

Weak definites with Dutch hyponyms of 'restaurant'.



Bachelor thesis (7,5 ECTS)
Cognitieve Kunstmatige Intelligentie

Author: Myrthe Tielman – 3345343

Supervisor: Joost Zwarts

Date: 08-06-2011

Abstract

This thesis is in the field of linguistics, semantics to be precise. A class of definite articles exists which behaves differently than the classical account would predict. These definites we call weak definites. This thesis will look at the definites which appear with hyponyms of restaurant. For these words an experiment was set up to determine which words can appear with a weak definite and which with a regular definite. The results from this experiment were then analysed. This analysis was done with two factors in mind which might play a role in determining whether or not a noun can appear with a weak definite. These factors are entrenchment and inclusivity. The results from this analysis were then presented in the conclusions.

Contents

1. Introduction	4
1.1. Weak definites	4
1.2. Narrowing the problem	4
1.3. Relevance to AI	5
1.4. Goals	5
1.5. Structure	6
2. Research so far	7
2.1. Traditional approach	7
2.2. Cognitive Linguistic approach	10
3. Experimental research	12
3.1. Introduction	12
3.2. The words chosen	12
3.3. The test	13
3.3.1. Subjects	13
3.3.2. Materials	13
3.3.3. Procedure	15
3.3.4. Results	15
3.4. Implications for further research	16
4. Entrenchment	19
4.1. Definition	19
4.2. Relevance	19
4.3. Corpus research	19
4.4. Analysis	20
4.4.1. 'Origin' nouns	20
4.4.2. Other nouns	21
5. Inclusivity	24
5.1. Definition	24
5.2. Relevance	24
5.3. Hierarchy of the words	24
5.4. Analysis	25
6. Conclusion	28
6.1. Conclusions	28
6.1.1. Experiment	29
6.1.2. Entrenchment	29
6.1.3. Inclusivity	29
6.2. Further research	29
6.2.1. Possible factors	29
6.2.2. Revision of the experiment	30
References	32

1. Introduction

1.1 Weak definites

The articles used in the Dutch language are *de* (the), *het* (the) and *een* (a/an). Of these articles, 'een' can be said to be indefinite while *de* and *het* are definite articles. The difference between these articles is that the definite article is traditionally said to be used when a noun is either familiar or has a unique referent (or both), while the indefinite article neutral with respect to familiarity and uniqueness and is used only to give a noun an argument status.¹ Familiarity here means the hearer has been previously acquainted with the noun in question and (following the maxims of Grice) the speaker knows this. Uniqueness means there is only one instantiation of the noun in the given context. Examples of uniqueness and familiarity are shown in 1 and 2 respectively. In 1. *Queen of England* is unique, because it has only one referent. In 2. *man* (the second token) is familiar, because it has been mentioned before.

1. *The Queen of England ate an egg.*
2. *A man and woman walked in the park, the man sneezed.*

Of course the uniqueness and familiarity analysis is very rough, and there has been much discussion about the details, but almost all theories agree that a certain amount of uniqueness and/or familiarity is involved. As has been pointed out by several authors though, the problem with this definition for the definite article is that it does not apply to all cases found in natural language (e.g. Löbner 1985; Ojeda 1993; Birner & Ward 1994; Poesio 1994; Epstein 1999; Barker 2005; Carlson and Sussman 2005; Levinson 2006). Definite articles are used even when the corresponding noun is neither unique nor familiar. We call these definite articles 'weak definites'. The contrast between 'regular' definites and weak definites is shown in 3. and 4. While in 3. *het hotel*, *het restaurant* and *het tijdschrift* seem to refer to one specific instantiation of the noun, in 4. *het ziekenhuis*, *de chinees* and *de krant* do not. The definites in 4. we call 'weak'. More background on weak definites will be given in chapter 2.

3.

- a. *Jan ging naar het hotel.*
 - i. John went to the hotel
- b. *Marie ging naar het restaurant.*
 - i. Mary went to the restaurant
- c. *Anne las het tijdschrift*
 - i. Anne read the magazine

4.

- a. *Jan ging naar het ziekenhuis*
 - i. John went to the hospital
- b. *Marie ging naar de Chinees*
 - i. Mary went to the Chinese
- c. *Anne las de krant*
 - i. Anne read the newspaper

1.2 Narrowing the problem

Because of the scope of the problem of weak definites, this thesis will focus only on a certain category of nouns. This narrowing of the problem will hopefully result in a successful search and give us an outcome which will explain all cases of weak definites within this category.

The category which I will be focussing on is that of restaurants. I have chosen to look only at Dutch nouns. One of the reasons is that it can be difficult to ascertain whether a definite is weak. This however is something which is crucial to the research. I therefore believe that testing should preferably be done with people in their native language. Because of the location of the research, Dutch is the preferred language. Aside from this

¹ Based on van der Klis (2010), Postma (2010) en Le Bruyn, de Swart & Zwarts (2010)

practical reason for looking at Dutch nouns, the Dutch language also has a property which is interesting. Many Dutch nouns translate into two separate English nouns, the Dutch language tends to 'stick' different words together. A random example is *schoolklas*, which translates to *school class* in English. Because of this property, more single nouns exist in Dutch for words which indicate a category of something else. Where in English types of restaurants will often be expressed by a noun and an adjective, in Dutch it is a single noun. These English nouns with adjective could however simply be seen as a noun with a space in-between. In any case this phenomenon makes the Dutch nouns easier to use for corpus research because they exist as a single lemma while in English the search would be for a noun combined with an adjective.

This thesis will thus focus on Dutch nouns for restaurants. I will not look to synonyms of *restaurant*, but to hyponyms. For the purpose of the research, I distinguish four categories of hyponyms of *restaurant*.

1. Name of the restaurant/ company
 - a. *La Place, van der Valk, McDonald's, Burger King, etc.*
2. Kind of food they serve
 - a. *Wokrestaurant, Pizzeria, Bistro, Rotisserie, Lunchroom, etc.*
 - i. Stir-fry restaurant, Pizzeria, Bistro, Rotisserie, Lunchroom
3. Origin of the food they serve
 - a. *Chinees, Italiaan, Griek, Turk, etc.*
 - i. Chinese restaurant, Italian restaurant, Greek restaurant, Turkish restaurant
4. Manner in which they serve
 - a. *Afhaalrestaurant, Snelbuffet, Sterrenrestaurant, Wegrestaurant, etc.*
 - i. Takeaway, counter, star restaurant, wayside restaurant

1.3 Relevance to AI

The major objective in the field of AI is to recreate (human) intelligence in another (computer) system. Language has always been an important factor in this achieving this objective, because humans seem to think and communicate mostly in language. Despite people's ease in learning language however, it has proven to be very difficult to implement language in a natural way in computers. The goal of this process is to find out how language works in humans and to formalise the results into rules we can implement in computers. The study of weak definites is a part of this. It would of course be possible to simply implement a list of all nouns that can appear with a weak definite. This is however, not the preferred solution. Optimal would be to implement the underlying rules that govern weak definites. In order to achieve this, these rules must be found first. With this thesis I hope to contribute to this search.

1.4 Goals

The goal of this thesis is to contribute to the research of weak definites. I hope to achieve this by giving a complete analysis of the way in which several factors play a role in determining if a noun can appear with a weak definite within the chosen category. The factors I will examine are entrenchment and inclusivity. One step in achieving this goal is to examine whether the categories given can provide an answer to the question if the noun can appear with a weak definite. Do all nouns in one category behave the same? If this is the case, this will be helpful in the further research. If not, it would be interesting to see if a categorization can be found which does predict the behaviour of the nouns correctly. In my conclusion, I hope to be able to say for every factor which role they play in this process.

1.5 Structure

In the coming chapters, I will examine several possible factors which might play a role in determining whether a noun can appear with a weak definite. These possible factors are entrenchment and inclusivity. Entrenchment here is a measure for the acquaintance people have with a word. The more a word is used, the more entrenched it becomes. Inclusivity refers to the relations the words have with each other. Some words might be hyponyms of others and this could effect whether it can appear with a weak definite. In order to start examining these factors it is important to know which nouns behave in which way. With some nouns it is obvious if they can appear with a weak definite, but some others seem more difficult. To get the right data I will therefore start by researching what the general opinion of native Dutch speakers is about these nouns. After this I will be able to analyse these results with the previously mentioned factors in mind.

2. Research so far

2.1 Traditional approach

The definite article started getting attention in research at the beginning of the 20th century in the work of philosophers like Russell and Frege. The classical view on definites stems from the work of Russell (1905), who proposed that the definite article signals the uniqueness and existence of the referent of the noun. This claim has some problems though, the most obvious one that a sentence like *the present king of France is bald* would be logically false if there were no or multiple kings of France, because in that case the uniqueness and existence criteria are not fulfilled. This does not however, seem to be a correct analysis of the sentence. A possible solution for this problem was given by Strawson and Frege, who took uniqueness and existence to be presuppositions necessary to evaluate the sentence instead of truth-conditions. Another problem with the strong uniqueness requirement is found in sentences like 1. Here it is obvious that the uniqueness does not hold for the entire world, but only for the given context.

1. Put that book into the bookcase.

A possible solution to this problem was introduced by Kamp(1981) and Heim(1982). Although their theories involve more, the element which has been most preserving throughout research is that of familiarity. The familiarity condition is based on the intuition of Christophersen (1939) and states that the use of the definite article is only felicitous if the noun is presumably familiar to the hearer. (Aguilar, 2008)

Most current theories about the definite article use these notions of uniqueness and familiarity in some way. Several authors pointed out however, that some uses of the definite article are neither unique nor familiar. For these uses, the term 'weak definite' is used. Within this class of definites, two groups can be distinguished. The difference is illustrated in 2. and 3.²

2.
 - a) John got these data from the student of a linguist. (*Poesio 1994*)
 - b) John was hit on the arm. (*Ojeda (1993)*)
 - c) Cal Thomas is the co-author of *Blinded by Might*. (*Levinson 2006*)
3.
 - a) Switch the light on. (*Löbner 1985*)
 - b) Sue took her nephew to the hospital. (*Carlson and Sussman 2005*)
 - c) Take the elevator to the 4th floor. (*Birner & Ward 1994*)

There is a subtle difference between the examples in 2. and those in 3. The first group we call relational weak definites and is characterized by the form [the N of [a N]]. The noun taking the weak definite is relational and takes a genitive NP with an indefinite as its argument. This class of weak definites has been studied most. The examples in 3. we call idiosyncratic weak definites and at first sight they seem to behave more like bare singular nouns. This group of idiosyncratic weak definites has received less attention than the relational weak definites in the literature. In this paper, I will focus only on idiosyncratic weak definites and from this point 'weak definite' will always refer to idiosyncratic weak definites.

Some of the main research into weak definites has been done by Carlson and Sussman(2005). They tested the intuitions about definites of 16 native speakers of English and found results which confirmed the semantic difference between weak and strong definites. A way to test whether a definite is weak is to see if a noun which is only used once in a sentence can refer to two different instantiations. Take for example *de*

² Examples collected by Ana Aguilar, Uniqueness & weak definites in Spanish, 2008

krant (the newspaper) and *het tijdschrift* (the magazine). The sentence *Anne las de krant en Marie ook* (Anne read the newspaper and so did Marie) is perfectly acceptable even when Anne and Marie are reading different newspapers. If we would substitute *de krant* with *het tijdschrift* however, the sentence would come to mean that Anne and Marie are reading the same magazine. The interpretation where two different instantiations of the noun are read would become impossible. This shows us that *tijdschrift* can not appear with a weak definite, while *krant* can. Sentences like these were used by Carlson and Sussman in their experiment. In their article they argue that weak definites should be treated as a separate class of definites. Weak definites have long been a problem case in finding an account for weak definite. To treat them as a separate class may therefore also help research into regular definites.

In their article, Carlson and Sussman argue that the class of weak definites share their properties with the class of bare singular nouns in English. A bare singular noun is a noun which appears without an article in a context where other nouns would have an article. Example of bare singular nouns are given in 4.

4.

- a) Mary went to school.
- b) John is in bed.
- c) The ship is at sea.

Carlson and Sussman provide a list of properties shared by weak definites and bare singular nouns. The first property is that the nouns in both classes are lexically restricted. The factor deciding if a noun can appear in the class is a lexical feature of the noun itself. A second feature is that they allow neither postnominal nor prenominal modification. As soon as the noun is modified by a description or adjective, it can no longer appear in the class. The third feature mentioned is that both in both classes, the nouns are semantically richer because of appearing with a weak/no definite article. A sentence like *John is in bed* means more than simply that John is in his bed. No one would use this sentence if John were standing on his bed to replace his lamp, the sentence implies something more; namely that John is sleeping. The same holds for sentences with weak definites. The fourth feature is that the items in the class have to appear with a set class of other lexical items. You can't say *John is on bed*, for instance and *Mary is behind the store* has no weak definite, while *Mary is at the store* does. The final factor mentioned by Carlson and Sussman is that although it is difficult to decide if the nouns in the classes are definite or indefinite, they do seem to take the narrowest scope with respect to other operators in the sentence.

In their article, Carlson & Sussman refrain from giving a complete analysis of weak definites, instead focussing on proving that they exist as a separate class from 'regular' weak definites. The second part of their article therefore examines this difference. The participants were shown a sentence in which a person engaged in an action with an item. (i.e. 'John read the newspaper'). The definite with this item was either weak or strong. The picture showed both a singleton item of the noun in the sentence and a group of these items. The participants were then asked to point out the item most likely to be the one engaged in the action in the picture. Their results confirm their hypothesis; they show that people do interpret weak definites different from regular definites. When shown a sentence with a regular definite, people were much more likely to choose the (unique) singleton item. With the weak definite, people more often chose an item from a group. This shows that when confronted with a weak definite people do not presuppose that the noun has a unique referent, as opposed to the regular definite, where they do. When this test was done on regular indefinites it also showed that weak definites were not treated the same as indefinite articles. These results can therefore be interpreted to be evidence that weak definites form a separate group of definites, distinguishable from both regular definites and indefinites.

Further work into the weak definite has been done by Aguilar and Zwarts (2010), who attempt to give an analysis of weak definites. They start their discussion of weak definites by making the theoretical choice to treat weak definites like regular definites in

assuming that they refer in a unique way. Under this assumption, two possible accounts for the weak definite present themselves. A short explanation of the accounts as given in the article is as follows:

1. **Minimal situation strategy:** Weak definites refer to ordinary objects that are unique in sufficiently restricted minimal situations.
2. **Abstract referent strategy:** Weak definites refer to abstract objects that have a context-independent uniqueness.³

Aguilar and Zwarts reject the Minimal situation strategy. The problem with this strategy becomes clear with sentences like *Lola took the train to work*. The problem with this sentence is that it is perfectly acceptable to assume that Lola took multiple trains. If the Minimal situation strategy were correct this would only be feasible if it were possible to pick out a minimal situation in which there is only one train. This however, seems impossible with the given sentence. The Minimal situation strategy is thus unable to deal with the fact that weak definites can refer to multiple objects in the same situation. Another problem with this account is that it is unclear how it can explain the special properties of weak definites (e.g. those described by Carlson and Sussman).

Aguilar and Zwarts thus turn to the Abstract referent strategy, which they explore further. They state that weak definites behave like singular definite generics when it comes to the reference to kinds. Singular definite generics are another class of definites which pose a problem for the traditional uniqueness approach. A generic definite in a sentence like *The lion roars when it smells food* does not refer to a unique object, but to a kind (in this case the species lion). According to Dayal (2004), these generic definites refer to uniquely identifiable and well-established kind individuals that are derived by combining the narrow meaning of the definite determiner with a common noun. Aguilar and Zwarts propose that this description holds for weak definites as well. In a sentence like *Mary reads the newspaper* this means that *the newspaper* refers to the general and unique type or kind of which all newspapers are instantiations.⁴ Following this analysis, some of the problematic properties of weak definites can be explained.

Firstly, the fact that weak definites allow no modification can be explained by the rule that the kinds need to be well-established. A modifier generally limits the kind to a certain sub-group, which is often less established. In those rare cases where a weak definite does allow modification, it is always when even the sub-group is well-established. This can explain why *the old doctor* is not weak, but *the eye doctor* is. Secondly we can explain why weak definites can appear as subjects of generic sentences, like in the sentence *The newspaper brings us our daily news*. The assumption that weak definites refer to kinds clearly explains why they can be a subject of a sentence which makes a general statement about a kind of object. The third property which can be cleared up is the stereotypicality effect. Because weak definites refer to well-established kinds, it is not strange that stereotypicality appears when they are used.

Aside from reference to kinds, weak definites also refer to instantiations of kinds. Aguilar and Zwarts follow the work of Carlson (1997) and assume that a realization relation R exists between unique objects and the kinds of which they are an instantiation. This realization can either be a unique entity or the sum of several entities. Aguilar and Zwarts also address the enrichment weak definites have due to their stereotypicality. They assume the enrichment in a sentence like *Alice went to the hospital* means that the event described in this sentence is a subset of the event 'going to the hospital' which is defined by adding the demand that a stereotypical function of hospitals was involved.

³ Weak definites and reference to kinds, Aguilar and Zwarts, 2010

⁴ Weak definites and reference to kinds, Aguilar and Zwarts, 2010 (p.6&7)

2.2 Cognitive-linguistic approach

Although most work in the field of definites has been done following a linguistic approach, another point of view also exists. This approach stems from cognitive science and thus has a very different approach to the problem.

The view follows from the observation that neither familiarity or uniqueness can give a full explanation of the uses of the definite article in English. Instead of focussing on inherent properties of the article, it looks at the cognitive state of the user. According to this approach, different cognitive domains exist. Language users constantly update the configuration of their domain according to the discourse heard. The set of possible configurations is limited by linguistic and contextual factors. Morphemes give rules by which the conceptual spaces are constructed. This is where the definite article also plays a role. Research into this role has been done by Richard Epstein, who has developed his own theory of the workings of the definite article. According to Epstein, the definite article is linked to the "accessibility" of a discourse referent. The accessibility of a noun refers to memory. The easier a noun can be retrieved in memory, the higher the accessibility is. Accessibility thus depends on previous mention of the noun, but also on how commonly a noun is used and the context. Definite articles are said to be used mostly with nouns with a low accessibility. According to Epstein, the definite article is used to signal to the hearer that an access-path exists, thus that knowledge necessary to interpret the phrase correctly is available. Because of the low accessibility associated with the definite article, we can also conclude this access-path can be complex. The function of the definite article is often to signal uniqueness and/or familiarity of the noun, which is also the traditional view on the definite article. Epstein includes other functions however, like signalling the noun will be prominent in the following discourse, to give emphasis to the noun, to signal a change in the point of view and to signal that the noun represents a value of a role.

Epstein does give examples of definites which are neither familiar or unique. Most of his examples though, seem to be neither relational nor idiosyncratic. The difference is mostly due to context. In Epstein's examples, context is vitally important to the felicitousness. With idiosyncratic weak definites however, this does not seem to be the case. The noun is essential in deciding whether a definite is weak and the verb also seems to play a role, but other than that context is unimportant.

1. Now Foreman's feet were planted. Now Moorer made the big mistake. He crouched ever so slightly. His chin was on a straight line with Foreman's feared right. It came straight and true and Moorer never had a chance. (*New York Times*, 7 November 1994, p. B13)
2. "If you plan to go to the hospital for, say, elective surgery, have a frank talk with your doctor about how long your recovery may be and what you will and will not be able to do." (*New York Times*, 19 November, 2010, P. B6)

1. is an example used by Epstein. In this sentence 'the big mistake' is the phrase which is neither familiar or unique. 2. Is an example I found myself, for consistency also taken from the New York Times, with 'the hospital' as weak definite. These sentences clearly show the importance of context. In 1., if you would delete the sentence and just keep 'made the big mistake', this would come to mean one specific instantiation of a big mistake. In 2. However, 'go to the hospital' can still mean any hospital. In 1., the context is what makes the definite something else than unique or familiar. Epstein argues that in this example, 'the big mistake' is a specific instantiation of a role, which is why the definite article is used. In 2. however, it seems the not the context but noun itself which is critical in deciding on the use definite article. In this thesis, I will only focus on the second phenomenon. Epstein however, gives no examples of this kind, and thus no explanation.

Although interesting, I believe further research in this direction will not provide the answers sought in this thesis. The main reason for this is the context dependence. In cognitive linguistics, context is vitally important in providing answers to linguistic problems. With weak definites however, the phenomenon exists even when virtually no context is given. It thus seems that the answers to this problem must be sought not in the context dependent cognitive linguistics, but in the traditional approach of logical linguistics.

3. Experimental research

3.1 Introduction

As mentioned in the introduction, this thesis will focus only on Dutch hyponyms of 'restaurant'. For this class of nouns an attempt will be made to explain why they can or cannot appear with a weak definite. The first step in this process is to decide on a finite list of words to examine. The second step is to determine for each word whether it can appear with a weak definite. After the first two steps the words can be analyzed, which will happen in the following chapters.

3.2 The words chosen

The words chosen for this paper are all hyponyms of 'restaurant'. Because the number of nouns which fall into this category is quite large, I will focus on a limited list of nouns. The results found with this set of words will hopefully be general for hyponyms of 'restaurant' so that we can explain what governs all hyponyms with regard to weak definites.

Selecting the words was done with the goal to incorporate as much diversity as possible in order to be able to say as much as possible about the differences found in the experiment. The first step was to see if any categories of nouns could be found. For this purpose the added meaning of the nouns was analysed. All hyponyms of restaurant add meaning to the term 'restaurant' in the way that they not only refer to a place where you can order/eat food, but say something more about that place. Any word for a restaurant which does not do this must be classified as a synonym or hypernym. What exactly the hyponyms specify differs per word however. A list of hyponyms for restaurants showed that four different categories can be made. The first category specifies which company owns the restaurant. A common example is 'McDonalds'. The second and third category both specify what food is served. The second category specifies the kind of food quite literally, like 'Steakhouse', while the third category specifies the origin of the food, such as 'Chinese'. The fourth and final category specifies the manner in which the food is served, as in 'Takeaway'. In selecting the nouns to use for the experiment, an equal amount of words was chosen for each category. Further selection of the nouns was done taking into account the possible factors which might play a role in deciding if a noun can appear with a weak definite. This is a very long list of course, but in this thesis I will focus on the aspects of entrenchment and inclusivity. The selection of nouns should therefore reflect nouns which differ in these aspects. In other words, both common and less common restaurants should be chosen. This aspect reflects entrenchment. Inclusivity was incorporated by making sure words were chosen which included others. An example are 'Italian' and 'Pizzeria', as all pizzerias are serve Italian food.

Taking all the above into account, the words chosen are the following:

Name of the Restaurant	Kind of food they serve	Origin of the food they serve	Manner in which they serve
La Place	Pizzeria	Chinees	Afhaalrestaurant
Vd Valk	Pannenkoekenrestaurant	Griek	Snelbuffet
McDonalds	Wokrestaurant	Italiaan	Sterrenrestaurant
Burger King	Grillroom	Turk	Wegrestaurant
AC Restaurants	Steakhouse	Indiër	Drive-inrestaurant
Applebee's	Crêperie	Thai	Eethuis

Carestel	Rotisserie	Nigeriaan	Zelfbedieningsrestaurant
Fifteen	Bistro	Zweed	Petit restaurant

An English version of this list:

Name of the Restaurant	Kind of food they serve	Origin of the food they serve	Manner in which they serve
La Place	Pizzeria	Chinese restaurant	Takeaway
Vd Valk	Pancake restaurant	Greek restaurant	Counter
McDonalds	Stir-fry restaurant	Italian restaurant	Starred restaurant
Burger King	Grillroom	Turkish restaurant	Wayside restaurant
AC Restaurants	Steakhouse	Indian restaurant	Drive-in restaurant
Applebee's	Pancake restaurant	Thai restaurant	Eating-house
Carestel	Rotisserie	Nigerian restaurant	Self-service restaurant
Fifteen	Bistro	Swedish restaurant	Luncheonette

All the names of restaurants are existing restaurant chains, some better known than others in the Netherlands. The 'kind of food' and 'manner of serving' category nouns are almost all taken from the list of hyponyms of 'restaurant' from the van Dale dictionary. 'Pannenkoekenrestaurant' is the only exception, this is one of the more common types of restaurants in the Netherlands though.

3.3 The test

For some of the nouns chosen it is easy to determine whether or not they can appear with a weak definite. Some others though, are more difficult. It is important to have a good idea of which nouns appear in which way. After all, it is not possible to analyse why noun phrases act a certain way if you don't know how they act in the first place. Although I have my own intuitions about these words, for this thesis I have chosen to research other peoples' intuitions in an experiment. Introspection is a method much used in linguistic theory and it has provided many results. It is also a method which has many critics however, especially from the field of psycholinguistics which promotes experimental research instead. Although introspection has the advantage of being a much more time-efficient method, experimental research has the benefit of preventing generalizations being made with false examples. For a scientist using introspection it is easy to presume a phrase has a certain meaning purely because it fits the hypothesis. The judgment of a single person can always be clouded by the believe in a certain theory, no matter how honourable the person. Because the subject in this thesis is a very subtle difference in meaning which is not always clear, this mistake is easily made. It is for these reasons that I have chosen to examine the intuitions of many people. Hopefully getting results from a large number of native-language users will give a better image of which words can appear with a weak definite and which can not in the Dutch language.

3.3.1 Subjects

The test subjects of this experiment were all native speakers of Dutch. This was an important factor for this experiment, for if we want to get the right results it is important we test people who have strong intuitions for Dutch. What we are testing is after all, a very subtle difference in meaning which might not be picked up so easily by non-native speakers. Other than this, the test subjects were not screened, but due to the circumstances several generalisations can be made. All test subjects were students at Utrecht University. Also, as all tests were conducted during language-related courses, namely 'Taal & betekenis' (Language & meaning), 'Inleiding algemene taalwetenschap'

(Introduction to general linguistics) and 'Natuurlijke taalverwerking' (Natural language processing). This has as a consequence that most test subjects study a discipline in which linguistics is important.⁵ Of course this means that our test subjects are not representative for an average native Dutch speaker. This however, should not have a negative effect on the results of this experiment.

3.3.2 Materials

The test I have designed uses a construction first introduced in the work of Carlson & Sussman. They pointed out that weak definites allow for a sloppy reading of the noun in an elliptical context.⁶ The example first given by Carlson & Sussman is the following.

- Mary heard about *the riot on the radio*, and Bob did, too.⁷

Carlson & Sussman pointed out that in this sentence, Mary and Bob had to have heard about the same riot, but could have heard this on two separate radio's. 'Riot' in this sentence does not allow for a sloppy reading, and its article is thus strong. 'Radio' however, does allow for a sloppy reading where 'the radio' refers to two different radios. The article with 'the radio' can therefore be supposed to be weak.

This construction can be used as an easy test to find out if an article is weak, even without the person doing it knowing anything about weak definites or their definition. Because of this property, and the fact that it can easily be translated into Dutch, I have decided to use this for my test. The construction I will be using will be as shown below.

- Marie ging eten bij de/het [noun] en Anne ook.
Mary had dinner at the [noun] and Anne did, too.

Though the English version of this sentence needs the article to be left out with some words (McDonalds for instance), this is not the case for Dutch, thus not making this a problem. The test subjects will be given this sentence, followed by the question; 'Will Anne & Mary be eating at the same restaurant?'. The possible answers to this question will be 'yes', 'no' and 'it is both possible'. For weak definites we expect the answer to be 'it is both possible', for strong definites we expect it to be 'yes'. The answer 'no' should never be chosen, and is only included because the list of answers feels more complete this way, it is more common to have 'yes', 'no' and 'either' as answers than to have only 'yes' and 'yes or no'.

A possible problem with this test is that people will answer 'yes' simply because the noun inserted is very rare, so Anne & Mary will have to have gone to the same place. With nouns like *Chinees* (Chinese) this will not be a problem, but with *Zweed* (Swede) for instance, it might be. The solution for this problem is to include a context with each question. The context given is the following:

- 'Uit: Dinerrubriek van de lokale krant van Anne & Marie':
'... afgelopen maand is de/het 2^e [noun] geopend in Lutjebroek, een mooie aanvulling op de/het al bestaande [noun].'

- 'From: Dinner section of the local newspaper of Anne and Mary':
'... last month the second [noun] has been opened in Lutjebroek, a nice addition to the existing [noun].'

⁵ See appendix 1. for a full list of the disciplines of the test subjects.

⁶ Zwarts & Aguilar Guevara, Weak definites and reference to kinds (2010)

⁷ Carlson & Sussman, Seemingly indefinite definites (2005), page 2.

These sentences should point out that there is more than one instantiation of the noun in the area. Both the 'new' and the 'existing' instantiation of the noun are mentioned, because when mentioning only the new restaurant, people might interpret the following sentence as referring to this restaurant. This would influence the results strongly and give much more strong readings, which is of course not acceptable.

Because it can be difficult to see whether a noun can appear with a weak definite, the test will not only include the hyponyms of restaurant mentioned. To test whether a person generally has a correct intuition in sentences of the given kind, four other words will also be included. These words are *krant* (newspaper), *tijdschrift* (magazine), *dokter* (doctor) and *hotel* (hotel). The sentences in which these nouns are given will of course be altered slightly to fit the noun. For these words, we know that *krant* and *dokter* can appear with a weak definite, while *tijdschrift* and *hotel* can not. If the test subject does not answer the questions with these words correctly, we know the rest of the results will also be less reliable. This can then be taken into account while analysing the results.

Test subjects will be given half of the nouns selected, plus the four additional nouns. This gives a total of 20 per person, which is chosen purely for the practical reason that people might not want to answer 36 questions, or might get sloppy near the end which could influence the results. The subjects will be given an introduction to the test including example sentences of a weak and a strong definite before the test, to introduce them to the subject. This example will be the following:

- Bij de zin '**Marie las de krant en Anne ook**' hoeven Anne en Marie niet per se dezelfde krant te lezen.
- Echter, bij '**Marie las de roman en Anne ook**', lezen Anne en Marie wel degelijk dezelfde roman.
- With the sentence '**Mary read the newspaper and so did Anne**', Anne and Mary don't have to read the same newspaper
- However, with '**Mary read the novel and so did Anne**', Anne and Mary do need to be reading the same novel.

It will then be pointed out that the following pages have sentences like the examples, which can behave in either way shown. The introduction also tells the subjects that no 'right' or 'wrong' answers exist, so that they should follow their own intuitions. The subjects will then be asked to fill in the other sentences.

The complete introduction, as well as both versions of the test can be found in Appendixes 2.1, 2.2 and 2.3.

3.3.3 Procedure

The procedure of this experiment was simple. The subjects were given the test on paper and informed that they would be participating in an experiment for a bachelor thesis. They were then asked to read the instructions and make the test. Because all necessary instruction was given in the introduction, this was all which was needed. The participants were not given a time limit, so everyone could fill in the test without time pressure.

3.3.4 Results

For each word the number of times each answer was given was recorded. The percentage of times the answer was given was then calculated. With these percentages the highest answer was calculated, as well as the difference between the percentage of times the highest answer was given with the follow-up answer. These scores can be seen in table 1. A 'both' in the column 'highest answer' means this word was mostly perceived as

being weak, a 'yes' means the word was mostly perceived as strong. For a complete overview of the results, see Appendix 3.

Word:	Highest answer:	Difference in % with the follow-up:	Word:	Highest answer:	Difference in % with the follow-up:
Pannenkoekenrestaurant	Yes	49,18	Turk	Both	70,49
Petit restaurant	Yes	27,87	McDonalds	Both	70,37
Sterrenrestaurant	Yes	20,37	Pizzeria	Both	67,21
Eethuis	Yes	9,26	Burger King	Both	59,26
Crêperie	Yes	9,26	Chinees	Both	57,38
Steakhouse	Yes	8,20	Applebee's	Both	55,74
Wegrestaurant	Yes	5,56	Fifteen	Both	52,46
Zelfbedieningsrestaurant	Yes	4,92	Vd Valk	Both	50,82
Wokrestaurant	Yes	3,70	Rotisserie	Both	50,82
			Italiaan	Both	37,04
			Zweed	Both	36,07
			La Place	Both	29,63
			Thai	Both	29,51
Tijdschrift	Yes	72,17	AC Restaurants	Both	22,95
Hotel	Yes	30,43	Drive-inrestaurant	Both	21,67
			Grillroom	Both	18,52
Dokter	Both	69,57	Nigeriaan	Both	16,67
Krant	Both	62,61	Carestel	Both	16,67
			Bistro	Both	15,09
			Indiër	Both	14,81
			Griek	Both	11,11
			Snelbuffet	Both	3,28
			Afhaalrestaurant	Both	1,85

Table 1 – Overview results experimental research

Several observations can be made from this table. The first is that for all control items the expected answer was given most. The second observation is that for the tested nouns only 8 out of 32 had most people say they were strong. 26 out of 32 were said to be weak more often. It is interesting to point out that out of 8 nouns with *restaurant* in their name (not counting AC Restaurants), 6 were said to be strong more often. This also means that 6 out of 8 nouns which were said to be strong have *restaurant* in their name. The final and most important observation though is that for the vast majority of the nouns, the difference between the first ranking answer and the follow-up is very small. Only about 1/3 of the nouns have a difference of 30% or more between the number of people giving the leading answer and the number of people giving the follow-up answer.

3.4 Implications & conclusions

Before a further analysis of the results can be given, some implications need to be pointed out. Originally the results were expected to show which words can appear with a weak definite and which can not. In order to have reliable answers to this question though, the number of times one answer was chosen over another needs to be significant. The first step thus needs to be to decide which answers will be treated as such. To do this, a fairly simple procedure was followed. As shown in Table 1, for each noun the difference in percentages between the most frequent answer and the second most given answer was calculated. This measure can be used to show how much the test-subjects agreed on an answer. For the purpose of this research, I have chosen to only treat answers which were given at least 30% more than the follow-up answer as significant. This percentage was chosen looking at the control items. As shown in Table 1, *hotel* had a percentage of 30,43. If we would thus choose a standard of over 30%, *hotel* would cease to be significant, which is not something we want as this is an item we were sure of before the test began. We do however want to be able to make some strong claims based on the results, which means we want to choose our percentage as high as

possible. Based on this minimal difference of 30% we get a list of 12 items which are significant, and two more which are so if the results were rounded to whole numbers. Out of a total of 32 nouns, this means that little more than 1/3 of the nouns are significant. From this several approaches can be designed. The first is that for the rest of this thesis, only those words which have a significant answer will be considered. This approach would severely limit the range of the possible conclusions though. If we look at the list of nouns this would leave us, we will see that only two nouns from the category of 'kind of food' are left and none from the category of 'serving manner'. The conclusions will thus only be able to say something about nouns in half of the categories as not enough examples can be given in the other categories. This would severely weaken the results. A second approach is thus to still look at all words, but constantly take into account the significance of the answer we have. This is the approach I will follow in order to say as much as possible about the results. While following this approach it will be possible to look not only to whether or not a noun can appear with a weak definite, but also how strong this answer is. It may very well be that for example the entrenchment of the noun has an effect on how certain people are whether it can appear with a weak definite or not. This aspect may be lost when following the first approach, which is another reason to choose the second approach.

With results which are in a way unsatisfying as these, it is always necessary to see if confounding factors haven't played a role in the test. For this test two possible confounding factors were found. The first possible confounding factor is the context given to the subjects. This context was given to ensure that the subjects knew there were at least 2 instantiations of the noun in the area and it was therefore physically possible to visit two different restaurants. In this context, both an old and a new restaurant was mentioned to keep the priming effect to a minimum. It would falsify the results if people believed *Anne* and *Marie* visited the same restaurant because only one restaurant was mentioned before. This priming effect may however have played some role anyway. Because the context was about the new restaurant (being opened and being an addition to the existing one), some subjects may still have seen this as a sign only this restaurant was visited. This would mean that the confounding factor has caused more people to say the definite was strong. If this were the case, it is even more remarkable that so much more nouns were said to go with a weak definite however. The second possible confounding factor lies in the question. In the question a definite is used with the name of the restaurant. A strong definite however always needs a clear referent. If this referent is not mentioned before, it needs to be known to the hearer that only one exists, because of the requirements of uniqueness and familiarity. With the given question however, no unique referent was clear. This could have caused some people to conclude that the definite was weak, because a weak definite does not need a unique referent. Although this also could have played a role, it has probably happened at a very subconscious level since most people are not aware of when exactly a strong definite can or cannot be used.

Before analysing the nouns some conclusions can be drawn from the results which are not noun-specific. As shown in Table 1, the results were far from conclusive. Even taking into account the possible confounding factors this is a strong observation. For some nouns the leading answer was given almost as much as the follow-up answer. This shows that for some nouns it is very unclear to people whether or not they can appear with a weak definite. Carlson & Sussman (2005) showed in their research that weak definites form a separate class and share properties with bare singular nouns. They didn't specify anything about the borders of this class of weak definites though. It may be because of this that up till now these borders were inexplicitly supposed to be hard. All examples shown in research so far showed a definite to be either weak or strong. This implicates that all weak definites are grammatically decided to be either weak or strong and the other interpretation is wrong. This implication is made stronger by the comparison with bare singulars, which are in fact a grammatical class with hard borders. A noun can either appear without a definite or not, and using it in any other way is grammatically

incorrect. The results from this test show something different though. From the un-conclusiveness of the results we can draw the conclusion that the borders of the class of weak definites are gradual to some extent. Some words can conclusively be said to be strong and others to be weak, but a third category of nouns exists which fall into the 'grey area' of weak definites. This grey area may even be person-specific and depending on personal experience, but more research would need to be done to make any strong claims about this. In the rest of this thesis I will use the term WDC (Weak-Definite conclusiveness) to refer to the area of the class of weak definite a noun is in. If a noun is in the grey area, the WCD will be low. If a noun is very clearly strong or weak, the WCD will be high.

4. Entrenchment

4.1 Definition

The term 'entrenchment' stems from the work of Ronald Langacker, a cognitive scientist. The term is used for the scale of how familiar people are with a word. The more entrenched a word is, the more familiar people are with that word. In Langacker's words:

"Every use of a structure has a positive impact on its degree of entrenchment, whereas extended periods of disuse have a negative impact. With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit; moreover, units are variably entrenched depending on the frequency of their occurrence (*driven*, for example, is more entrenched than *thriven*)."⁸

A way to test the entrenchment of a word is thus to simply look how often a word is used. Researching this in 'real life' would be extremely difficult, because it would entail keeping track of every piece of language heard. It is possible though, to look at the amount of tokens of a noun in a corpus. Genre and age of the text would have to be taken into account, but this is a relatively simple way to compare the entrenchment of nouns.

4.2 Relevance

Entrenchment is relevant to the problem of weak definites because previous research has shown that weak definites are mostly used in phrases which describe a typical or habitual activity (Carlson and Sussman, 2005). Because these activities are typical, you expect the phrases describing them to be sufficiently entrenched. After all, a phrase describing a typical or habitual activity is usually used often, since these activities take place often, thus making the phrase more entrenched. This gives rise to the theory that a noun must also be sufficiently entrenched to be able to appear with a weak definite. The most commonly used nouns in examples of weak definites are nouns with relatively high entrenchment, which supports the previous theory. Of course, this could also be simply because these are the nouns most easily thought of when creating examples, so further research is necessary.

4.3 Corpus research

I will test the entrenchment of the nouns chosen for this thesis by means of counting the tokens of these nouns in a corpus. This will hopefully give each noun a number by which we can compare them. The corpus chosen for this purpose is the 38 Million words corpus from 1996. This corpus is chosen purely because of its size, this is the largest Dutch corpus available which meets the requirements of not being too old or too specific. (A historical or legal corpus will not give the right data for this research, nor will a corpus from the forties, for instance). In this corpus, the tokens where the given noun refers to a restaurant are counted. This is especially relevant for the class of the origin of the food, since these terms can refer to people as well as to places to eat. For the sake of completeness the tokens which referred to people were counted separately, as well as the number of times '... Restaurant' appeared. A count of the tokens in the corpus gave the results as shown in Table 2 and 3.

⁸ Ronald Langacker – Foundations of Cognitive Grammar (1987), p 59

Company Name	Nr. of Tokens	Kind of food	Nr. of Tokens	Manner of Serving	Nr. of Tokens
McDonalds	34	Pizzeria	12	Wegrestaurant	44
Vd Valk	16	Bistro	6	Eethuis	28
Burger King	2	Pannenkoekenrestaurant	2	Afhaalrestaurant	1
La Place	0	Grillroom	2	Snelbuffet	1
AC Restaurants	0	Steakhouse	1	Sterrenrestaurant	1
Applebee's	0	Wokrestaurant	0	Zelfbedieningsrestaurant	1
Carestel	0	Crêperie	0	Drive-inrestaurant	0
Fifteen	0	Rotisserie	0	Petit restaurant	0

Table 2 – Entrenchment according to the corpus research, Company name, Kind of food and Serving manner category

Origin Food	Nr. of Tokens	Nr. of tokens persons	Nr. of tokens with restaurant
Chinees	0	0	47
Italiaan	1	289	7
Turk	0	0	7
Griek	3	19	2
Indiër	0	0	2
Nigeriaan	0	35	0
Thai	0	4	0
Zweed	0	0	0

Table 3 - Entrenchment according to the corpus research, Origin category

4.4 Analysis

Before the data can be analyzed, some guidelines for this analysis must first be made. Entrenchment is a quality which can best be used in comparisons, there is no set rule for when the entrenchment is low and when it is high. I will therefore only look at the differences found, not at the numbers themselves. While comparing there is a problem though with the words from the category 'Origin of the food'. These terms are most used in phrases like *I'm going to the [restaurant]*, which appear mainly in conversation. The corpus used has no conversational component though. Comparing these words with the others would therefore give the wrong image. This means that the nouns in question will only be compared with nouns in the same category and will thus be treated separately from the rest. In all comparisons nouns with 3 or less tokens in the corpus will be treated as not significant. This is because very few tokens in the corpus can be a coincidence. A large portion of the corpus exists of newspaper texts and the three tokens registered for *Griek* for instance were all referents to an event happening at one specific restaurant. This is the reason only numbers of 4 or more will be given any credit, for the purpose of this comparison everything lower will be treated as a 0.

4.4.1 'Origin' nouns

In table 4 the complete results for these nouns can be found. This table shows that for all words, most people believed it to be weak. For some words the difference with the number of people believing it to be strong was greater than for others though. This means the WDF of the nouns does differ.

Word:	Yes	No	Both	% Yes		% No		Highest answer:	Difference with the follow-up.
				%	%	%	%		
Griek	21	6	27	38,889	11,111	50,000		Both	11,11111
Indiër	22	2	30	40,741	3,704	55,556		Both	14,81481
Nigeriaan	21	3	30	38,889	5,556	55,556		Both	16,66667
Thai	21	1	39	34,426	1,639	63,934		Both	29,5082
Zweed	19	1	41	31,148	1,639	67,213		Both	36,06557
Italiaan	16	2	36	29,630	3,704	66,667		Both	37,03704
Chinees	12	2	47	19,672	3,279	77,049		Both	57,37705
Turk	8	2	51	13,115	3,279	83,607		Both	70,4918

Table 4 – Results words in the Origin category

If we look at the numbers we got from our corpus research we see that the number of times the word was used as name for a restaurant is very small. Because the corpus has a large section which came from newspapers, a restaurant can also be named several times linking it to a single event. This means that a difference of two or three is not significant and that one token found can be a coincidence. This list therefore tells us very little. It is for this purpose that the number of persons and the number of times the noun appeared as an adjective with *restaurant* were also counted. In the list for persons we see that *Italiaan* and *Nigeriaan* are higher than expected. This can be explained by the sports section of the corpus, where athletes are often referred to by their nationality. Although the word for a person can be seen as a homonym of the word for a restaurant, this might still effect the entrenchment and is therefore a factor to keep in mind. For this analysis though, I will focus mostly on the third list. The number of times the noun appears as an adjective for *restaurant* is a good indication of how much people talk about certain types of restaurants. It is therefore closely linked to how much people use only the adjective to describe the restaurant. This is also the list which has the most realistic numbers and is still somewhat significant.

Based on the corpus research and the data from the experimental research we can make several observations. The first is that there might be a relation between the entrenchment of a word and the WDC. (The WDC is based on the percentages from the column *Difference with the follow-up*). The three nouns which were found most in the corpus are all also the top 3 of nouns with the highest WDC. The five nouns which were found least in the corpus are therefore also the five nouns with the lowest WDC. There is a problem with this hypothesis though, namely the noun *Zweed*. *Zweed* is not found at all in the corpus, but still has a relatively high WDC. This means that 1/8 of the nouns does not show this relation, while 7/8 does. Because of the small number of nouns one single noun not following the hypothesis is still a large set-back. The following conclusion can therefore be drawn. A relation between a high WDC and a high entrenchment is probable, but due to contradicting evidence more research would need to be done to make a more conclusive claim.

A second observation which can be made is that based solely on these tables, very little can be said about whether or not entrenchment effects whether or not a noun is weak. To say something about the effect of entrenchment we need to be able to make a comparison, which is here made impossible because for all nouns the most given answer was that the noun was weak. This problem does not exist for the other nouns however, making this a problem which will be solved when analyzing these nouns.

4.4.1 Other nouns

In table 5 the complete results for all the nouns not yet analysed can be found. This table shows that the majority of the words was believed to be weak. All nouns in the category 'Company name' were said to be weak, while for the categories 'Serving manner' and 'Kind of food' the most given answer varies. It can also be noted that for a relative large

number of nouns, the difference between the most given answer and the follow up is not significant. This means many nouns in this list have a low WDC.

Word:	Yes	No	Both	% Yes		% No		Highest answer:	Difference with the follow-up.
				Yes	% No	Both	% Both		
Pannenkoekenrestaurant	45	1	15	73,770	1,639	24,590	Yes	49,18033	
Petit restaurant	39	0	22	63,934	0,000	36,066	Yes	27,86885	
Sterrenrestaurant	31	2	20	57,407	3,704	37,037	Yes	20,37037	
Eethuis	29	1	24	53,704	1,852	44,444	Yes	9,259259	
Crêperie	29	1	24	53,704	1,852	44,444	Yes	9,259259	
Steakhouse	32	2	27	52,459	3,279	44,262	Yes	8,196721	
Wegrestaurant	28	1	25	51,852	1,852	46,296	Yes	5,555556	
Zelfbedieningsrestaurant	32	0	29	52,459	0,000	47,541	Yes	4,918033	
Wokrestaurant	27	2	25	50,000	3,704	46,296	Yes	3,703704	
Afhaalrestaurant	25	3	26	46,296	5,556	48,148	Both	1,851852	
Snelbuffet	28	3	30	45,902	4,918	49,180	Both	3,278689	
Bistro	22	1	30	41,509	1,887	56,604	Both	15,09434	
Carestel	21	3	30	38,889	5,556	55,556	Both	16,66667	
Grillroom	21	2	31	38,889	3,704	57,407	Both	18,51852	
Drive-inrestaurant	23	1	36	38,333	1,667	60,000	Both	21,66667	
AC Restaurants	23	1	37	37,705	1,639	60,656	Both	22,95082	
La Place	18	2	34	33,333	3,704	62,963	Both	29,62963	
Rotisserie	15	0	46	24,590	0,000	75,410	Both	50,81967	
Vd Valk	14	2	45	22,951	3,279	73,770	Both	50,81967	
Fifteen	13	3	45	21,311	4,918	73,770	Both	52,45902	
Applebee's	12	3	46	19,672	4,918	75,410	Both	55,7377	
Burger King	9	4	41	16,667	7,407	75,926	Both	59,25926	
Pizzeria	8	4	49	13,115	6,557	80,328	Both	67,21311	
McDonalds	7	2	45	12,963	3,704	83,333	Both	70,37037	

Table 5 – Results words in the Company name, kind of food and serving manner category

Comparing this table with the findings from the corpus, we can make several observations. There are two comparisons possible. The first is the comparison of the entrenchment with the WCD, the second the comparison of the entrenchment with the most given answer. This analysis will start with the first comparison. For the nouns in the category 'Origin' a relation between the WCD and the entrenchment of a noun was visible. This relation can also be seen with this set of nouns. The two nouns with the highest WCD (*Pizzeria* and *McDonalds*) are both also the most entrenched word of their category and none of four nouns with the lowest WCD were found a significant amount of times in the corpus. Again though there are some nouns which defy this trend.

Wegrestaurant is one of the main problems with this hypothesis. This noun has the highest entrenchment of all words, but scores among the five lowest when it comes to the WCD. The conclusion is again that though a relation between WCD and entrenchment is probable, this set of nouns does not provide conclusive evidence to support this hypothesis fully.

The second comparison to be made is between the entrenchment of a noun and whether or not this noun is weak. Looking at the data from the corpus research and the experimental research, a relation between the two is difficult to find. The two most entrenched words have different scores. *McDonalds* is mostly said to be weak, *Wegrestaurant* mostly to be strong. Taking the other words into account does not help in finding a possible relation. We can therefore conclude the entrenchment of a word does not effect whether or not it can be used with a weak definite. This seems somewhat

surprising, as stereotypicality is an important quality of weak definites. Most examples of weak definites give nouns which have a relatively high entrenchment. The reason that the expected relation was not found here might be that the entrenchment of a noun is less important than the entrenchment of the phrase. It has been pointed out before that whether or not a definite is weak depends not only on the noun, but also on the sentence in which the noun is used. In 1. the definite is obviously weak, but in 2. it is not even though the same word has been used.

1. Hij leest de krant.
He reads the newspaper.
2. Hij scheurt de krant in twee stukken.
He tears the newspaper in two.

These examples show that the context in which the noun is used also effects whether or not a noun is weak. Roughly speaking, weak definites appear only in sentences where the activity is stereotypical for the object. This explains why *reading* and *newspaper* have a weak definite, while *tear* and *newspaper* do not. This quality of weak definites might explain why no relation between entrenchment of a noun and 'weakness' was found. An hypothesis would be that although the entrenchment of a noun plays no role, the entrenchment of a whole sentence does. This would mean that sentences which are used more often also have more weak definites in them. Important is here to keep in mind which comes first, entrenchment or a weak definite. In this case it would be the high entrenchment of a sentence which causes the definite to get a weak interpretation.

5. Inclusivity

5.1 Definition

Inclusivity refers to the situation where a certain noun 'includes' another. In this situation the word being included is the hyponym, while the 'including' word is the hypernym. Inclusivity appears in language every time one noun refers to a subclass of the class another noun refers to and is very common. An example keeping in line with this thesis is the couple *restaurant* and *pizzeria*, where *restaurant* is a noun which can be used for all pizzeria's. In set notation:

A is included in B if $a \in A \rightarrow a \in B$ for all a in A.

Translated back to natural language, A & B refer to the nouns in question, a refers to all possible instantiations. If a is included in A, this means that the word A can refer to the instantiation a.

5.2 Relevance

Inclusivity is very obvious when looking at the words examined in this thesis. It is quite easy to create a hierarchy from the words where some include others. One of the goals in analysing weak definite is to explain why *restaurant* can not appear with a weak definite, but *Chinees* can (when used as a term for the restaurant). An hypothesis for this case is that inclusivity plays role. Because the hyponym of 'restaurant' can appear with a weak definite, the hypernym (restaurant) can not. Of course, this is a very rough hypothesis, but it is one worth looking into. Inclusivity is something which doesn't only play a role in linguistics, but in cognitive science as well. Because the problem of weak definites is very much a problem of semantics, factors which play a role in the semantic field are especially interesting to look into.

5.3 Hierarchy of the nouns

In order to examine the factors of inclusivity on this problem, we first need to examine how the words behave in terms of inclusivity. I therefore created a hierarchy of the selected words. The first hierarchy I created is based not on the nouns themselves so much as on the categories. Though not all the nouns include others, it is possible to imagine that nouns in a certain category can include those in another, but not the other way around. This does not fit the official description of inclusivity, rather it is the possibility of inclusivity.

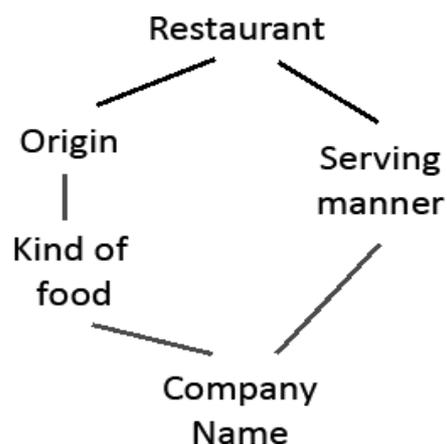


Figure 1 – Hierarchy showing the inclusivity of the categories

In figure 1 the hierarchy is shown. The grey lines represent possible relations, the black lines certain relations. The lines connecting the categories mean that for a noun in the lower category, all instantiations can also be instantiations of one single noun in the higher category. For instance, it is possible to imagine that all Carestel restaurants are drive-in restaurants. The same holds for pizzerias, all these are Italian restaurants. In this figure a certain symmetry can be found. Restaurant is a hypernym of the words in all categories, Company names are possible hyponyms of the words in all categories. The categories standing between 'Company name' and *Restaurant* can be divided in two categories. One path says something about the way the food is served, the other about the food itself. The categories 'Origin' and 'Kind of food' are both placed on the path which says something about the food itself. 'Origin' is placed above 'Kind of food' though, simply because it is possible to imagine that a kind of food is served in one country alone but not that all food from a certain county is of the same kind. It should be noted however that this hierarchy is of a different sort than the rest of the figure. The reason which places 'Origin' above 'Kind of Food' has more to do with world knowledge, while the other hierarchies are mainly the way they are for semantic reasons.

5.4 Analysis

In table 6 the results for all nouns from the experimental research are shown, grouped per category. The results from this table will be compared with figure 1 in two ways. First a comparison between the place in the hierarchy and the WCD will be made and afterwards a comparison between the place in the hierarchy and the most given answer. Because the hierarchy only says something about the categories and not about the nouns themselves it is necessary to consider the results of the experimental research per category.

Word:	Yes	No	Both	% Yes	% No	% Both	Highest answer:	Difference with the follow-up.
Petit restaurant	39	0	22	63,934	0,000	36,066	Yes	27,86885
Sterrenrestaurant	31	2	20	57,407	3,704	37,037	Yes	20,37037
Eethuis	29	1	24	53,704	1,852	44,444	Yes	9,259259
Wegrestaurant	28	1	25	51,852	1,852	46,296	Yes	5,555556
Zelfbedieningsrestaurant	32	0	29	52,459	0,000	47,541	Yes	4,918033
Afhaalrestaurant	25	3	26	46,296	5,556	48,148	Both	1,851852
Snelbuffet	28	3	30	45,902	4,918	49,180	Both	3,278689
Drive-inrestaurant	23	1	36	38,333	1,667	60,000	Both	21,66667

Pannenkoekenrestaurant	45	1	15	73,770	1,639	24,590	Yes	49,18033
Crêperie	29	1	24	53,704	1,852	44,444	Yes	9,259259
Steakhouse	32	2	27	52,459	3,279	44,262	Yes	8,196721
Wokrestaurant	27	2	25	50,000	3,704	46,296	Yes	3,703704
Bistro	22	1	30	41,509	1,887	56,604	Both	15,09434
Grillroom	21	2	31	38,889	3,704	57,407	Both	18,51852
Rotisserie	15	0	46	24,590	0,000	75,410	Both	50,81967
Pizzeria	8	4	49	13,115	6,557	80,328	Both	67,21311
Griek	21	6	27	38,889	11,111	50,000	Both	11,11111
Indiër	22	2	30	40,741	3,704	55,556	Both	14,81481
Nigeriaan	21	3	30	38,889	5,556	55,556	Both	16,66667
Thai	21	1	39	34,426	1,639	63,934	Both	29,5082
Zweed	19	1	41	31,148	1,639	67,213	Both	36,06557
Italiaan	16	2	36	29,630	3,704	66,667	Both	37,03704
Chinees	12	2	47	19,672	3,279	77,049	Both	57,37705
Turk	8	2	51	13,115	3,279	83,607	Both	70,4918
Carestel	21	3	30	38,889	5,556	55,556	Both	16,66667
AC Restaurants	23	1	37	37,705	1,639	60,656	Both	22,95082
La Place	18	2	34	33,333	3,704	62,963	Both	29,62963
Vd Valk	14	2	45	22,951	3,279	73,770	Both	50,81967
Fifteen	13	3	45	21,311	4,918	73,770	Both	52,45902
Applebee's	12	3	46	19,672	4,918	75,410	Both	55,7377
Burger King	9	4	41	16,667	7,407	75,926	Both	59,25926
McDonalds	7	2	45	12,963	3,704	83,333	Both	70,37037

Table 6 – Results from the experimental research

When looking at the WCD of the nouns per category, several observations can be made. The first is that great differences exist within categories, the standard deviations are all almost the same as the average. This would suggest that the WCD of a noun has no relation with what category a noun is in. When comparing the average WCD and the standard deviation (as shown in Table 7) of different categories though, we do see some differences.

Category	Average	Standard deviation
Serving Manner	11,85	10,37
Kind of food	27,75	24,28
Origin	34,13	35,62
Company Name	44,74	46,16

Table 7 – Average WCD and standard deviation per category

Table 7 shows that the average WCD of the category 'Company Name' is much higher than that of the category 'Serving manner'. Interesting is also that there also seems to be a relation between the number of weak nouns in a category and the WCD. The more weak nouns, the higher the WCD. The question now is if the hierarchy of figure 1 can explain these differences.

Comparing the WCD scores with the hierarchy shows that a relation between the WCD and the place in the hierarchy is not probable. The fact that company name has the highest WCD and the lowest rank in figure 1 might suggest that the higher the WCD, the lower the rank of the category. The WCD scores of the other categories do not support

this conclusion though. Especially 'Origin' and 'Kind of food' do not obey this principle, as 'Origin' has the higher WCD, but 'Kind of food' is placed lowest in figure 1. This hypothesis can thus be rejected. When comparing figure 1 with table 7, no other possible relations present themselves. We can therefore conclude that these data do not support any relation between WCD and inclusivity.

The second comparison to be made is the one between inclusivity and whether or not a noun is said to be weak. Again it is necessary to compare the categories instead of single nouns. The relevant data for this comparison is found in table 8. The average WCD is included in this table because it gives information about how significantly weak or strong a noun was.

Category	Nr. Of Weak nouns	Nr. Of Strong Nouns	Average WCD
Serving Manner	3	5	11,85
Kind of food	4	4	27,75
Origin	8	0	34,13
Company Name	8	0	44,74

Table 8 – Nr. Of weak & strong nouns per category

Comparing table 8 with figure 1 several observations can be made. We know the word *Restaurant* is strong. This word is placed highest in the hierarchy. Noticeable is now that the 'weakest' category in the list, namely 'Company name' is placed lowest in the hierarchy. This suggest that the more strong nouns exist in a category, the higher the placement in figure 1. The problem with this analysis is the category 'Origin'. This is the only category which seems out of place. Both 'Kind of Food' and 'Serving manner' obey this principle fully. It is of course possible to imagine that for 'Origin', other factors interact with inclusivity resulting in different scores in the experiment. For one, we know that the placement of 'Origin' and 'Kind of food' is solely due to world knowledge. This ordering is much less due to semantic reasons than the others. This alone might explain why 'Origin' acts different than the other categories. The conclusion we can draw from this is that a relation between the 'weakness' of a noun and the place in the hierarchy is probable. It seems like the weaker the noun/category, the lower the place in the hierarchy. Further research would need to be done into the category 'Origin' though, to be able to draw any certain conclusions.

6. Conclusion

6.1 Conclusions

This thesis has investigated a set of words and analyzed their behaviour with respect to weak definites. This process has involved two parts, an experimental part and the analysis of the results of this experiment. Because of the scope of the problem it has not been possible to analyze the data for all possible factors, only entrenchment and inclusivity have been selected in this thesis. In this chapter the conclusions from both the experiment and the analysis will be presented.

6.1.1 Experiment

The experimental part of this research was done in order to have reliable data for the further analysis. The expectation was that from this experiment a clear image would arise of which nouns could appear with a weak definite and which could not. The results were much less clear than expected though. Although a setback, this has enabled us to draw some conclusions from the results of the experiment alone.

The first conclusion is of course which words can appear with a weak definite, and which can not. Because the answers were inconclusive for many of the nouns, only nouns with a WCD of at least 30% will be considered significant enough to present as results. The results for these words is shown in table 9.

Word:	Category	Conclusion	WCD in %
Pannenkoekenrestaurant	Kind of food	Strong	49,18033
Italiaan	Origin	Weak	37,03704
Zweed	Origin	Weak	36,06557
Rotisserie	Kind of food	Weak	50,81967
Vd Valk	Company name	Weak	50,81967
Fifteen	Company name	Weak	52,45902
Applebee's	Company name	Weak	55,7377
Chinees	Origin	Weak	57,37705
Burger King	Company name	Weak	59,25926
Pizzeria	Kind of food	Weak	67,21311
McDonalds	Company name	Weak	70,37037
Turk	Origin	Weak	70,4918

Table 9 – Significant results

As seen in this table, the vast majority of the significant answers said the noun to be weak. From this table we can also conclude that for the category 'Origin' and 'Company Name' at least half of the nouns are significantly weak, since 8 nouns per category were used in the experiment. Of the category 'Kind of food' one noun is significantly strong, two are weak. No noun from the category 'Serving manner' got a significant result. When taking into account all results we see that for the category 'Origin' and 'Company Name' all nouns were said to be weak, while for the other categories roughly half was weak and half strong. It is also noticeable that the categories with only weak nouns have a much higher average WCD than the others. This suggests that a relation between WCD and the 'weakness' of the nouns might exist, the weaker the noun, the higher the WCD.

Aside from the conclusions derived from the 'weakness' of the noun, we can also draw some conclusions from looking at the WCD only. In the existing literature the class of weak definites has always been treated as a class with a fixed border. From the work of Carlson & Sussman(2005) we know that weak definites are in fact treated as a separate

class of nouns. No research has been done however to study the border of this class. Nouns have been treated as either weak or not, but this research shows this might be the wrong approach. The vast majority of the words investigated gave no significant result. People had very different opinions on whether or not a noun was weak. This result suggests that the border of the class of weak definites is not as hard as it was supposed to be. Many nouns fall into a 'grey area' where a noun can either be strong or weak, depending on the interpretation of the hearer. The main conclusion we can draw from this experiment is thus that nouns exist of which it is simply not clear whether or not they can appear with a weak definite.

6.1.2 Entrenchment

The results from the experimental research have been compared with the entrenchment of the words according to the 38-miljoenwoorden corpus. Because results from this corpus for the 'Origin' category were very minimal, these words were analyzed separate from the other categories.

The first comparison made was between the entrenchment of the words and the WCD. This comparison showed a relation between the two is probable. The results show that most of the nouns with a relative high entrenchment also have a relative high WCD. The problem though is that some nouns exist which do not show this relation. The conclusion which can be drawn is that it is probable a relation exists between a high entrenchment and a high WCD, but more research would be needed to be certain. The second comparison made was between entrenchment of the words and the 'weakness'. This comparison showed no obvious relations. This might be because it is not the entrenchment of the nouns which plays a role, but the entrenchment of the entire sentence.

6.1.3 Inclusivity

Inclusivity was the other factor with which the results from the experiment were compared. The place in the hierarchy of all categories was compared with both the average WCD and the number of weak and strong nouns. The first conclusion from these comparisons is that a relation between the WCD and the inclusivity of the noun is not probable. No obvious trends were visible in the data. The other conclusion which can be drawn relates to the number of weak nouns in the category. In the data a trend is possible where the 'weaker' the category, the lower it is in the hierarchy. The only problem with this hypothesis is the category 'Origin', which although weak, is relatively high in the hierarchy. A possible explanation for this is that the placement of 'Origin' in the hierarchy is more due to world knowledge than the placement of the other categories. The conclusion we can draw from these observations is that a relation between a high place in the hierarchy and many weak nouns is probable, but that more research into the category 'Origin' is needed to draw any definite conclusions.

6.2 Further research

This thesis has been able to provide some answers which bring us closer to understand what governs weak definites. As with most research though, it has also raised more questions. In this section I will provide an overview of these questions and provide suggestions for further research necessary to answer them.

6.2.1 Possible factors

This thesis has analyzed the results from the experimental research by examining two factors which might play a role in deciding whether or not a noun can appear with a weak

definite. It has become clear though that entrenchment of the nouns and inclusivity of the categories are not the only factors which play a role.

When comparing the entrenchment of the nouns with the results of the experimental research it became clear that no obvious relations exist. A factor which is closely related though is the entrenchment of the entire sentence. Though weak definites are largely independent of context, the specific sentence in which they appear does play a role. Only sentences which denote a stereotypical activity of the noun have weak definites. It may therefore be important how often a noun is used in the specific sentence. In order to examine this, the entrenchment of not only the noun but the entire sentence would need to be compared with the weakness of the definite.

Another factor which is closely related to entrenchment is the presence in the environment. This factor is how many instantiations of the nouns exist in the environment of the speaker/hearer. This may be of importance because when people are not familiar with instantiations of the noun, they might suppose it to be strong simply because they cannot imagine more than one instantiation existing.

One of the main problems with testing this factor is that a person's environment is always different than that of other persons. This difference in experience between people might be very important in how definites are understood. Because the borders of the class of weak definites are gradual this gives people space to interpret definites as either strong or weak in the same situations. It is also possible that people have a class of weak definite which has a strong border, but that this class simply differs per person. People's personal experience might play a much bigger role with weak definites than previously assumed.

6.2.2 Revision of the experiment

In this thesis an experiment was performed in order to get a correct image of how people interpret definites with the noun in question. It has been pointed out before that some confounding factors might have influenced the results. In order to be able to make conclusions which are as certain as possible it is important to remove all doubt about the experiment. Due to time constraints it is not possible to re-do the test, but it is possible to provide suggestions as to how it should be done.

The first possible confounding factors is that of the context given. Because the context was given in a 'newspaper article' which had one of the instantiations of the noun as main subject, it is possible people took this to mean the definite was strong and referred to this instantiation. A possible solution to this problem is to give the context visually instead of textually. The goal of the context was to ensure the test subjects were aware that more than one instantiation of the noun exists in the area. The textual context could thus be replaced with a visual context like a picture of a street with several restaurants, pointing out the different restaurants in the area. Figure 2 is an example of what this could look like. This way no one instantiation would be primed more than another.



Figure 2 – Possible context experiment.

The second possible confounding factor had to do with the question itself. The problem with a sentence like 1 is that there is no referent for *the hotel*, but because *the* is strong,

it does need one in order to be interpreted. This may have caused people to assume the definite was weak, because a sentence like 2 with a weak definite can refer without a referent.

1. Mary works in the hotel and Anne too.
2. Mary is in the hospital and Anne too.

The solution to this confounding factor might in fact be the same as to the first confounding factor. When giving the subjects a visual context like figure 2, *de McDonalds* could refer to one of the restaurants on the image. If the definite were weak this could be any of the restaurants, if the definite were strong it would be the single restaurant placed apart from the others. This way all noun phrases would refer and the problem would be solved.

This solution also gives rise to another possibility. The test done by Carlson and Sussman(2005) involved the same kind of image. Instead of asking if a noun phrase can refer to multiple instantiations, they asked their participants to point out the restaurant Mary went to in a sentence like 3.

3. Mary went to the McDonalds

If the singleton restaurant was pointed out significantly more than the others this means the subjects interpreted the definite as strong. If all restaurants were pointed out equally as much, the subjects interpreted the definite as weak. This test worked very well for the nouns Carlson and Sussman used, but because of the great dependence on many test subjects it might not work as well when the interpretation of the nouns differs per person.

References

Aguilar, A. (2008). Uniqueness & weak definites in Spanish

Aguilar A. and Zwarts J. (2010) Weak definites and reference to kinds

Abbott, B. (2000). Definiteness and identification in English. *Pragmatics in 2000: Selected papers from the 7th International Pragmatics Conference*, pages 1-15.

Birner, B. and Ward, G. (1994). Uniqueness, familiarity, and the definite article in English. *In Berkeley Linguistics Society*, volume 20, pages 93- 102.

Carlson, G. and Sussman, R. (2005). Seemingly indefinite definites. *Linguistic evidence: Empirical, theoretical, and computational perspectives*, pages 26-30.

Epstein, R. (2002). The definite article, accessibility, and the construction of discourse referents. *Cognitive Linguistics*, 12(4), 333-378.

Langacker, R. (1987). Foundations of Cognitive Grammar

Appendix 1.

Disciplines of test subjects:

Literatuurwetenschap	Literature science	21
Cognitieve kunstmatige intelligentie	Cognitive artificial intelligence	21
Spaanse taal & cultuur	Spanish language & culture	13
Taalwetenschappen	Linguistics	12
Taal & cultuurstudies	Language & culture studies	10
Portugese taal & cultuur	Portuguese language & culture	6
Liberal arts & sciences	Liberal arts & sciences	5
Engelse taal & cultuur	English language & culture	4
Keltische talen & cultuur	Celtic languages & culture	4
Wiskunde	Mathematics	4
Franse taal & cultuur	French language & culture	2
Nederlands	Dutch	1
Duitse taal & cultuur	German language & culture	1
Italiaanse taal & cultuur	Italian language & culture	1
Vertalen	Translating	1
Onderwijskunde	Didactics	1
Pedagogische wetenschappen	Pedagogic sciences	1
Culturele antropologie	Cultural anthropology	1
Master Taal, mens & maatschappij	Master Language, humans & society	1
Onbekend	Unknown	5

Appendix 2.1 Test version 1

1.

Marie las de krant en Anne ook.

Vraag:

Lazen Anne & Marie dezelfde krant?

- Ja
- Nee
- Het kan beide

2.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e Griek geopend in Lutjebroek, een mooie aanvulling op de al bestaande Griek.'

Marie ging eten bij de Griek en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

3.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e grillroom geopend in Lutjebroek, een mooie aanvulling op de al bestaande grillroom.'

Marie ging eten bij de grillroom en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

4.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e Burger King geopend in Lutjebroek, een mooie aanvulling op de al bestaande Burger King.'

Marie ging eten bij de Burger King en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

5.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is het 2^e wokrestaurant geopend in Lutjebroek, een mooie aanvulling op het al bestaande wokrestaurant.'

Marie ging eten bij het wokrestaurant en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

6.

'Uit: Lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e dokterspraktijk geopend in Lutjebroek, een mooie aanvulling op de al bestaande dokterspraktijk.'

Marie ging naar de dokter en Anne ook.

Vraag:

Gingen Anne & Marie naar dezelfde dokter?

- Ja
- Nee
- Het kan beide

7.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e McDonalds geopend in Lutjebroek, een mooie aanvulling op de al bestaande McDonalds.'

Marie ging eten bij de McDonalds en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

8.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is het 2^e eethuis geopend in Lutjebroek, een mooie aanvulling op het al bestaande eethuis.'

Marie ging eten bij het eethuis en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

9.

'Uit: Lokale krant van Anne & Marie':

'... afgelopen maand is het 2^e hotel geopend in Lutjebroek, een mooie aanvulling op het al bestaande hotel.'

Marie ging naar het hotel en Anne ook.

Vraag:

Gingen Anne & Marie naar hetzelfde hotel?

- Ja
- Nee
- Het kan beide

10.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e crêperie geopend in Lutjebroek, een mooie aanvulling op de al bestaande crêperie.'

Marie ging eten bij de crêperie en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

11.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e La Place geopend in Lutjebroek, een mooie aanvulling op de al bestaande La Place.'

Marie ging eten bij de La Place en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

12.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is het 2^e afhaalrestaurant geopend in Lutjebroek, een mooie aanvulling op het al bestaande afhaalrestaurant.'

Marie ging eten bij het afhaalrestaurant en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

13.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e Indiër geopend in Lutjebroek, een mooie aanvulling op de al bestaande Indiër.'

Marie ging eten bij de Indiër en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

14.

Marie las het tijdschrift en Anne ook.

Vraag:

Lazen Anne & Marie hetzelfde tijdschrift?

- Ja
- Nee
- Het kan beide

15.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is het 2^e sterrenrestaurant geopend in Lutjebroek, een mooie aanvulling op het al bestaande sterrenrestaurant.'

Marie ging eten bij het sterrenrestaurant en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

16.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e Italiaan geopend in Lutjebroek, een mooie aanvulling op de al bestaande Italiaan.'

Marie ging eten bij de Italiaan en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

17.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is het 2^e wegrestaurant geopend in Lutjebroek, een mooie aanvulling op het al bestaande wegrestaurant.'

Marie ging eten bij het wegrestaurant en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

18.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e Nigeriaan geopend in Lutjebroek, een mooie aanvulling op de al bestaande Nigeriaan.'

Marie ging eten bij de Nigeriaan en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

19.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e Carestel geopend in Lutjebroek, een mooie aanvulling op de al bestaande Carestel.'

Marie ging eten bij de Carestel en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

20.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e bistro geopend in Lutjebroek, een mooie aanvulling op de al bestaande bistro.'

Marie ging eten bij de bistro en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

Appendix 2.2 Test version 2

1.

Marie las de krant en Anne ook.

Vraag:

Lazen Anne & Marie dezelfde krant?

- Ja
- Nee
- Het kan beide

2.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is het 2^e pannenkoekenrestaurant geopend in Lutjebroek, een mooie aanvulling op het al bestaande pannenkoekenrestaurant.'

Marie ging eten bij het pannenkoekenrestaurant en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

3.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e AC geopend in Lutjebroek, een mooie aanvulling op de al bestaande AC.'

Marie ging eten bij de AC en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

4.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e Chinees geopend in Lutjebroek, een mooie aanvulling op de al bestaande Chinees.'

Marie ging eten bij de Chinees en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

5.

'Uit: Lokale krant van Anne & Marie':

'... afgelopen maand is het 2^e hotel geopend in Lutjebroek, een mooie aanvulling op het al bestaande hotel.'

Marie ging naar het hotel en Anne ook.

Vraag:

Gingen Anne & Marie naar hetzelfde hotel?

- Ja
- Nee
- Het kan beide

6.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e steakhouse geopend in Lutjebroek, een mooie aanvulling op de al bestaande steakhouse.'

Marie ging eten bij de steakhouse en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

7.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e Thai geopend in Lutjebroek, een mooie aanvulling op de al bestaande Thai.'

Marie ging eten bij de Thai en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

8.

Marie las het tijdschrift en Anne ook.

Vraag:

Lazen Anne & Marie hetzelfde tijdschrift?

- Ja
- Nee
- Het kan beide

9.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is het 2^e petit restaurant geopend in Lutjebroek, een mooie aanvulling op het al bestaande petit restaurant.'

Marie ging eten bij het petit restaurant en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

10.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e van der Valk geopend in Lutjebroek, een mooie aanvulling op de al bestaande van der Valk.'

Marie ging eten bij de van der Valk en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

11.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is het 2^e snelbuffet geopend in Lutjebroek, een mooie aanvulling op het al bestaande snelbuffet.'

Marie ging eten bij het snelbuffet en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

12.

'Uit: Lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e dokterspraktijk geopend in Lutjebroek, een mooie aanvulling op de al bestaande dokterspraktijk.'

Marie ging naar de dokter en Anne ook.

Vraag:

Gingen Anne & Marie naar dezelfde dokter?

- Ja
- Nee
- Het kan beide

13.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is het 2^e zelfbedieningsrestaurant geopend in Lutjebroek, een mooie aanvulling op het al bestaande zelfbedieningsrestaurant.'

Marie ging eten bij het zelfbedieningsrestaurant en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

14.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e Applebee's geopend in Lutjebroek, een mooie aanvulling op de al bestaande Applebee's.'

Marie ging eten bij de Applebee's en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

15.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e pizzeria geopend in Lutjebroek, een mooie aanvulling op de al bestaande pizzeria.'

Marie ging eten bij de pizzeria en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

16.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e Zweed geopend in Lutjebroek, een mooie aanvulling op de al bestaande Zweed.'

Marie ging eten bij de Zweed en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

17.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e Fifteen geopend in Lutjebroek, een mooie aanvulling op de al bestaande Fifteen.'

Marie ging eten bij de Fifteen en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

18.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e rotisserie geopend in Lutjebroek, een mooie aanvulling op de al bestaande rotisserie.'

Marie ging eten bij de rotisserie en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

19.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is de 2^e Turk geopend in Lutjebroek, een mooie aanvulling op de al bestaande Turk.'

Marie ging eten bij de Turk en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

20.

'Uit: Dinerrubriek van de lokale krant van Anne & Marie':

'... afgelopen maand is het 2^e drive-inrestaurant geopend in Lutjebroek, een mooie aanvulling op het al bestaande drive-inrestaurant.'

Marie ging eten bij het drive-inrestaurant en Anne ook.

Vraag:

Gingen Anne & Marie eten bij hetzelfde restaurant?

- Ja
- Nee
- Het kan beide

Appendix 2.3 Test Introduction

Inleiding

Dit is een testje naar het gedrag van de lidwoorden 'de' en 'het'. Die gedragen zich niet altijd hetzelfde.

Een voorbeeld:

Bij de zin '**Marie las de krant en Anne ook**' hoeven Anne en Marie niet per se dezelfde krant te lezen.

Echter, bij '**Marie las de roman en Anne ook**', lezen Anne en Marie wel degelijk dezelfde roman.

Op de komende pagina's staan meer van dit soort zinnen. De vraag is of deze zinnen zijn als de zin met 'krant', of als de zin met 'roman'. Het antwoord kan verschillen per vraag. Er bestaan geen 'goede' of 'foute' antwoorden bij deze test; het is belangrijk dat u het invult volgens uw eigen intuïtie.

Introduction

This is a test into the behaviour of the articles 'de' and 'het'. They don't always behave the same.

An example:

With the sentence '**Mary read the newspaper and Anne too**', Anne and Mary don't necessarily have to read the same newspaper.

However, with the sentence '**Mary read the novel and Anne too**', Anne and Mary do in fact read the same novel.

On the following pages are more of this kind of sentence. The question is if these sentences are as the sentence with 'newspaper', or as the sentence with 'novel'. The answer can differ per question. There are no 'right' or 'wrong' answers in this test; it is important that you fill it in following your own intuition.

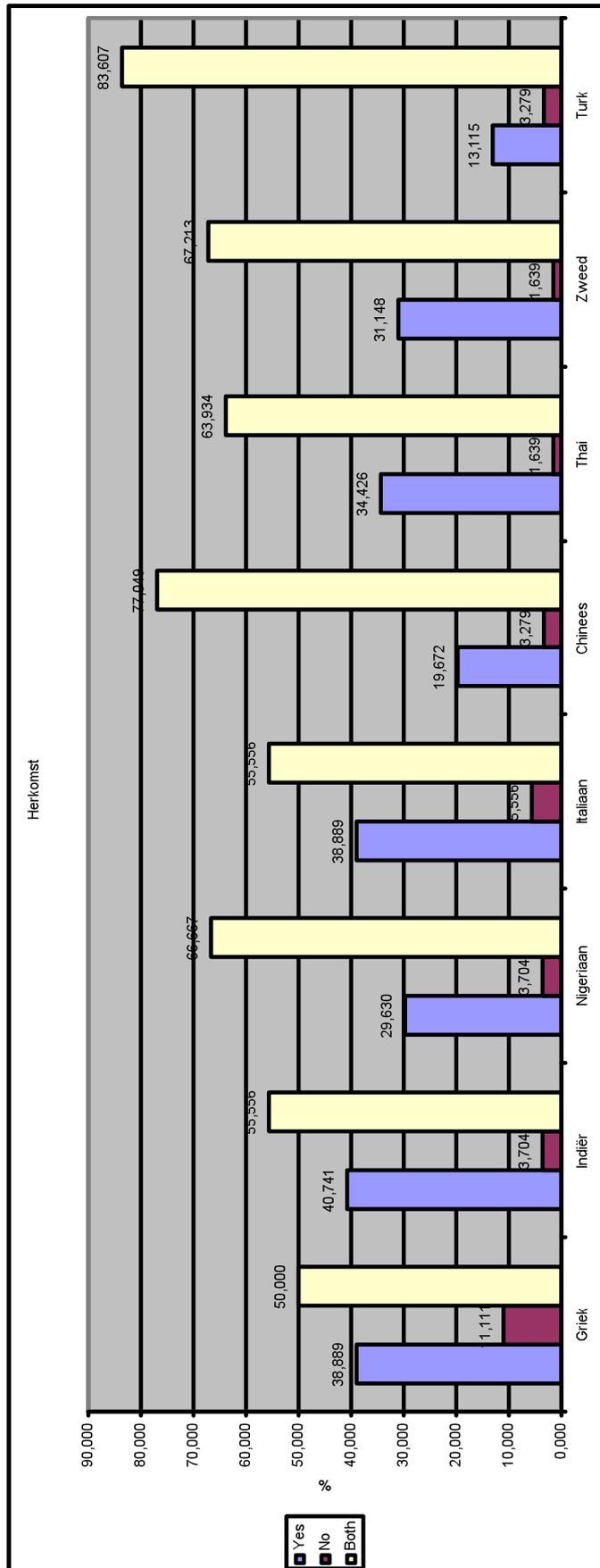
Appendix 3. Results, sorted from strong to weak.

Word:	Yes	No	Both	% Yes	% No	% Both	Highest answer:	Difference with the follow-up.
Pannenkoekenrestaurant	45	1	15	73,770	1,639	24,590	Yes	49,18033
Petit restaurant	39	0	22	63,934	0,000	36,066	Yes	27,86885
Sterrenrestaurant	31	2	20	57,407	3,704	37,037	Yes	20,37037
Eethuis	29	1	24	53,704	1,852	44,444	Yes	9,259259
Crêperie	29	1	24	53,704	1,852	44,444	Yes	9,259259
Steakhouse	32	2	27	52,459	3,279	44,262	Yes	8,196721
Wegrestaurant	28	1	25	51,852	1,852	46,296	Yes	5,555556
Zelfbedieningsrestaurant	32	0	29	52,459	0,000	47,541	Yes	4,918033
Wokrestaurant	27	2	25	50,000	3,704	46,296	Yes	3,703704
Afhaalrestaurant	25	3	26	46,296	5,556	48,148	Both	1,851852
Snelbuffet	28	3	30	45,902	4,918	49,180	Both	3,278689
Griek	21	6	27	38,889	11,111	50,000	Both	11,11111
Indiër	22	2	30	40,741	3,704	55,556	Both	14,81481
Bistro	22	1	30	41,509	1,887	56,604	Both	15,09434
Nigeriaan	21	3	30	38,889	5,556	55,556	Both	16,66667
Carestel	21	3	30	38,889	5,556	55,556	Both	16,66667
Grillroom	21	2	31	38,889	3,704	57,407	Both	18,51852
Drive-inrestaurant	23	1	36	38,333	1,667	60,000	Both	21,66667
AC Restaurants	23	1	37	37,705	1,639	60,656	Both	22,95082
Thai	21	1	39	34,426	1,639	63,934	Both	29,5082
La Place	18	2	34	33,333	3,704	62,963	Both	29,62963
Italiaan	16	2	36	29,630	3,704	66,667	Both	37,03704
Zweed	19	1	41	31,148	1,639	67,213	Both	36,06557
Rotisserie	15	0	46	24,590	0,000	75,410	Both	50,81967
Vd Valk	14	2	45	22,951	3,279	73,770	Both	50,81967
Fifteen	13	3	45	21,311	4,918	73,770	Both	52,45902
Applebee's	12	3	46	19,672	4,918	75,410	Both	55,7377
Chinees	12	2	47	19,672	3,279	77,049	Both	57,37705
Burger King	9	4	41	16,667	7,407	75,926	Both	59,25926
Pizzeria	8	4	49	13,115	6,557	80,328	Both	67,21311
McDonalds	7	2	45	12,963	3,704	83,333	Both	70,37037
Turk	8	2	51	13,115	3,279	83,607	Both	70,4918

Control Items:									
Tijdschrift	98	1	15	85,217	0,870	13,043	Yes	72,17391	
Hotel	74	2	39	64,348	1,739	33,913	Yes	30,43478	
Krant	5	19	91	4,348	16,522	79,130	Both	62,6087	
Dokter	5	15	95	4,348	13,043	82,609	Both	69,56522	

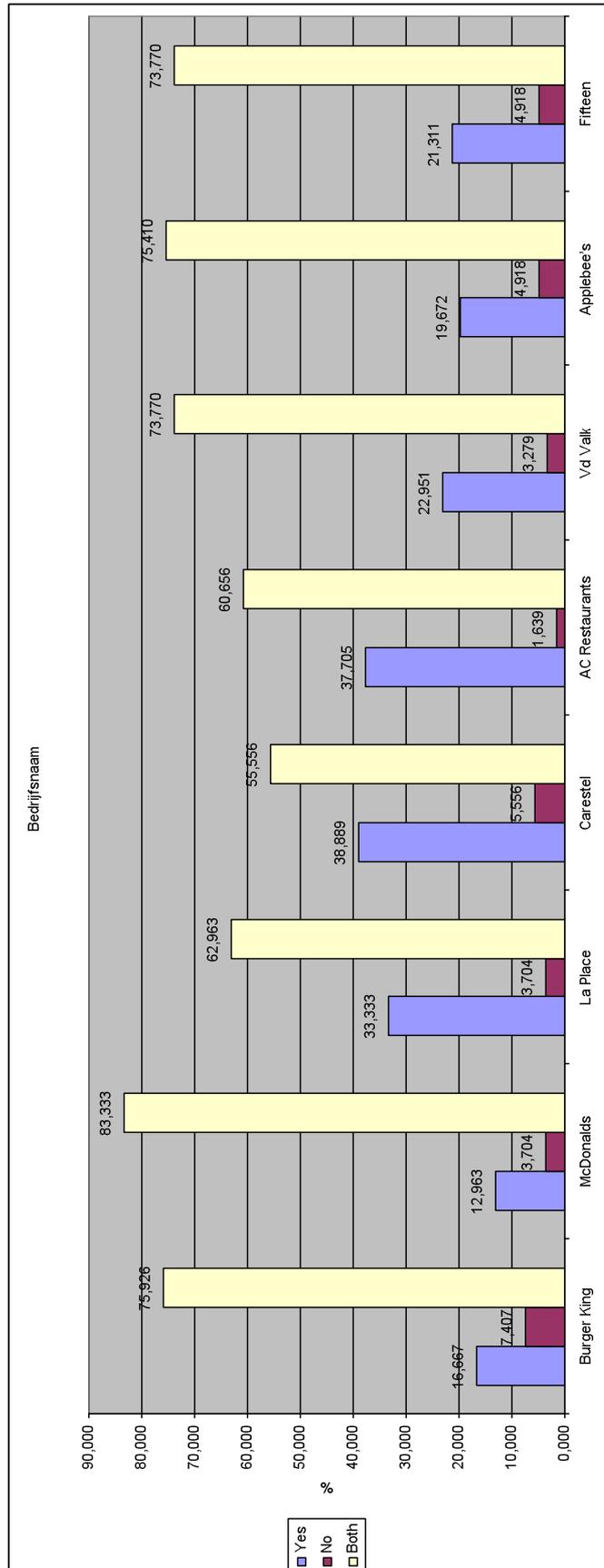
Appendix 3.2.1 Results – Graphs

Answers in percentages – Words for the origin of the food.



Appendix 3.2.2 Results – Graphs

Answers in percentages – Words for the company name.



Appendix 3.2.3 Results – Graphs

Answers in percentages – Words for kind of food.

