Ambiguity comprehension in the English language

By: Johanna Fröjmark

D-level essay

Spring 2005

Lund University

Index

1. A description of the study

This study focuses on the psycholinguistic problems that people may encounter when it comes to ambiguity.

The most common words in English are ambiguous (Murray, 1990). An example is *pen*, which can both mean 'tool to write with' or 'small cage for animals'. To be able to distinguish the meaning of an ambiguous word, people naturally look at the context in which the word appears. So a sentence like: *The farmer filled the pen*, which is ambiguous, will then be easier to understand if the sentence: *The livestock was hungry* is read earlier in the text.

But how do we solve these problems if we do not have a context? What is the most natural thing to do when we cannot resolve an ambiguous sentence? And is it easier for some people to understand the correct meaning of an ambiguous sentence than others?

These are some of the questions I am going to deal with in this study.

I will also look at so called garden path sentences (Deane, 1992), sentences that lead us to interpret a sentence in one way when starting to read it, and another when it is finished. For example: *Without her donations failed to appear*. Here the 'her' might trigger our minds into seeing a possessive form rather than a personal pronoun, due to the following word 'donations'. But a disambiguated version with just a change of pronoun could be, for example: *Without him donations failed to appear*.

When studying these garden path sentences using eye-tracking, I will be able to show where and what it is in a sentence that makes us understand the words differently.

This study focuses on ambiguity in psycholinguistics, and an eye-tracking test has been performed on two groups of people. The aim was to be able to locate the exact millisecond the human mind finds out the correct meaning of the ambiguous word. The test and how it was done will be explained later.

My hypothesis is that people who have a better knowledge of English will be a lot faster in revealing the correct reading of an ambiguous sentence than people who are less knowledgeable, since English has more ambiguous words than Swedish.

This study is the complete version and follow-up of my previous A60 study about testing this experiment.

2. Sentence interpretation

How do we interpret and analyze sentences containing ambiguity and garden paths, and how can we be misled when it comes to ambiguity? When using garden path sentences, why do we normally read the sentence wrong the first time?

2.1 The Immediacy principle

People naturally start interpreting a sentence immediately when they begin reading it (Townsend & Bever 2001). Normally we interpret sentences accurately, and as we have finished the sentence we have understood it correctly. In a common sentence like: *Sam loaded the boxes on the cart*, the typical reader interprets the words one by one as soon as he reads or hears them, making 'Sam' the subject, 'loaded' the verb 'the boxes' the object and 'on the cart' is then a prepositional phrase, completing the sentence. This is a common way to interpret a sentence for us, and will here be described as the *immediacy principle* (Townsend & Bever 2001). According to this view, as has been said, the words in a sentence are interpreted as soon as they are encountered. That means that every word in a sentence (such as the one above) will hypothetically carry forward the interpreted meaning until the end of the sentence. So when an ambiguous sentence is encountered, such as, for example: *John kicked the ball to the left* (where John could either have kicked the left one of two balls, or kicked the ball to the left instead of to the right), the interpreter will choose one of the meanings and maintain it until proven wrong.

2.2 Garden paths

One particularly clear phenomenon that exemplifies the immediacy of parsing is called garden paths (Deane, 1992). "To lead someone down the garden path" has been the common description of sentences that deceive the reader and by first sight making a sentence incomprehensible.

For example, in a sentence like: *The old train the young*, the reader normally interprets the word 'train' as a noun rather than a verb, which makes the correct interpretation impossible. If readers did not commit themselves to an immediate interpretation, they would not be tricked by the garden paths.

Frazier's Garden Path Theory (Murray, 1990) makes two basic assumptions about the process of combining words into phrases. First of all, the processing resources are limited, which means that the human mind tries to minimize complexity in order to preserve the resources, that is, choosing the simplest way of interpreting the sentence (immediacy principle). Second, he uses only syntactic category information, that is, whether the words are nouns, verbs or adjectives. Therefore the Garden Path Theory suggests that, as each word is received, the listener selects the structure that provides a minimal structure change when integrating the word into the previous structure. Due to the limitation of processing resources, only the first interpretation is maintained, and is called "minimal attachment".

For example, if we encounter a sentence like: *While Mary was knitting the sock fell of her lap*, the first part 'while Mary was knitting' is interpreted as adverbial, subject and predicate, and then, to receive a minimal structure change, 'the sock' is interpreted as the object of the subclause instead of the subject of the main clause, since it is easier to expect an object after a predicate than a new subject. When the word 'fell' is encountered, this first interpretation proves to be wrong, and the reader must perform a new analysis.

There are different kinds of garden path sentences (Garman, 1990). I will describe three of the most common ones.

2.2.1 *Sentential complement*

First we have a garden path in a sentence with sentential complement: *John knew the answer was wrong.*

This sentence is a garden path, the reader believes (at least when the words are received one by one) that the sentence is complete when he reaches the word 'answer'. The Garden Path Theory explains this by noting that 'knew' can have a noun phrase object (such as for example 'the answer'), or it might have a sentential complement (for example 'the answer was wrong'). Since the sentential object construction requires rules for expanding the verb phrase in the complement clause, and the noun phrase object construction does not, the reader uses the simplest construction which then is to make 'the answer' a noun phrase object.

2.2.2 *Reduced relative clause*

A more complex garden path sentence can be found in the reduced relative clause: *The horse raced past the barn fell.*

This is a more difficult sentence, since the reader, using only syntactic category information, adopts the active structure for 'the horse raced past the barn' as soon as he reaches 'raced'. When 'fell' is then reached, it then requires re-analyzing, and it becomes apparent that 'raced past the barn' is really a relative clause that has been reduced from the full relative clause; *the horse that was raced past the barn fell.*

2.2.3 *Impossible garden paths*

There are also ambiguous sentences that are hard to interpret, since there is no real solution until further information has been given. An example is: *John told the girl that Bill liked the story*. Either it can be interpreted that John tells a girl that his friend Bill likes a story, or it can be read so that John tells a story to a girl who his friend Bill likes. That is, the first subject provides a complement object, while the second is a relative clause. But according to the Garden Path Theory, the reader is more likely to interpret the sentence as containing a complement object, since it is the simplest, most uncomplicated way.

3. The eye-tracking experiment

3.1 Hypothesis for the results of the eye-tracking experiment:

People who are higher educated in the English language or are native speakers of English will earlier resolve an ambiguous sentence than a native Swedish speaker who has a lower education in English. This is measured by the time it takes for the test subject to fixate on the correct picture without turning their gaze away, from the beginning of the sentence until its end.

3.2 The test subjects

The people tested were mostly students between the ages 20-30, but there were also representatives from non-students and persons older than 30.

I have chosen to test the difference in understanding ambiguity between higher educated in (or native speakers of) English, and lower or not educated in English at all. Therefore I have 10 people who are either native speakers, have lived a year or longer in an English speaking country or studied more than 40 points of English at the University, who belong to the group of "Well Educated in English". Then I also have 10 people who have only studied English up to 40 points at the University or have not studied at all, who are considered "None- or Low Educated in English".

3.3 The experiment

The experiment consists of two parts. The first part is the eye tracking experiment, and the second is a sheet of paper with questions about the subject (the question sheet can be found in appendix 1, pp.22).

As the subjects arrived, they were asked to sit down in front of a computer screen. They were told that they were free to leave the room and to abort the experiment whenever they wished.

Then they were introduced to the camera, which is mounted on a bicycle helmet. As the camera is in place, the test subjects were asked to fixate their eyes on a number of calibrating points appearing in different places on the screen, one at a time.

After the calibrating is done, the instructions for the experiment were given orally, to prevent any possible misunderstandings.

The instructions are the same for every one of the test subjects:

You will be shown groups of pictures with three pictures in each group (subjects are shown where they will appear, one in the top part of the screen and two below, in a triangle). At the same time, you will hear a sentence read in slow English from the speakers. You are supposed to look at the picture you find best fits together with the sentence. Fixate on it until you change your mind about it, to avoid unnecessary flickering. After each of the picture groups have disappeared, a fixation mark will appear in the middle of the screen – this is to make you look exactly in the middle of the screen, so you won't look at any particular picture as they appear.

All of the sentences contain one ambiguous word each, and some of them (no. 3, 9 and 20) are garden path sentences. Below is an example of what the test subjects saw and heard:

The sentence: *The old man hacked into the data base*

These pictures all symbolize the word ‘hack’, and the correct picture to look at is number 2 (the picture on top of the screen is number 1, the one to the left is number 2 and the picture in the lower right corner of the screen is number 3. This does not matter to the test subjects), the computer, since the word ‘hack’ in this sentence means ‘to break into a computer network illegally’. To make an example of how a test subject may reason when this picture group appears and the sentence starts, I will describe an imagined, possible scenario:

The test subject looks in the middle of the screen as the sentence starts, due to the fixation mark before the picture group appears, and as the pictures come the sentence starts: *The old man…* The subject looks at the picture of the man coughing, due to the word ‘man’.
As the sentence continues *…hacked…*, the test subject has to choose from the three pictures, since all of them can be referred to by ‘hacked’. In this imagined scenario, he chose the pickaxe. When the sentence continues with “...into the data base”, the subject realizes that the computer is the correct picture to look at, and moves his eyes to the second picture where he stays until the pictures disappear and a fixation cross comes up.

This experiment will show exactly where the majority of the test persons move their eyes to look at a different picture.

After the experiment on the computer is completed, the subjects are asked to answer the questions on the question sheet, and then they were provided with a small award (varying from testing day to testing day).
The whole experiment took about 15-20 minutes to complete.

3.4 Stimulus

A description of the experiment.

3.4.1 *Collecting stimulus*

The stimulus consists of 20 sentences, all containing one ambiguous word with three different readings. The words were either homophones/homonyms (words with different meanings and/or spellings, but the same pronunciation) or just words with metaphorically different readings. Fig 1, 2 and 3 show an example of three pictures symbolizing the word "base (bass)". The sentence heard simultaneously was "The second bass was not what I had in mind either, the musician said to the guitar salesman."

Fig 1

Fig 2

Fig 3

Fig 1 shows a bass guitar, and since "bass" is pronounced the same way as "base", it is a homophone.

Fig 2 shows a kissing couple, which has a metaphorical meaning since "second base" is slang for kissing.

Fig 3 is a baseball field with four bases, named Base 1, 2, 3 and 4. This picture represents the word "base" from baseball.

3.4.2 *Finding symbolizing pictures*

The next step was to find pictures to symbolize the different meanings of the sentences. Three pictures for each group, and 20 groups, which equals 60 pictures in total. Two of the sentences were later removed, due to technical problems.

The pictures were considered to be symbolizing the ambiguous word in the sentences if they could not be mistaken for any other of the three meanings in that sentence. To be totally sure that no one would misunderstand the symbolism, the first six test subjects from the testing group (described in my first study) were asked to state afterwards if they had misunderstood anything (since most of the test subjects were Swedish, I did not prioritize to find native English speakers to test this as well, but instead I asked the first native English speaking subject after he was finished).

Both the pictures and the sentences to which they belong can be found at the end of the study as appendix 2, pp23. Note that sentence number 20 has a word with a mismatching picture (the one with a picture of violets and the ambiguous word is “vile/viol”), in order to see if anyone would change their gaze already at the first part “viol-” because they expect “–et” to end the word.

3.4.3 *The making of the computer program*

The third step was to make the computer program that was used to create the eye tracking experiment. For this I used the program E-prime, where I also programmed details about the experiment, such as the placement of the pictures (a triangle on the screen with the same size of space reserved for all the pictures) the duration of the fixation mark (1,5 seconds), sentences and picture groups (the pictures stayed for approximately one second after the sentence was finished), and the computer showing the groups of pictures randomly, so that the same picture group wouldn’t be first for all the subjects. This is to preventing that the first mistakes the subjects may make on the first picture group (when they are not fully aware about what will happen), will not be made on the same group for every subject.

3.4.4 *Recording the sentences*

Since I could not find any native speakers before the recording (which was time scheduled), I recorded them myself. This will in all probability not affect the results of the experiment, since the author has passed several pronunciations tests at the University of both Gothenburg and Lund.

The sentences were recorded in a soundproof room and were read slower than normal to create some time for the mind to react on the words one by one. There was no particular intonation in the sentences.

When the recording was complete, the sound files were inserted in the E-prime program.

3.4.5 *Procedure*

An assistant put on the camera-helmet, and we ran the experiment in the humanist’s lab at Lund University. The camera sent signals to a computer in another room, and the data was logged as planned.

3.5 The eye-tracker

The eye-tracker is a small camera, mounted on a bicycle helmet, that measures your right pupil and corneal reflex in it, and in that way it spots exactly where you are looking on the screen. It sends the signals to another computer, which calculates the coordinates of that spot and logs them and at which time they appeared with 20 millisecond intervals.

The subject is considered fixating on a picture when his/her gaze stays for longer than 120 milliseconds.

3.6 Data analysis

To analyze the data, it is copied into Excel, creating columns of numbers (see appendix 3, pp30). In another log file received from the stimuli-computer, the order in which the picture groups appeared to each of the test subjects is found. In Excel a function was created, that showed which pictures the test subject has looked at during the time that each of the picture groups have appeared.

I created three columns, one for each picture in a group, and then the created Excel function shows a 1 for each time the test subject had looked at the picture in that column, or if not, a 0. Where only 0's have appeared, the test subject has looked outside the frames of the pictures, closed his/her eyes or looked outside the screen.

This is what the data looks like after such an analysis:

Picture1	Picture 2	Picture 3	Picture group	Time in ms per group
1	0	0	17	0
1	0	0		20
1	0	0		40
1	0	0		60
0	0	1		80
0	0	1		100
0	0	1		120
0	0	1		140
0	0	1		160
0	0	1		180
0	0	1		200
0	0	1		220
0	0	0		240
0	0	0		260
0	0	0		280
0	0	0		300
0	0	0		320
0	1	0		340
0	1	0		360
0	1	0		380
0	1	0		400

Fig 4

Fig 4.
Table showing an example of a log file of a test subject's data from the eye-tracking equipment.

To get the results the mean value of all the data of when the test subjects have looked at the correct picture must be calculated. That is, if picture 2 on the screen (for example the computer, in the example with the sentence: *The old man hacked into the data base*) was correct in a sentence that was 400 milliseconds long like in Fig 4, it means that the test subject would have revealed the correct meaning in the 340th millisecond since that is where the 1's are starting in the column for Picture 2. Having the number 340 plus the corresponding numbers from the other test subjects, a mean value can be calculated that shows the time it took for all the subjects in one group to reveal the ambiguity. I will hereby refer to this time as "time of realization".

4. The results

Since my first study on the results of the first six test subjects proved my hypothesis, I had a very positive feeling about the results of the real experiment on the 20 test subjects. After collecting all the data, the analysis was made. Due to technical problems, two of the sentences (no. 2 and 17) could not be used in the analysis.

4.1 Data calculation

Comparing the mean values in times of realization from the subjects from the test group of higher educated, to the mean value from the lower educated, a diagram can be created that shows which group is faster on realizing the correct meaning of a sentence in a particular picture group (some of these diagrams can be found in the study as appendix 4, pp31).
But that data is more useful if a statistic test is made first, to see whether it can be statistically proven that one group is faster than the other.

First, a mean value of all the times of realization for both of the groups must be calculated, to receive a total mean of how fast the ambiguity was revealed in all of the 18 sentences separately. The reason why the values must be calculated separately is because all the sentences were of different length. Table A below shows those 18 mean values:

Total mean
4728
4992
6625
5388
6285
5637
7008
4039
4410
8206
7382
7552
6454
8037
6208
5229
7272,45
4395

Table A

Table A: Shows the total mean value of times of realizations, that is, the standard time it takes to reveal the ambiguity for all of the subjects in all the 18 sentences.

After that, each of the test subjects must get a percentage of their time of realization compared to the total mean value, that is, how fast each subject was compared to what was standard for that sentence (shown by the mean value). Therefore each of the subjects' times of realization was divided by the total mean of each sentence. This produced 10 numbers in percent for each group of test subjects times 18 for the 18 sentences. That makes two columns with 180 numbers in each, and these ran through a statistics-test (T-test) in Excel show whether the results can be statistically proven correct. That means that you can say that there are enough differences between the groups to say that the difference is not only due to chance.

The two tables below show the 360 numbers (180 per table) representing the time of realization per sentence for each of the test subjects. They also show the total lengths of the sentences from the 18 picture groups. Where the same numbers have occurred both in the individual time of realization and in the total length of the sentence, it means that the subject has looked at the wrong picture when the sentence was finished.

Well educated group

↓ Sentences

Subjects →

	1	2	3	4	5	6	7	8	9	10	Total lenght of sentence
1	3720	4680	4500	6560	4900	3420	4900	4760	4540	4700	6700
2	5920	5160	4320	6700	4120	2860	6700	3660	5300	3760	6700
3	7040	6520	10100	5360	3820	6720	3780	4800	8980	6180	10100
4	5480	4940	4800	4740	5700	6380	5820	4700	3880	4360	7200
5	6080	6000	6300	4380	6300	7900	7900	4840	5620	6320	7900
6	5600	4440	4560	4180	6900	4620	6840	4500	7960	3140	8000
7	7540	7500	5160	8560	5440	5400	5560	5160	5420	9360	10000
8	5020	2860	3360	4000	3280	3680	4240	4820	3680	3520	7100
9	2460	6100	3520	4880	2240	5820	5700	2080	4640	4000	6100
10	5360	6520	6380	6740	10600	6740	7000	6800	10580	6540	10600
11	7520	7880	7600	7620	9120	1760	8540	7560	7380	7920	9800
12	7700	5220	9180	7360	9340	4660	9500	9760	10000	7120	10000
13	7260	3740	8280	7560	8600	8600	1320	1540	8460	3880	8600
14	8340	8300	7080	8260	10100	6740	8540	10100	8120	9360	10100
15	7680	5780	5540	8200	5820	5780	5880	7540	5540	5760	8200
16	4700	4940	6900	3400	4700	3600	6600	4980	4860	2300	8100
17	6540	3260	6140	10100	5900	3500	4420	10100	10100	6020	10100
18	4400	4040	4660	4280	4100	4660	5060	4120	4700	4920	6700

Table B: Shows the times of realization received from the educated group of test subjects, and the total length of the 18 sentences.

Table B

Non- or low educated group

↓ Sentences

Subjects →

	1	2	3	4	5	6	7	8	9	10	Total lenght of sentence
1	4880	6700	3660	3080	4860	6700	4620	4680	3720	4980	6700
2	4920	5800	3640	5940	6680	4380	6700	5740	3980	3560	6700
3	1580	9340	6440	5820	5820	10020	9980	8260	6200	5740	10100
4	5600	5580	6140	5980	6040	6820	6080	5860	3600	5260	7200
5	7700	5420	4320	7900	5080	6220	6160	5720	7640	7900	7900
6	4640	7180	2580	8000	6160	7120	5380	6180	5040	7720	8000
7	5600	8580	5600	9980	10000	8160	6140	7660	4420	8920	10000
8	4320	3560	3600	2060	3400	6920	4920	6220	3580	3740	7100
9	6100	5620	2620	2820	5840	5640	5200	4340	4540	4040	6100
10	10580	10580	6500	10420	10420	8460	10600	6240	10580	6480	10600
11	7260	9800	8260	5200	6880	7700	8740	5800	7880	7220	9800
12	10000	9340	2520	3620	10000	7640	7460	9020	5220	6380	10000
13	8340	7580	4360	3780	8600	8580	8600	7600	3940	8460	8600
14	2180	8960	8520	10100	8220	6600	7960	8260	7260	7740	10100
15	6440	5260	6360	6360	6340	5060	6160	7260	5480	5920	8200
16	0	7240	5420	8000	6700	5200	7840	6400	4880	5920	8100
17	10000	9320	6180	8820	7620	9729	5600	10100	6000	6000	10100
18	1580	5100	4100	2260	5760	4320	4720	5760	3920	5440	6700

Table C

Table C: Shows the times of realization received from the non- or low educated group of test subjects, and the total length of the 18 sentences.

The mean values were divided with the total mean value per sentence (shown earlier in table A).

The subjects' individual times of realization (Tables B & C) were then divided by the total mean value (Table A), one by one, which equals 180 numbers for each group of subjects, showing a percentage of each test subject's time of realization. The 360 numbers (180 for each group) were then run through a t-test (all the figures can be found in appendix 5, pp.31-32). Here follows a demonstration of the calculation described above, performed on the first picture group:

Test subject	Times of realization	% Educated group	Mean value	% Non-educated group	Times of realization
1	3720	0,786802	4728	1,032149	4880
2	4680	0,951777		1,41709	6700
3	4500	0,989848		0,774112	3660
4	6560	1,387479		0,651438	3080
5	4900	1,036379		1,027919	4860
6	3420	0,72335		1,41709	6700
7	4900	1,036379		0,977157	4620
8	4760	1,006768		0,989848	4680
9	4540	0,960237		0,786802	3720
10	4700	0,994078		1,053299	4980

The figures of times of realization for the educated group is shown in blue, the figures in the column for the non-educated are shown in yellow. The blue and yellow numbers are divided with the mean value (shown in green), one by one, equaling ten numbers for each group (shown in the two columns "% Educated group" and "% Non-educated group"). These numbers are the percent for each subject's time of realization. When it is lower than 1, it means that this subject was faster than what was standard, and when it is higher than 1 it means that the subject was slower. This means that in picture group 1 there were 6 subjects from the educated group that were faster than the standard time of realization, and 5 from the non-educated group, which can also be seen when comparing the times of realization with the mean value (that is, there are six of the blue numbers that are lower than the mean value, and five of the yellow numbers).

4.2 T-test results

The 180 numbers from each test group were inserted in the t-test as variable 1 and variable 2, and the outcome was the following table:

t-test: Two-Sample Assuming Equal Variances			
	Variable 1	*Variable 2*	
Mean	0,96441177	1,035588	1
Variance	0,06146976	0,076217	2
Observations	180	180	3
Pooled variance	0,06884346		4
Hypothesized Mean Difference	0		5
df	358		6
t Stat	-2,57351248		7
P(T<=t) one-tail	0,00523448		8
t Critical one-tail	1,6491208		9
P(T<=t) two-tail	0,01046895		10
t Critical two-tail	1,96661404		11

Table D: Shows the t-test results

Table D

Rows 1,2 and 3 show the mean, variance (the variance is a measure of how spread out a distribution is) and number of observations for each variable.

Row 4 presents the "pooled" variance, that is, the weighted average of the separate sample variances (i.e., for both samples together).

Row 5 shows the hypothesized mean difference (usually zero).

Row 6 shows the "degrees of freedom"; the number of independent pieces of information that go into the estimate of a parameter. In general, the degrees of freedom of an estimate is equal to the number of independent scores that go into the estimate minus the number of parameters estimated as intermediate steps in the estimation of the parameter itself. For example in this case there are 360 observations (180 numbers from each test group), and that minus the two parameters (the two test groups) gives us 358.

Row 7 presents the *t* statistic (the higher the absolute value, the less similar the means of the two samples are).

Row 8 shows the one-tailed probability that the *t* statistic calculated for the data is lower than or equal to the critical *t*-value, given in row 9 (a one-tailed test is used if the hypothesis is that the mean of sample 1 is either higher *or* lower than the mean of sample 2; a two-tailed test is used if your hypothesis is that the means of the two samples differ, no matter which one is higher and which is lower, which is the case in this experiment).

Rows 10 and 11 show the probability and critical *t*-value for two tails.

What is interesting to us is row 10, P(T<=t) two-tail (inside the dotted square), which shows the number 0,01046895. To receive a significant result, this number (t) must be lower than the significance level (also called alpha level), which is conventionally 0,05. In hypothesis testing, the significance level is the criterion used for rejecting the null hypothesis (that is, that the mean difference should be 0). The significance level is used in hypothesis testing as follows: First, the difference between the results of the experiment and the null hypothesis is determined. Then, assuming the null hypothesis is true, the probability of a difference that large or larger is computed. Finally, this probability is compared to the significance level. If the probability is less than or equal to the significance level, then the null hypothesis is rejected and the outcome is said to be statistically significant. (David Lane's Academic Home Page, see references for more information)

As can be seen, the number inside the dotted square is much lower than 0,05, which means that this is significant data.

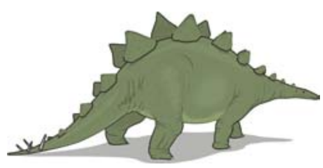
4.3 Point of disambiguation

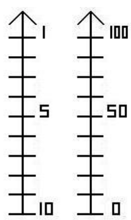
Is there a particular word or phrase in a sentence that makes it possible to resolve ambiguity? How fast can you reveal ambiguity without the result being due to chance?

The point of disambiguation is the point in every sentence where the subjects change their gaze to the correct picture without turning away, until the pictures have disappeared. That is, the word in the sentence that makes the ambiguity disambiguated.

Those words were found by listening to the sound files and stopping them at a time received from the mean values of the 10 subjects in each group, that is, if the mean value of a sentence 10000 ms long was 6540, the sentence was stopped in the 6540th millisecond. The word before that point is considered the point of disambiguation. This shows quite a clear picture of which word was the most important one to make a sentence comprehensible.

The two sentences below show the different points of disambiguation for the two groups: the words for the well educated group are written in **bold** letters, and the ones for the non- or lesser educated are written in *italics*.





1. Those scales must belong to **animals** from the *jura* period.

2. The second bass was not what I had in mind either, the **musician** said to the *guitar salesman*.

By analyzing these sentences, it becomes clear that the subjects from the well educated group have revealed the ambiguity not only earlier than the non- or lesser educated group, but at the earliest stage possible without being due to chance. There is no way that anyone is able to know which picture is the correct one before they have reached the word ‘animals’ in sentence one, and ‘musician’ in sentence two, since none of the words before reveal anything about the ambiguous word. For example, the ambiguous word in sentence one is ‘scales’, but neither ‘must’ nor ‘belong to’, which are the following words, reveal anything about the word ‘scales’. The educated group revealed the ambiguity as fast as they possibly could in 15 of the 18 sentences.

5. Discussion

This study has followed up my previous study about the testing of this experiment. I received significant data both from the six first test subjects, and the 20 tested in this study, which gives me reason to believe that I have provided future researchers in this field with a good start.

The results of my analyses show that people well educated in English are faster in revealing ambiguity in sentences with homonyms or garden paths, and they reveal it at the first point possible, but it is not quite obvious why. It is also hard to know whether the results had been different if the sentences had been read in normal speed, but I doubt it since the difference between the two groups should be the same even if you change the speed of the sentences. There is no reason to believe that there would be different results between the groups due to speed.

I believe that the reason that people with a high education are faster in revealing ambiguity might be due to English being a language full of homonyms and homophones, which makes a language richer without having to come up with too many new words when the language expands. One would think that a language with many different meanings of just one word would be hard to understand, but with this study I have proven that the understanding of ambiguity follows the knowledge of the language, and not the other way around.

So what could this knowledge be used for? For one thing, to help creating new languages, which is not only something children do, but a real hobby to some people, and a way of learning and analyzing language structure.

When creating a completely new language, it might very well be useful to know that homonymy will not be a problem to the learners.

It might also help when it comes to learning a new language. Why is it that we understand ambiguity better when we have a better knowledge of the language? Can there be changes earlier in the learning of a language, which make the comprehension of ambiguity faster to receive?

Another thing these analyses may become an asset to, is the research about dyslexia. Is ambiguity a big or small problem for dyslectics? Maybe one can find interesting results when using my method of looking at pictures to experiment on dyslectics. Since there is no reading involved, there is no room for such problems. My method would be an easy and still reliable way of seeing whether dyslectics have an easier or harder time resolving ambiguity than non-dyslectics.

References

Literature:

Deane, Paul D. 1992, *Grammar in mind and brain: explorations in cognitive syntax,* Mouton de Gruyter.

Garman, Michael. 1990, *Psycholinguistics,* Cambridge: Cambridge University Press.

Singer, Murray. 1990, *Psychology of language: an introduction to sentence and discourse processes,* Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Townsend, David J. & Bever, Thomas G. 2001, *Sentence comprehension: the integration of habits,* Cambridge, MA: MIT Press.

Internet:

Performing a t-test (April 25th 2005)
http://eeb.bio.utk.edu/weltzin/GenEcol03/Stats/ttest.htm

Using Excel to do Basic Statistical Tests (April 25th 2005)
http://depts.alverno.edu/nsmt/stats.htm

David Lane's Academic Home Page; Glossary (April 25th 2005)
http://davidmlane.com/hyperstat/glossary.html

Appendix 1

Question sheet

1. Are you male or female?

___

2. How old are you?

___

3. What is your first language?

___

4. What is your parents' first language?

___

5. (If not native speaker) How much have you been studying the English language? (grundskola, gymnasie, university 20p, 40p, 60p, 80+p, living in an English speaking country for more than 1 year, 5 years, 10 years)

___

6. How often do you (approximately) read or speak English language in your life? (more than just a few words) (every day, once a week, once a month, once a year, almost never)

___

Appendix 2

The picture groups and the sentences used in the eye-tracking experiment. The picture inside the square is the correct picture for that sentence. Under each picture is an explanation of the word


(hack: computer hacking)


(hack: coughing)


(hack: hacking with a pickaxe)

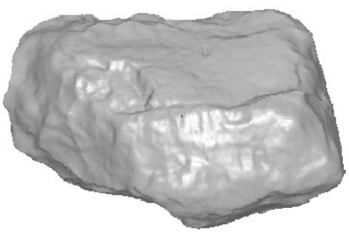
(1)
The old man hacked into the data base.


(hit: crashing into s.th)


(hit: hitting s.th with s.th)


(hit: hitting someone)

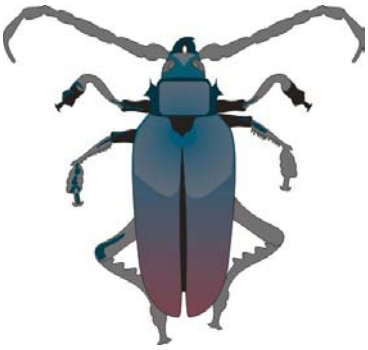
(2)
The hit was very hard, when the team's best player collided with the telephone pole.


(rock: big stone)


(rock: type of music)


(rock: back-and-forth movement)

(3)
As they heard the rock fall, they ran for cover.

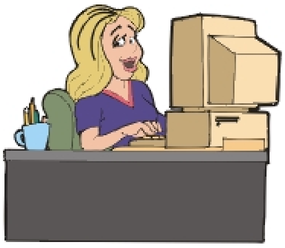


(bug: insect)

(bug: small microphone)

(bug: computer malfunction)

(4)
The intruder found a bug, so he tried to fix the computer program before continuing.



(save: to save on a data base)

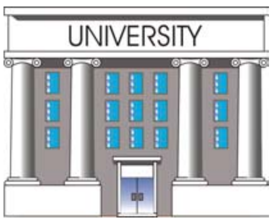
(save: rescue)

(save: store for later use)

(5)
The woman saved a little of what she had done on the computer.

(course: part of a meal)

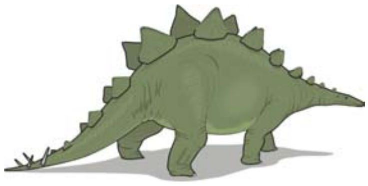
(course: route)

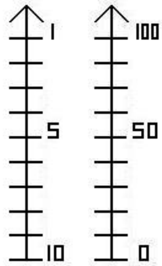


(course: lesson/meeting)

(6)
I think this course is the wrong one to take, the captain mumbled.

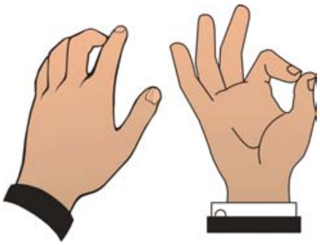

(scales: measuring instruments)


(scales: rigid plates on animals)



(scales: indicators)

(7)
Those scales must belong to animals from the jura period.


(hands: body parts)


(hands: your cards in a card game)

(hands: pointers on clock)

(8)
You have to look at the hands of that weird clock, the little boy said laughing.


(bear: animal)

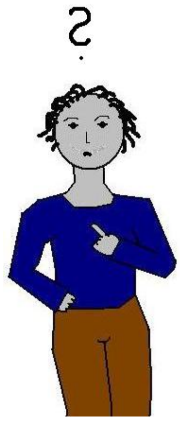

(bare: naked)


(bear: move while holding s.th up)

(9)
I want that bare little child to have a blanket!

(Aye: scotch for 'yes')

(eye: body part)

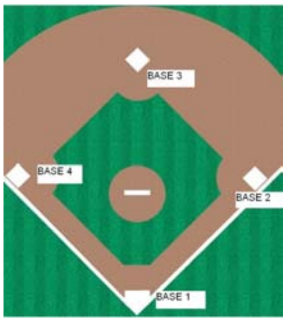


(I: myself)

(10)
I, said the person, have no idea.

(bass: low toned guitar)

(second base: slang for kissing)



(base: place in baseball)

(11)
The second bass was not what I had in mind either, the musician said to the guitar salesman.

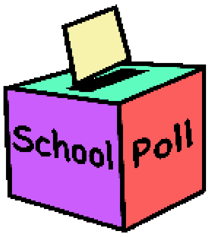
(Chile: country)

(chilli: spicy fruit)

(chilly: cold)

(12)
Isn't that chili? the mother asked her daughter who had just picked up a red little fruit.

(Pole: man from Poland)



(poll: voting)

(pole: big wooden rod)

(13)
That pole looks very big, are you sure you can get it into the ground by yourself?

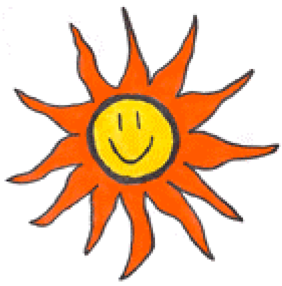
(reign: royal authority)

(rein: straps to control horse)

(rain: water from the sky)

(14)
The rain looks like it will end soon, since the drops are so small.

(raze: destroy)



(rays: beams of light)

(raise: to lift s.th)

(15)
Raze? I will not raze anything! I will be careful! the boy said.

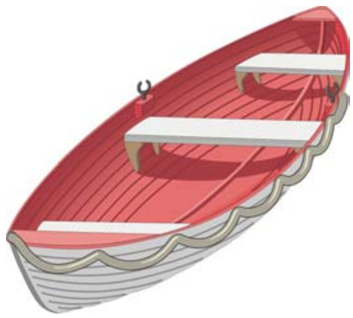
(ring: sound from telephone)


(ring: jewellery)


(wring: twisting squeeze)

(16)
I didn’t like that ring, the customer said to the telephone salesman.

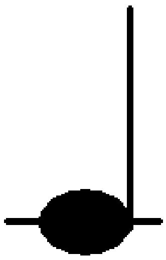

(rode: imperfect of ride)


(rowed: imperfect of row)


(road: way)

(17)
That one rowed all the way over the Thames by himself!


(see: to watch/look at)


(C: note in music)


(sea: ocean)

(18)
Is it a C that I hear, or is it maybe a D? said the musician.

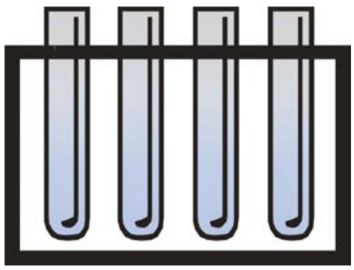
(T: 20th letter in alphabet)

(tee: starting hole in golf)

(tea: beverage)

(19)
So this is tee. Then that must be fairway, and that must be the bunker.

(vile: nasty)



(vial: small glass container)

(violet: flower)

(20)
Over there you can see a vile and dangerous beast.

Note that ‘violet’ is a mismatching picture. This was done in order to see if anyone would change their gaze already at the first part “viol-” because they expect “–et” to end the word.

Most of the pictures where found in a computer program called CorelDraw. The creators allow the use of their images as long as you add a source. Some images where also found on the internet, and one was taken from The Simpson’s homepage (http://www.thesimpsons.com).

Note: Due to technical problems, two of the sentences (no. 2 and 17) have been removed from the analyses.

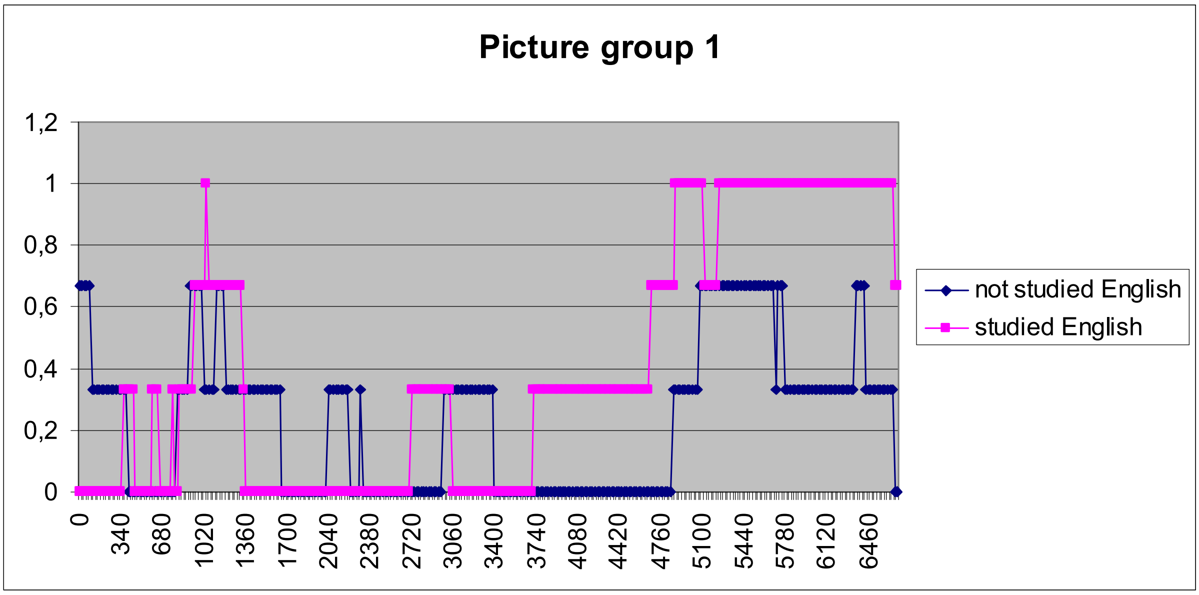
Appendix 3

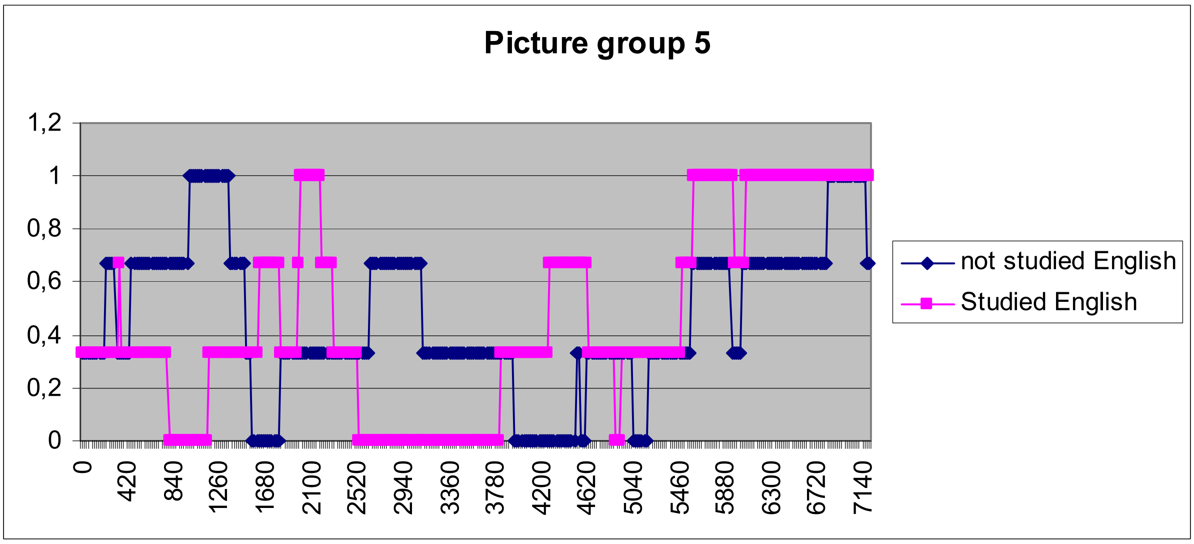
Extract of data from the eye-tracker.

#FileVersion:	2								
#Fileformat:	2558								
#Subject:									
#Date:	March 31, 2005								
#Description:									
## of Pts Recorded:	9667								
#Offset Of Calibration Area:	0	0							
#Size Of Calibration Area:	400	325							
#Sample Rate:	50								
#									
#Time	Set	Pupil H	Pupil V	C.R. H	C.R. V	ScreenH	ScreenV	Diam H	Diam V
0	-1	4667	4043	4873	4491	204	173	1496	1401
19	-1	4662	4052	4873	4499	198	174	1475	1399
40	0	4662	4064	4873	4507	199	174	1466	1400
60	0	4662	4071	4873	4512	200	177	1473	1403
80	0	4662	4071	4873	4512	199	178	1441	1399
100	0	4662	4071	4870	4512	201	177	1442	1398
120	0	4658	4068	4864	4512	204	177	1473	1398
140	0	4658	4059	4864	4507	204	175	1466	1397
160	0	4658	4053	4864	4500	202	178	1466	1367
180	0	4658	4047	4864	4495	201	180	1441	1368
200	0	4656	4044	4864	4493	204	178	1466	1368
220	0	4656	4043	4864	4493	203	176	1473	1369
240	0	4656	4039	4864	4493	205	173	1443	1398
260	0	4657	4035	4872	4489	199	175	1442	1368
280	0	4659	4031	4874	4487	199	170	1439	1399
300	0	4670	3988	4877	4461	206	157	1436	1372
320	0	4707	3929	4890	4434	225	131	1431	1366
340	0	4708	3918	4895	4421	221	132	1432	1339
360	0	4708	3915	4895	4420	223	128	1440	1365
380	0	4708	3914	4895	4420	222	130	1436	1367
400	0	4705	3911	4894	4418	222	128	1432	1335
420	0	4700	3911	4888	4417	223	126	1437	1338
440	0	4700	3911	4888	4417	223	127	1436	1334
460	0	4700	3912	4888	4417	222	127	1442	1339
480	0	4706	3912	4890	4417	222	125	1442	1367
500	0	4715	3912	4895	4417	225	126	1405	1338
520	0	4733	3879	4907	4416	234	99	1434	1339
540	0	4758	3792	4928	4358	239	73	1446	1337
560	0	4758	3792	4928	4358	241	73	1433	1364
580	0	4758	3792	4928	4358	239	76	1441	1337
600	0	4752	3790	4926	4354	239	75	1443	1365
620	0	4748	3790	4919	4354	241	74	1410	1335
640	0	4739	3790	4911	4354	242	77	1447	1367
660	0	4739	3786	4905	4354	247	74	1440	1335
680	0	4739	3769	4905	4342	248	68	1442	1340
700	0	4739	3714	4905	4308	249	48	1441	1340

Appendix 4.

Two of the diagrams showing eye movement. The higher the line goes in the diagrams, the more of the test subjects have looked at the correct picture.





Appendix 5

The 18 total mean values of the times of realization for the 20 test subjects and the individual 360 percentages of the mean values for each subject

Pic group	% Educated group	Mean	% Non-educated group	Pic group	% Educated group	Mean	% Non-educated group	Pic group	% Educated group	Mean	% Non-educated group
1	0,786802	4728	1,032149	4	1,017075	5388	1,039347	7	1,075913	7008	0,799087
	0,951777		1,41709		0,890869		1,035635		0,736301		1,224315
	1,387479		0,774112		0,879733		1,139569		1,221461		0,799087
	1,036379		0,651438		1,057906		1,109874		0,776256		1,424087
	0,72335		1,027919		1,184113		1,12101		0,770548		1,426941
	1,036379		1,41709		1,080178		1,265776		0,793379		1,164384
	1,006768		0,977157		0,872309		1,128434		0,736301		0,876142
	0,989848		0,989848		0,916852		1,087602		1,070205		1,093037
	0,960237		0,786802		0,720119		0,668151		0,773402		0,630708
	0,994078		1,053299		0,809206		0,976244		1,335616		1,272831
2	1,185897	4992	0,985577	5	0,967383	6285	1,225139	8	1,242882	4039	1,069572
	0,865385		1,161859		1,002387		0,862371		0,831889		0,881406
	1,342147		0,729167		0,696897		0,687351		0,990344		0,89131
	0,825321		1,189904		1,002387		1,256961		0,812082		0,510027
	0,572917		1,338141		1,256961		0,808274		0,911117		0,841793
	1,342147		0,877404		1,256961		0,989658		1,049765		1,713295
	0,733173		1,342147		0,770088		0,980111		1,193365		1,218123
	1,033654		1,14984		0,954654		0,910103		0,708096		1,539985
	1,061699		0,797276		0,894193		1,215593		0,911117		0,886358
	0,753205		0,713141		1,005569		1,256961		0,871503		0,925972
3	1,062642	6625	0,238491	6	0,993436	5637	0,823133	9	0,557823	4410	1,38322
	1,524528		1,409811		0,808941		1,273727		0,798186		1,274376
	0,809057		0,972075		0,741529		0,45769		1,106576		0,594104
	0,576604		0,878491		1,224055		1,419195		0,507937		0,639456
	1,01434		0,878491		0,819585		1,09278		1,319728		1,324263
	0,570566		1,512453		1,213411		1,263083		1,292517		1,278912
	0,724528		1,506415		0,798297		0,954408		0,471655		1,179138
	0,984151		1,246792		0,787653		1,096328		1,38322		0,984127
	1,355472		0,935849		1,412099		0,894093		1,052154		1,029478
	0,93283		0,866415		0,557034		1,369523		0,907029		0,9161

10	0,653181	8206	1,289301	13	1,124884	6454	1,292222	16	0,898833	5229	0
	0,77748		1,289301		1,282925		1,174465		1,319564		1,384586
	0,82135		0,792103		1,171367		0,67555		0,65022		1,036527
	1,291738		1,269803		1,332507		0,585683		0,898833		1,529929
	0,82135		1,269803		1,332507		1,332507		0,688468		1,281316
	0,853034		1,030953		0,204524		1,329408		1,262192		0,994454
	0,828662		1,291738		0,238612		1,332507		0,952381		1,499331
	0,794541		0,760419		0,579486		1,177564		0,944731		1,223943
	1,289301		1,289301		1,310815		0,610474		0,929432		0,933257
	0,796978		0,789666		0,601178		1,310815		0,439855		1,132148
11	1,018694	7382	0,983473	14	1,037701	8037	0,271245	17	0,899284	7272,45	1,375052
	1,029531		1,327554		0,880926		1,114844		0,844282		1,281549
	1,032241		1,118938		1,027747		1,060097		1,388803		0,849782
	1,235438		0,704416		1,256688		1,256688		0,811281		1,212796
	0,238418		0,931997		0,838621		1,02277		0,481268		1,04779
	1,156868		1,043078		1,062586		0,821202		0,607773		1,337789
	1,024113		1,183961		1,256688		0,990419		1,388803		0,770029
	1,067461		0,785695		1,032724		1,027747		0,448267		1,388803
	0,999729		1,067461		1,010327		0,903322		1,388803		0,825031
	1,07288		0,978055		1,164614		0,963046		0,827782		0,825031
12	1,019597	7552	1,324153	15	1,237113	6208	1,037371	18	1,001138	4395	0,359499
	1,215572		1,236758		0,892397		0,847294		1,060296		1,16041
	0,974576		0,333686		1,320876		1,024485		0,973834		0,932878
	1,236758		0,479343		0,9375		1,024485		0,932878		0,514221
	0,617055		1,324153		0,931057		1,021263		1,060296		1,31058
	1,257945		1,011653		0,947165		0,815077		1,151308		0,982935
	1,292373		0,987818		1,214562		0,992268		0,937429		1,073948
	0,691208		1,194386		0,931057		1,169459		0,919226		1,31058
	1,324153		0,691208		0,892397		0,882732		1,069397		0,891923
	0,942797		0,844809		0,927835		0,953608		1,119454		1,23777