6.13 Lexically sensitive disambiguation techniques

Gaustad T., van Noord G.J.M.

Most people are quite unaware of how vague and ambiguous human languages really are, and they are disappointed when computers are hardly able to understand language and linguistic communication. To illustrate the problem of ambiguity, consider the noun "party". It can refer to (at least) 4 different things:

- an organization to gain political power (political party)
- a band of people associated temporarily in some activity (search party/party of three)
- a group of people gathered together for pleasure (birthday party)
- a person involved in legal proceedings (third party rights)

Without any further information, a list of possible senses is all we can do to try and decide what "party" refers to. One could also argue that all these meanings are related and could be subsumed in a more general sense, namely "group of people". However, for various applications, such as Information Retrieval queries, it is nevertheless important to be able to distinguish between the different senses of the word "party".

Now let us consider the meaning of "party" in the following sentence: *The guests had a lot of fun at John's party last night.* It is quite clear to the human reader that the only possible reading here is the social gathering for pleasure. It is interesting to note that most people are not even aware of the potential ambiguity contained in this sentence. People are so skilled at resolving potential ambiguities that they do not realize that they are doing it. Since we do not exactly know how people resolve ambiguities, it is even more difficult to teach a computer to do the same thing. Especially if more than one ambiguous word is present in a sentence, the number of potential interpretations of the sentence "explodes": the number of interpretations is the product of all possible meanings of the words. Assume that only "guest" and "party" are ambiguous in the example sentence, and that they both have 4 senses. This brings the number of possible interpretations to 16. Imagine what happens if there are more senses to take into account...

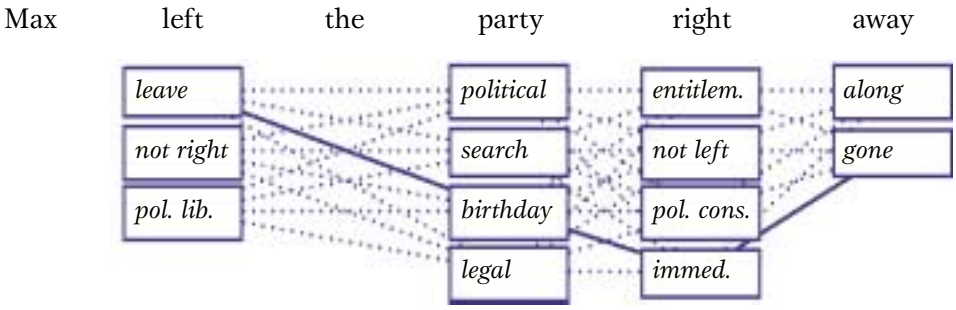

Figure 19 › Figure illustrating the possible interpretations of the sentence *Max left the party right away*. The dotted lines show all possible combinations of senses for all ambiguous words; the black line indicates the correct path.

The only way to determine the meaning of a word in a particlular usage is to examine its context. The context can be seen as the words surrounding the ambiguous word, in this case "party". A word such as "guest" might be a good cue for a particular sense of "party". But words surrounding the ambiguous word is not the only kind of information that is available. Underneath the simple words lies information on whether a word in the context is a noun or a verb (its syntactic class), on whether that same word plays the role of subject or object, on the syntactic structure of the entire sentence. All this information is certainly available to people in the process of disambiguation and a combination of all these different kinds of information together with general knowledge about the situation and the world is used to rule out improbable readings.

The goal of our project is to develop a computer tool which is able to automatically determine the meaning of a particular ambiguous word in context. In order to achieve this, we make use of the

information contained in the context, just as humans do. So we use the words surrounding the ambiguous word, and additional underlying information such as syntactic class and structure to build a statistical language model. This model is then used to determine the meaning of examples of that particular ambiguous word in new contexts.

Investigating the use of the different kinds of information available from the context might help us understand which of this knowledge is prominently used by humans in determining the meaning of words