuries. However, linguists, philosophers, or other experts dealing with this issue could not find a general model of language processing that has been widely accepted in the scientific community.

Received Date : 14.02.2005
Accepted Date: 11.06.2005

The humans are bestowed with the incredible capability of understanding or interpreting the given text or discourse immediately. Although it seems to be a very simple task, the details of this mysterious event are not so clear. There are many striking processes that must be considered seriously in the human understanding system. First of all, this task is completed in a very short period of time. Sometimes even reading or hearing a few words or sentences from a very long text or discourse might be enough for understanding the whole subject. People can understand, or at least get the main point without having the full information about the topic and can fill in some parts that are missing. Another important fact is that, the natural languages do not have a static structure and they are subject to changes by the time, but humans can also deal with the dynamicity of the natural languages successfully.

When we consider the children who acquire this ability at the very early ages, the task can be thought to be a very simple one that can be simulated by machines. However, natural languages are inherently ambiguous and difficult to be processed by computers. This is one of the main reasons that enforced us to interact with the computers mostly by the help of the standardized tools. In other words, we have to communicate with the computers by typing texts via keyboards, clicking the mouse, using graphical user interfaces or touch screens and so on. These tools are somehow the best solution that can be offered for the time being, but it will be very nice and, of course, more preferable to have man-machine interaction by using natural languages, especially the native language of your own.

Obviously, having a better communication with the computers is not the major reason behind the researches about natural language processing (NLP). There are many other application areas such as text summarization, translation, information extraction, speech analysis and synthesis and so on where we need to understand text and discourse. However, it is not possible to have this opportunity now and it seems that it will not be possible in the near future without recognizing the complexity of the human languages.

George Miller [26] expresses this fact in a very nice way with the following words: "Most people are unaware how vague and ambiguous human languages really are, so they are disappointed when computers fail to understand linguistic communication. … As more and more documents are stored in computers, the machines' inability to understand the information they hold restricts their usefulness to both business and government. Computers are not to blame for this situation. Language itself is at the root of it. "

The words are generally the building blocks of the natural languages, therefore understanding the issues related to the words are important for the other tasks that depend on the words. Although, in the literature there have been many discussions on what a word or its meaning is, it is really difficult to give a concise definition for the word or its meaning. In the history language is thought to be nomenclature and this view has been very influential on Western thought for a long time. Nomenclaturism has been taken from the Bible sources and from Plato. Names for things were accepted as the names given or explaining the natural properties for the things they belonged to. The reason for a word having a meaning is that word's being a representative for something else in this view. In Saussure period [35], this view has been criticized and the approach of words standing for an idea in the mind has been offered against words standing for something which independently exists in the world, i.e. nomenclaturism. One of the interesting approaches in this field was from Wittgenstein [8, 41]. He suggests that different types of word/meaning relationships correspond to different types of games[1]. He established an analogy between the word meanings and the game types relationships. He claimed that there are many different types of relationships between word and meaning similar to many different types of games. Moreover, it has been strongly emphasized that, there is no single characterization about the word-meaning relationship. Same fact can be observed when we tried to find a single set of properties that we can give a game definition. In this view it has been proposed that the meaning is nothing other than a way we have collectively decided to speak and write. Since it's a by-product of cooperative work and meaning is nothing more than a shared language-game. Therefore what we need is to clarify the way we use ordinary language.

Many different views about the word senses exist in the history. Aristotle adopted the word object relationship and ignored context. On the other hand, Wittgenstein [41] claimed that the usage was important rather than the meaning of the word and assumes that there are no predefined word senses, but usage of a word in a context determines meaning. Bloomfield [5] claimed that there is a strong relationship among the meaning and the frequencies of words. Alternatively some

[1] Game Theory : The meaning of a word or phrase or proposition is nothing other than the set of (informal) rules governing the use of the expression in actual life. Like the rules of a game, these rules for the use of ordinary language are neither right nor wrong, neither true nor false. They are merely useful for the particular applications in which we apply them.

believed that meaning of a word was related with the syntagmatic use in discourse [3].

1.1. Word Sense disambiguation

Word sense disambiguation (WSD) has been affected from these thoughts on word and meaning to a great extent. Even defining what this task is very difficult due to many different views about the word and the meaning. Generally it has been described as the task of assigning the most appropriate meaning to a polysemous word within a given context. When we are dealing with the complexity of languages, one of the difficulties we have to consider is the syntactic or semantic lexical ambiguity which is an important problem at the bottom level of NLP applications. WSD is an essential and unavoidable task especially for the language understanding systems and it is necessary or at least helpful in one way or another for many other applications as an intermediate step [20].

The process of sense disambiguation can be given as follows: The WSD programs take natural language sentences as the input and the expected output of such programs are the assignment of sense tags. These tags are assigned to the ambiguous words and generally senses are taken from a dictionary or another similar source such as Longman Dictionary of Contemporary English (LDOCE) [6]) or WordNet [16]. WordNet is the commonly used lexical database and in many wsd researches it is the primary source for obtaining the set of applicable senses for the ambiguous word. It is really a wide-coverage, public-domain dictionary containing about 95,000 English word forms and refined sense distinctions for words. WordNet includes semantic taxonomies to provide a rich source of knowledge for NLP applications. Word senses does not include independent entities due to several semantic relations connecting these entities. Considering the advantages of such a resource, there are many efforts for building similiar resources for the other languages [37].

Words in natural language are known to be highly ambiguous. The words that have high frequencies are generally the words that have higher number of senses on the average. In the Wordnet dictionary, the average number of senses for each noun for the most frequent 121 nouns in English is 7.8. This average is considerably higher, strictly speaking it is 12.0, for the most frequent 70 verbs supporting the ideas that the most ambiguous words are the verbs[29]. The important point that has to be emphasized at this point is This set of 191 words is estimated to account for about 20 percent of all word occurrences in any English free text. Therefore WSD is a difficult and hard to master task in NLP. However, once it has been achieved it can be useful for many other tasks.

Senseval-1 in 1998 [22] was a cornerstone in wsd researches WSD has received growing attention from the Natural Language Processing community. Applying wsd in various fields lead to promising results such as the improved precision by about 4.3 percent [36] on part of the TREC corpus[2] and the better quality of machine translation by WSD techniques [13]. These works are good indicators of the utility of WSD in practical NLP applications.

WSD has many difficulties when its all aspects are considered deeply. Defining word senses as the mental representations of different words is a reasonable assumption. However, we do not know much about the mental representations, since it is hard to design experiments to learn what they really are. Humans can categorize word usage intuitively, but the agreement is not so high. Using dictionary definitions for tagging senses is an alternative way, but it is only helpful for nonuniform distributions where one sense is highly distinguishable. Additionally, definitions can often be vague. When we considered the frequencies of the words, it is observed that the higher the frequencies, the higher the disagreement rate is, so selecting words based on frequency would bias results. In text or discourse, it is common for humans to have a simultaneous activation of different senses, so this may lead to high levels of disagreement. For example the proper nouns may cause problems, e.g. Gül (it can be a name, surname, or rose), Kara (surname, military group, land, black) etc. Clustering approaches use the strategy of taking only course-grained distinctions among word senses (e.g. the ones that may not be ignored in translations across languages) into account.

The early work on wsd concentrated on hand coded knowledge ([21], [19]). However, this can be laborious and time consuming. Additionally,

[2] TREC is a standard information-retrieval test collection

manual systems always suffer from the scalability. The alternative to this approach is the corpus-based methods that became more popular due to these problems. Machine-learning techniques are used to automatically acquire disambiguation knowledge. Sense-tagged corpora and large-scale linguistic resources such as online dictionaries are the fundamental components of a typical wsd system.

1.2. Feature Selection

The set of features that is to be used has to be considered in the early design phases of a wsd system. Finding the appropriate set of features is crucial. In many artificial intelligence applications features were studied carefully and variable sets of features have been successfully used. Nowadays, the trend is for automatic feature selection. In [1] feature selection has been achieved by using searching algorithms in the domain of cloud types classification and increased performance was obtained. Many different efficient search algorithms on several synthetic datasets for the detection of optimal feature subsets were used in [28]. The description of a linguistic and cognitive biased approach that considered the application of instance based learning with automatic feature selection for relative pronoun resolution can be found in [10]. Another interesting work is about selection of different features for each instance in the training set. This was achieved by using a context sensitive feature selection algorithm [14]. Decomposable probabilistic models plus Naive Bayes algorithms for the same issue can be found in [7]. A lazy learner that used automatic feature selection has been presented in [25] with improved WSD results.

Surrounding words in a given window size and their part of speech (POS) [7], keywords [29] or bigrams in the context [34] and various syntactic properties [17] etc are some possible candidates that can be employed in WSD. Various kinds of feature representing different knowledge sources have been used in supervised WSD research. The ones that are included in [30] are surrounding words, local collocations[3] ([21], [42], [44], [29]), syntactic relations, POS and morphological forms ([29], [6]). The wide variety set of features given in [25] are current ambiguous word, current part of speech, contextual features (the words and parts of speech of K words surrounding current word), collocations formed with maximum K words surrounding, head of noun phrase, sense specific keywords (maximum MX keywords occurring at least MN times), bigrams (maximum MX bigrams occurring at least MN times) are determined for all training examples, the verbs, nouns, named entities, prepositions, pronouns, determiners before and after the ambiguous word. Among these, the most effective features were selected as the current ambiguous word, current part of speech, contextual features and collocations which are also the features most frequently mentioned in the literature.

Weaver [40] in his Memorandum states the necessity of WSD in machine translation and emphasizes the basic ideas of the WSD approaches from then on as follows:*" If one examines the words in a book, one at a time as through an opaque mask with a hole in it one word wide, then it is obviously impossible to determine, one at a time, the meaning of the words. [...] But if one lengthens the slit in the opaque mask, until one can see not only the central word in question but also say N words on either side, then if N is large enough one can unambiguously decide the meaning of the central word. [...]The practical question is: "What minimum value of N will, at least in a tolerable fraction of cases, lead to the correct choice of meaning for the central word?"*

This view emphasizes the importance of N neighboring words, however the decision function does not solely depends on those N words in human language processing system. These N words are necessary but not sufficient component of the process.

We believe in that the solution of all these discussions and the WSD process is understanding the human language processing system. Today we know only a little about this system. Psychologists say that when a human hears or reads an ambiguous word, more than one senses of the word can be activated initially.

[3] A local collocation refers to a short sequence of words near w, taking the word order into account. Such a sequence of words need not be an idiom to qualify as a local collocation. Collocations differ from surrounding words in that word order is taken into consideration.

But in a very short period of time disambiguation takes place in the human brain and that person chooses the correct sense in the given context. There must be a mental representation reflecting the properties of the given context. This specific representation includes a linguistic component, a situational context and general knowledge components. The structure of this representation must be solved and coded for machines to implement fully automated word sense disambiguators, or other understanding related components.

These type of problems are not only a research are of linguistics, but also psychology, cognitive science, computer science, etc. Since this an interdisciplinary problem, having linguistic theories about human language processing system will not be enough, and problems related to computations and evaluation of the linguistic models and specific language structures must also be considered.

2. Application

In this section information about the data and feature selection issues will be examined.

2.1. Corpus

Computational methods improved in parallel to the electronic devices. Computational linguistics has been affected from these improvements positively. Turkish has very limited resources that can be used for these purposes, although there are many electronic resources available that can be used for devising and testing different models. Electronic dictionaries are available for Turkish, but they have many inconsistencies. Parsers, morphological analyzers and some other tools for Turkish language processing have been developped in recent years. However, some of them do not have a broad coverage or some others are not open to public. There are some ongoing projects for providing data for NLP applications in Turkish like METU Corpus Project[4][12].

Supervised methods are generally employed in these types of applications. We need corpora for training and testing the computer programs. There are available corpora for languages like English, Japanese, Spanish, etc especially in SENSEVAL[5] [8] project, but there is no corpus for Turkish yet. We are trying to build a corpus which has been sense tagged for a limited number of verbs and nouns. Evaluation process has the same type of problems; we need standardized corpora and methods for evaluation.

METU Turkish corpus has become available for academic purposes. It has two parts one is the main corpus and the other one is the treebank that consists of parsed, morphologically analyzed and disambiguated sentences selected from the main corpus. We have preferred to use the treebank part of the corpus for our purposes. Not all the sentences in the corpus are included in the treebank, since the disambiguation process has to be achieved manually. The sentences are given in XML format and provide many syntactic features that can be helpful for disambiguation. On the other hand, it has some disadvantages: First of all it is the first time this treebank is open to academic researches, naturally, there are some errors and inconsistencies that have to be corrected manually in the treebank. Additionally, only the syntactical annotation has been provided meaning that manual sense tagging has to be completed.

The texts in main corpus have been taken from different types of Turkish written texts published in 1990 and afterwards. It has about two million words. It includes 999 written texts taken from 201 books, 87 papers and news from 3 different Turkish daily newspaper. The distribution of texts are shown in **Error! Reference source not found.**. They have used XML and TEI(Text Encoding Initiative) style annotation and tried to obtain a corpus similiar to BNC(British National Corpus)

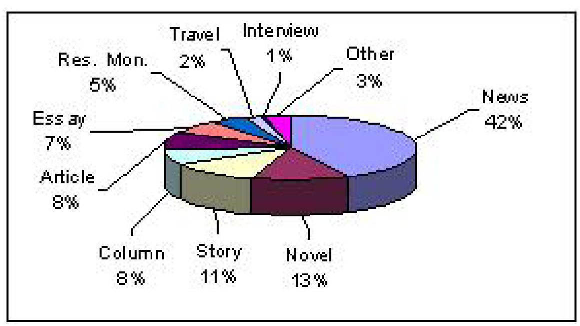

Fig 1: Distribution of texts in the corpus

[4] METU Turkish Treebank, http://www.ii.metu.edu.tr/~corpus/indextr.html

[5] SENSEVAL, Evaluation exercises for Word Sense Disambiguation Organized by ACL-SIGLEX, http://www.senseval.org

There are 6930 sentences in this treebank. These sentences have been parsed, morphologically analyzed and disambiguated. In Turkish, a word can have many analyses, so having disambiguated texts is very important. Additionally, some of the collocations, idioms etc. have been preprocessed.

The distribution of the texts is similar to the main corpus. We have examined the frequencies of the words as it is necessary to select appropriate ambiguous words for WSD. There are 5356 different root words and 627 of these words have 15 or more occurrences, and the rest have less. Therefore, most of the root words are so rare and not suitable for WSD experiments.

2.2. Affective Features

The disambiguation process is a mapping function from a set of given features plus our general knowledge to the senses of the given word. The mapping function is very sensitive to the selected features, and therefore precision and recall can be increased/decreased depending on the features that are going to be used. One possible feature can be collocational words (e.g. *hoş* in hoş geldiniz-wellcome or *karşı* in karşı geldi-he opposed). Other types of features can be the affixes, syntactic categories of the words preceding and succeeding the target word, POS etc. In our outgoing project we are trying to determine these effective features in WSD for Turkish.

There are many different features that can contribute to the meaning of a word. These features can vary according to the metaphorical usages, POS of the word, etc. In the following examples, some of these features ar emphasized:

Aklına bir soru **geldi** (A question arose to his mind)
Elimizden geleni yaptık.(We did our best)
Bebek artık **ele gelir** oldu. (The baby has grown enough (has gotten bigger))
Bağlama**yı eline aldı**. (He took the instrument (baglama: 3 stringed instrument))
Konu**yu ele aldı**. (He handled the subject (the matter))
Bir yar sevdim **el aldı**. (I loved someone and a stranger took him/her away from me)

All the syntactical features that can be selected from the corpus are provided in **Table 1.** The features that are observed as the most affective are indexed in the last column. Then we have selected a set of ambiguous words and their senses. These words and their senses along their distributions are given in Table 2. For each word we have devised 17 experiments that have included different features. The features included in each of these 17 sets were shown in Table 3.

3. Experimental Results

Considering the close relationship and analogies in the problem domain between machine learning and NLP applications, machine learning algorithms were used in WSD task. There is a system so called WEKA[6] [8] developed at the University of Waikato in New Zealand. The implementations are in Java which is compatible with our previous applications and includes many famous machine learning algorithms. The system provides many visualization tools and a detailed analysis of the output.

We have tried various machine learning algorithms and compared their results on different sets and observed that the performances are not fluctuating dramatically for different algorithms. However the results are very sensitive to the features employed. Therefore, we have tried all the experiments by AODE method which is an improved version of famous Naïve Bayes algorithm. The results of the tests are provided in Error! Reference source not found..

It is important how to interpret these results. First of all, the distributions are important. The most frequent sense is the baseline for evaluation. The results below this are not acceptable. The results above have to be considered relative to the distributions. If the senses are not uniformly distributed in the test sets, the improvement may vary from word to word. Another important point is the fluctuations in the test sets. The results are really very sensitive to the features. For the first word the results are changing from 37.66 through 90.90. Similar deviations are observed for the other words. The best results are not always obtained for the same set of features. Additionally using many features did not result in more accurate results. These observations indicate that selecting a sufficiently large set of features is critical and this set has to include only the necessary features but not more or less.

[6] WEKA: Data Mining Software in Java, http://www.cs.waikato.ac.nz/~ml/weka/index.html

Table 1: Features selected for experiments

Variable Name in arff file	Used for	Value	Index
Tümce no	Sentence number	2	
Dosya no	File number	0000221313.xml	
@attribute onceKok	Previous word root	BURA	F1
@attribute onceTur	Previous word POS	NOUN	F2
@attribute onceTuretme	Previous word inflected POS	NOUN	
@attribute onceHalEki	Previous word case marker	DAT	F3
@attribute onceİyelik	Previous word possessor	FL	F4
@attribute onceİliski	Previous-target word relation	OBJECT	
@attribute hedefKok	Target word root	GEL	
@attribute hedefTur	Target word POS	VERB	F5
@attribute hedefTuretme	Target word inflected POS	VERB	
@attribute hedefHalEki	Target word case marker	NULL	F6
@attribute hedefİyelik	Target word possessor	FL	F7
@attribute hedefİliski	Target-subsequent word relation	SENTENCE	
@attribute sonraKok	Subsequent word root	PUNC	F8
@attribute sonraTur	Subsequent word POS	PUNC	
@attribute sonraTuretme	Subsequent word inflected POS	PUNC	
@attribute sonraHalEki	Subsequent word case marker	NULL	
@attribute sonraİyelik	Subsequent word possessor	FL	
@attribute sonraİliski	Subsequent- Subsequent word relation	NULL	
@attribute anlam	Sense number	1	

Table 2: The words and their senses in the experiments

Word	Sense 1	Sense 2	Sense 3	Sense 4	Sense 5
yan	Burn (verb) 0.07	Be on, Shine (verb) 0.08	Near, next, side (noun) 0.47	Side by side, with somebody (adverb) 0.24	Aspect (noun) 0.14
kız	Girl (noun) 0.87	Get Angry (verb) 0.13			
kap	Grab, catch (verb) 0.79	Container, pot(noun) 0.21			
art	Back, rear (noun) 0.14	Following, successive (adverb) 0.18	Increase (verb) 0.69		
yüz	Face (noun) 0.28	Hundred (adjective) 0.43	For This reason, therefore (adverb) 0.26	Swim (verb) 0.03	
iç	Drink(verb) 0.26	Inside, interior, internal(noun, adverb, adjective) 0.74			

Table 3: Features used in the test sets

TEST SET	F1	F3	F4	F5	F6	F7	F8
AMBIGUOUSWORD1	+	-	-	-	-	-	-
AMBIGUOUSWORD2	-	+	-	-	-	-	-
AMBIGUOUSWORD3	-	-	+	-	-	-	-
AMBIGUOUSWORD4	+	+	-	-	-	-	-
AMBIGUOUSWORD5	+	-	+	-	-	-	-
AMBIGUOUSWORD6	+	+	+	-	-	-	-
AMBIGUOUSWORD7	-	-	-	+	-	-	-
AMBIGUOUSWORD8	-	-	-	-	+	-	-
AMBIGUOUSWORD9	-	-	-	-	-	+	-
AMBIGUOUSWORD10	-	-	-	+	+	-	-
AMBIGUOUSWORD11	-	-	-	+	-	+	-
AMBIGUOUSWORD12	-	-	-	+	+	+	-
AMBIGUOUSWORD13	-	-	-	-	-	-	+
AMBIGUOUSWORD14	+	-	-	+	-	-	-
AMBIGUOUSWORD15	+	+	-	+	-	-	-
AMBIGUOUSWORD16	+	+	+	+	-	-	-
AMBIGUOUSWORD17	+	+	+	+	+	+	+

Table 4: Test Results (Correctness given as %)

TEST SET	ART	İÇ	KAP	KIZ	YAN	YÜZ
1	37.66	58.48	75	88.8	37.5	60.34
2	72.73	76.79	80.55	91.2	53.12	54.31
3	70.13	74.55	80.55	88	46.88	47.41
4	72.73	79.02	75	88.8	53.12	63.79
5	72.73	76.34	80.55	85.6	49.22	60.34
6	67.53	79.02	80.55	88	51.56	64.65
7	87.01	100	100	100	55.47	73.28
8	76.62	83.04	77.78	95.2	52.34	87.93
9	70.13	82.14	80.55	88	46.09	62.93
10	87.01	100	100	99.2	57.81	96.55
11	87.01	100	100	100	56.25	75.86
12	87.01	99.55	100	99.2	62.25	96.55
13	76.62	61.16	86.11	90.4	52.34	75
14	89.61	100	100	100	57.03	83.62
15	89.61	99.55	100	100	57.03	87.93
16	87.01	100	100	100	53.12	87.93
17	90.90	92.85	91.67	95.2	71.18	87.07

4. Conclusion and Future Work

It is obvious that the syntactical clues and information can be helpful to some extent. Other than the syntax, we need at least a higher level of information for more accurate results. Sometimes, using syntax results in nothing and we have to consider some other information from one level above. We have sense distinctions that can be resolved only by using the general knowledge of the hearer, and nothing else. If we examine the following three examples it will be understood that the sense of the verb *gelmek* can not be disambiguated without the previous knowledge:

Taş ***yukarıdan geldi***. *(The stone fell down from upper part)*
Emir ***yukarıdan geldi***. *(The order is received from someone in an upper position)*
Adam ***yukarıdan geldi***.*(The man came from upstairs)*

In all these three examples we have similar structure, i.e. a noun followed by "yukarıdan geldi", but the meanings of the sentences are totally different. In these examples the word *yukarıdan* has the same root and morphemes and it just comes before the word *geldi.* The words before it have different roots but they are all nouns and they have the same POS in those sentences. However, the word *gel* has totally different meanings in all these three sentences. In the first sentence, it has the meaning to *fall down* that has been determined by the word *taş* since we have already known that the stone can fall down. The second sentence has the meaning *command* due to the word *emir*. The last one is more interesting since the sentence can be interpreted in two ways and we can disambiguate the whole sentence only by using a wider context. The first interpretation can be "The man came from upstairs" and the second can be "The man came from an upper position by favoritism".

These are some typical examples where we need some other types of information other than the syntax. At this point, the interdependencies among the words and the ontological features of the words can be necessary. The words can be classified into some general categories and some features can be attributed to them. Then these categories and their general properties can be used as a feature in the WSD process. Some typical categories are:

Önce (before), sonra (after), kış (winter), yaz (summer) ➔ zaman (time)
Mavi (blue), yeşil (green), kırmızı (red) ➔ renk (colour)
El (hand), ayak (foot), kol (arm) ➔ organ (body part)

The meaning of a word can be determined not only by the sentence it is used, but also by some of the previous sentences as the following examples:

Korktuğun başına gelebilirdi, ama **gelmedi.** **(**the thing that frightened you could happen, but it did not)
Bu at **hoşuna gitti** mi? **Gitmez** olur mu hem de çok **gitti.** (Did you like this horse? Why not, yes I did)

These examples emphasize the issues related to the human processing system in which some missing information can be inserted from the previous contexts. Then the disambiguation process takes place after this preprocessing. However, it can be difficult to determine the effect of the previous context. Since the first sentence has been affected from the sentence that comes just before this sentence, on the other hand the second one has to search further to detect this effect. The depth of the search can vary from context to context.

There are some other situations where some words that are related to the meaning of the target word are deleted in spoken language and this is very common especially in the newspapers. For example:

Galatasaray Fenerbahçe'ye 5 attı. (Galatasaray has recorded 5 goals against Fenerbahçe)
Bu maddede 46 ret, 414 kabul çıktı. (46 people voted against, and 414 for about this item)

The words *gol(goal)* and *oy(vote)* have been omitted in the sentences above. Whenever we want to disambiguate the words *at* and *çık* we have to consider these missing words. Otherwise, they can be assigned totally different meanings from the given context. Unfortunately we do not have a mechanism that can insert the missing information in these types of sentences.

WSD is an important problem in NLP systems. It has been investigated for many years, however there are still too many problems that have to be solved. Sense determination, feature selection, learning methodologies, evalutaion strategies are the major issues that have to be dealt with in this domain. We need a to find a model for the human language processing system. In order to do this we must have a coherent and plausible represantation model for the entities in the world analogous to the humans. Additionally we have to detect the components of this representation and the affective features seem to be one of these important components. If this can be completed and the problems stated above for computational stage can be overcome then we can have a very powerful disambiguation system..

REFERENCES

1. Aha, D.W., and Bankert, R.L., 1994. Feature selection for case-based classification of cloud types: An empirical comparison. In Proceedings of the AAAI'94 Workshop on Case Based Reasoning, pages 106-112, Seattle, WA.
2. Atalay, N. B., Oflazer, K., Say, B., 2003The Annotation Process in the Turkish treebank, in Proceedings of the EACL Workshop on Linguistically Interpreted Corpora-LINC, April 13-14, Budapest
3. Barwise, J., and Perry, J., 1983, Situations and Attitudes, Cambridge, MA: MIT Press.
4. Bilgin, O., Çetınoğlu, Ö., Oflazer, K., 2004, Building a WorldNet for Turkish, Romanian Journal of Information Science and Technology, Volume 7, Numbers 1–2, 163–172
5. Bloomfield, L., 1933, Language, New York: Henry Holt.
6. Bruce, R. and J. Wiebe. 1994. Word-sense disambiguation using decomposable models. In Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, pages 139-- 146.
7. Bruce, R., and Wiebe, J., 1999. Decomposable modeling in natural language processing. Computational Linguistics, 25(2):195-207.
8. Canfield J.V. (Editor), 1997, Philosophy of Meaning, Knowledge and Value in the 20th Century: Routledge History of Philosophy Volume 10. British Library Cataloguing in Publication data.
9. Cardie, C., 1993. A case-based approach to knowledge acquisition for domain-specific sentence analysis. In Proceedings of the Eleventh National Conference on Artificial Intelligence, pages 798-803, Washington, D.C.
10. Cardie, C., 1996, Automating feature set selection for case-based learning of linguistic knowledge. In Proceedings of the Conference on Empirical Methods in Natural Language Processing EMNLP, pages 113-126, Somerset, New Jersey.
11. Ciaramita M., and Johnson M., 2004, Multi-Component Word Sense Disambiguation Proceedings of Senseval-3: The Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, pp. 97-100.
12. Daeleman, W., 2002, Machine Learning of Language: A Model and a Problem, ESSLLI'2002 Workshop on Machine Learning Approaches in Computational Linguistics, August 5 - 9, 2002, Trento, Italy
13. Dagan, I., Itai, A., 1994, Word sense disambiguation using a second language monolingual corpus, Computational Linguistics, v.20 n.4, p.563-596.
14. Domingos, P., 1997, Context-sensitive feature selection for lazy learners. Artificial Intelligence Review, (11):227-253.
15. Escudero, G., Marquez, L., and Rigau, G., 2000, Naive Bayes and Exemplar-Based Approaches to Word Sense Disambiguation Revisited. In Proceedings of the 14th European Conference on Artificial Intelligence, ECAL
16. Fellbaum, C., 1998, WordNet: An Electronic Lexical Database. The MIT press
17. Fellbaum, C., Palmer, M., Dang, H.T., Delfs, L., and Wolf., S., 2001. Manual and automatic semantic annotation with WordNet. In WordNet and Other lexical resources: NAACL 2001 workshop, pages 3-10, Pittsburgh.
18. Gale, W., Church, K., 1992 and Yarowsky, D., Work on Statistical Methods for Word Sense Disambiguation." In Proceedings, AAAI Fall Symposium on Probabilistic Approaches to Natural Language, Cambridge, MA, pp. 54-60.
19. Hirst, G., 1987 Semantic interpretation and the Disambiguation of Ambiguity Cambridge University Press, England,

20. Ide, N., and Veronis, J., 1998, Introduction to the special issue on word sense disambiguation: The State of the Art, Computational Linguistics, 24(1), 1-40
21. Kelly, E. and Stone, P. (1975) Computer Recognition of English Word Senses, North Holland, Amsterdam.
22. Kilgarriff., A., SENSEVAL: An Exercise in Evaluating Word Sense Disambiguation Programs, In Proc. LREC, Granada, May 1998. Pp 581--588.
23. Leacock, C., Towell, G. and Voorhees, E. M., 1993, "Corpus-based statistical sense resolution." In Proceedings of the ARPA Human Languages Technology Workshop
24. Lee, Y. K., and Ng, H. T., 2002, An Empirical Evaluation of Knowledge Sources and Learning Algorithms for Word Sense Disambiguation. Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP-2002). pp. 41-48.
25. Mihalcea, R., August 2002, Instance Based Learning with Automatic Feature Selection Applied to Word Sense Disambiguation, in Proceedings of the 19th International Conference on Computational Linguistics (COLING 2002), Taiwan.
26. Miller, G., Ambiguous Words, Released: March 22, 2001, iMP Magazine: http://www.cisp.org/imp/march_2001/miller/03_01miller.htm
27. Mooney, R., 1996. Comparative experiments on disambiguating word senses: An illustration of the role of bias in machine learning. In Proceedings of the 1996 Conference on Empirical Methods in Natural Language Processing (EMNLP-1996), pages 82-91, Philadelphia.
28. Moore, A.W. and Lee, M.S., 1994. Efficient algorithms for minimizing cross validation error. In International Conference on Machine Learning, pages 190-198, New Brunswick.
29. Ng, H.T., and Lee, H.B., 1996. Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach. In Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL-96), Santa Cruz.
30. Ng, H. T., Zelle, J., Winter, 1997, Corpus-based approaches to semantic interpretation in natural language processing - Natural Language Processing, AI Magazine
31. Ng, H. T., 1997, Exemplar-Based Word Sense Disambiguation: Some Recent Improvements, In Procs. of the 2nd Conference on Empirical Methods in Natural Language Processing, EMNLP.
32. Orhan, Z., Altan, Z., 2003, Disambiguation of Turkish Word Senses By Supervised Statistical Methods, International XII. Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN 2003).
33. Pedersen, Ted and Bruce, R., 1997. A new supervised learning algorithm for word sense disambiguation. In Proceedings of the 14th National Conference on Artificial Intelligence (AAAI-97), Providence, RI.
34. Pedersen, T., 2001. A decision tree of bigrams is an accurate predictor of word sense. In Proceedings of the North American Chapter of the Association for Computational Linguistics, NAACL 2001, pages 79-86, Pittsburg.
35. Saussure, Ferdinand de. 1974 [1916]. Course in General Linguistics. Tr. Wade Baskin. Glasgow: Fontana & Collins. [Orig.: Cours de linguistique générale.Lousanne et Paris: Payot.]
36. Schutze, H., and Pedersen, J. 1995. Information Retrieval Based on Word Senses. In Proceedings of the Fourth Annual Symposium on Document Analysis and Information Retrieval, 161-175. Las Vegas, Nev.: University of Nevada at LasVegas.
37. Stamou, S., Oflazer, K., Pala, K., Christodoulakis, D., Cristea, D., Tufis, D., Koeva, S., Totkov, G., Dutoit, D., Grigoriadou, M., January 2002, BalkaNet: A multilingual Semantic Network for Balkan Languages, in Proceedings of the First International WordNet Conference, Mysore India
38. Tuğlacı, Pars, 1995, Okyanus Ansiklopedik Türkçe Sözlük, ABC Kitabevi Yayın ve dağıtım AŞ, İstanbul
39. Veenstra, A. van den Bosch, Buchholz, J., S., Daelemans, W., and Zavrel, J., 2000 Memory-based word sense disambiguation Computers and the Humanities, 34:171-177.
40. Weaver, W., 1949, Translation, Mimeographed,12 pp., July 15, 1949. Reprinted in Locke, William N. and Booth, A. Donald, 1955 (Eds.), Machine translation of languages, John Wiley & Sons, New York, 15-23.

41. Wittgenstein,L.,1953, Philosophical Investigations translated by G E M Anscombe Oxford: Basil Blackwell.
42. Yarowsky, D. 1993. One Sense per Collocation. In Proceedings of the ARPA Human-Language Technology Workshop, 266-271, Washington, D.C.: Advanced Research Projects Agency.
43. Yarowsky, D., 1992, Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora, In Proceedings, COLING-92, Nantes, pp. 454-460.
44. Yarowsky, D. 1994. Decision Lists for Lexical Ambiguity Resolution: Application to Accent Restoration in Spanish and French. In Proceedings of the Thirty-Second Annual Meeting of the Association for Computational Linguistics, 88-95, Somerset, N.J.: Association for Computational Linguistics.
45. Yarowsky, D., 2000, Hierarchical Decision Lists for Word Sense Disambiguation. Computers and the Humanities, 34(2):179-186.
46. Yılmaz, O., September, 1994, Design and implementation of a verb lexicon and sense disambiguator for Turkish, MS. Thesis, Bilkent University, Ankara, Turkey