Connecting the unconnectable: quantifying pragmatics

Presented to the
Leiden Institute for
Advanced Computer Science (LIACS)

Leiden University

in partial fulfillment of the
requirement for the degree of the
Master of Science

Media Technology

By

Monica Preller

Student ID: 1406930
Supervisor 1: Max van Duijn
Supervisor 2: Maarten Lamers

January 16, 2020

Leiden, The Netherlands

INDEX

ABSTRACT
The larger background theme of this thesis is whether computational methods can be used to abstract pragmatic language phenomena. This theme is explored through a theoretical analysis, based on existing work, plus an experiment in which participants' judgements of a set of sentence pairs are compared with a computerised analysis of these pairs.

The paper is ordered as follows. First, the concept of *pragmatics* as it is used in this thesis will be defined. This will be followed by a short introduction to the theory on language use by the language philosopher Grice, who specialised in pragmatics. Gricean maxims will be explained and what role they play in communication. Succeeding that, a review of previous attempts to convert human language interpretation to a calculative process (in part inspired by) Gricean theory will be presented. This includes a subsection to introduce the distributional semantics hypothesis, the theory that underlies the methodology. After that, the hypothesis for the empirical study is presented, followed by an explanation of the methodology and a discussion of particular methodological choices. A discussion of the results comes afterwards. In the concluding sections pitfalls and caveats of the study are addressed, and suggestions for future research are made.

Personal note
Scientific research serves the sole and noble purpose of retrieving new knowledge from abstract or concrete observations. This differs from art, that stands on its own and may even serve as the object of scientific studies.

The idea for this thesis, on the other hand, did originate from the standpoint that scientific research can also serve a purpose of making a statement that stands on its own. This resulted in the research presented here; in the process that led to this thesis - the study programme of Media Technology - the division of natural sciences and humanities was exploited, leading to personal frustration and (therefore possibly biased since n=1) perception of social disbalance of these two faculties. However, without social disbalance no frustration, and without frustration, no art.

It should be noted that in this thesis only a marginal portion of computational capability is explored. In spite of this, it is to be hoped that some part of the answer to the question of a machinal comprehension of such a humanitarian concept (language) is solved.

Introduction

If the Ford car is the symbol for technological innovation in the first decade of the twentieth century, a human-simulating machine would be that of the decade in which we live now. However, this picture of a robotized human has been around more than a hundred years (in literature, this possibly is the publication of the novel *The Nyctalope on Mars* by Jean de la Hire in 1911, starring a so-called nyctalope – an individual bearing organic and mechanical body parts simultaneously). Nevertheless, no technological device has been able yet to beat a human on all the fields they were built to play on. In communication, an increasing amount of computers interacting with humans are available to the public (Venislav, 2018; Labbe, 2019; Schebella, 2019). These are found in smartphone applications such as Siri, hardware devices like Google Home and in online customer service.

Computer models that aim to reproduce communication have a hard time taking into account certain aspects of interaction that are so obvious to dialogue partners. For example, if someone asks "Should I bring my ticket before departure?", this makes sense if the question is addressed to a steward at an airport. Contrastingly, the question seems odd if used on a cashier of a merry-go-round at a fair. Information contributing to this difference in utterance content resides in the domain of pragmatics, the invisible but inevitable factor for the communicator to take into account. A difference in pragmatic information can grow culturally, which makes it hard to trace or pinpoint. Consequently, it is difficult to incorporate this in computer models built for simulating human interaction. It forms a reason to why still most computerised agents that are built for interacting with humans come across as silly when really put to the test. Eventually, they would fail a test in which their humanity would be questioned, such as the Turing test. During the latter, a female needs to prove her dialogue partner to be male instead of the female he claims to be. In the first part of the test, this role is taken up by a man. In the second part, it is taken up by a computer. The computer passes the test if the female does not experience a difference in difficulty to prove her conversation partner to be non-female in the two different parts of the test.

As mentioned before, implementation of human aspects in technology are at the same time sought-after concepts, especially now that the corporate world thriving by efficiency has seen the possibilities of androids. Take the customer service example: if one would replace a (stereotypical female) callcenter employee with a chatbot, interaction would take place a lot faster and more efficient. The 'employee' would not get tired or emotionally drained by annoyed customers who are forcing their frustrations onto her or him: no, this ideal mechanical conversational agent could just keep on going without requiring any breaks. However, the digital agents as they are available to the public now cannot make distinctions between the child demanding "green stuff" or the adult, so they are lacking very important skills that humans have. Whatever would resolve this, it requires some form of 'translating' the qualitative information humans perceive in communication to a quantitative entity.

The topic and experiment of this thesis was inspired by the non-existent overlap between these sides: after all, how beautiful would it be if the sciences have able to have a framework for humanities, just as humans try to grasp the natural world through language? That way, a computer (or perhaps, quite more disrespectful, a 'computer minded' person even) would have a way of comprehending human language like a very skilled communicator would.

This thesis makes an attempt to connect something that is easily measurable by a computer algorithm and something that is mostly based on a human 'feeling'. By doing so, two research traditions from the worlds of the humanities and sciences are connected. It tests whether distance between words in a sentence can be used as a proxy for felt pragmatic relevance measured in a sample of speakers. This concept – the relevance maxim as established by the highly influential language philosopher Grice – carries a lot of information for both speakers. This will also be explained later on in this thesis. First of all, however, more of the relevance of the research needs to be explained. Therefore, a very brief history of conversion of human language to a machine is presented in the upcoming chapter. After that, this thesis' definition on the study area of pragmatics is to be read aiming to a founded understanding of the topic. As the theory by Grice is regarded as a field within pragmatics, but also a specific focus of this thesis' topic, attention is paid to this subject specifically in the next chapter. After reading on these three elements, the actual experiment that was carried out in order to produce this research's results is presented. This is followed by the results, conclusion and discussion. An appendix is added including research material.

Now, let us tiptoe in the water between the land of humanities and science by reading the first chapter.

CONTENT

1 Formalisation of language processing

1.1 Algorithmic models & brief history of natural language processing

1.1.1 Introduction

As written in the previous introduction, linguistics and computer science seem to be worlds apart to anyone unfamiliar with the topics. However, according to a leading computational linguist Jurafksy (2001: p. 3) early work in computer science drew very strongly from linguistic literature. Upon closer inspection of Grice (1975, p. 33), who was aiming to convert language processing into a calculated concept by providing reasoning paths for deducing meaning from sentences, a connection between linguistics and computer science becomes more contingent. See for example the fragment below, proposed by Grice (1989: p. 276) on how a speaker would expect a hearer to reason after uttering *D* (the affirmative form of a sentence with a definite description as a subject, for example "The house on the hill is red"):

1. Speaker *S* has uttered the negation of *D*; so s/he is speaking as if s/he were responding negatively to *S*, that is, to one who utters *D*
2. So *S* is fulfilling the expectations that *S* would have had about *H* (that is, s/he is accepting that there is no situation possible in which *D* takes place **and** does not occur or that it is because of the environment ['the world'] in which *D* exists is false)

Searle, another influential linguist, provided (almost literary) reasoning paths in which language would be processed for when a certain communicative act (in language philosophical terms: speech act) takes place. For example, he (1969: p. 57-61) states that a promise comes about as follows:

1. The utterance takes place through a channel disposing of sufficient means of input and output
2. A speaker *S* expresses the proposition that *p* in the utterance of a sentence *T*
3. In expressing that *p*, *S* predicates a future act *A* of *S*
4. The hearer *H* would prefer *S*'s doing *A* to his not doing *A*, and *S* believes *H* would prefer his doing *A* to his not doing *A*
5. It is not obvious to both *S* and *H* that *S* will do *A* in the normal course of events
6. *S* intends to do *A*
7. *S* intends that the utterance of *T* will place him under an obligation to do *A*
8. *S* intends to produce in *H* the knowledge (*K*) that the utterance of *T* is to count as placing *S* under an obligation to do *A. S* intends to produce *K* by means of the recognition of the previous, and he intends i-1 to be recognised in virtue of (by means of) *H*'s knowledge of the meaning of *T*
9. The semantical rules of the dialect spoken by *S* and *H* are such that *T* is correctly and sincerely uttered if and only if conditions 1-8 sustain.

If one would be able to convert the reasoning paths to a programming language, which is in its turn ran by a machine capable enough of understanding the nuances in the paths such as "future act" and "recognition", the machine would – in theory – be capable of producing natural language. Obviously, technological innovation has not reached the point yet in which natural language can be developed. Over time, inventions to do so have improved still. To make the purpose for writing of this thesis – investigating an opportunity for calculating pragmatics – more clear, background history of natural language processing is provided. After this, it should be clear why the research was worth taking place.

1.1.1.1 1940s

The first investigations on computational processing of language did not begin later than in the 1940s (Jones, 2001) with automated translation by a.o. Weaver (1949). At that time, the first mechanical computer was already older than a century with the invention of Charles Babbage's *Difference Engine* (Doron, 2002: p. 16). What's more, the first time mechanical processing of language (which is now generally known under the umbrella term of *natural language processing*) was published on was arguably in 1950 in the article "Computing Machinery and Intelligence" (Turing, 1950). In this a thought experiment is described that has a computer output coherent language. In spite of this, it is not described how the machine would exactly produce its contribution to the conversation, let alone whether this would be automated or not.

1.1.1.2 1950s

In the early 1950s, research on natural language processing was commercially invested in by the (now) multinational company International Business Machines Corporation (IBM). This was directed towards a project to translate Russian to English. Many scholars working on this project and the research field of natural language processing in general arrived from a background and established status in linguistic research. They were motivated by a belief that something practically useful could be realised, which resulted in enthusiasm and optimism surrounding translation research. Another cause for this was that it was attempted to use a new tool for non-numerical data-processing purposes, which was never done before. Therefore, nothing could really "go wrong" - it was a matter of trial and error. Most problems the researchers encountered were of syntactic and semantic nature with ambiguity playing the biggest role (Jones, 2001). Including pragmatics was not yet in sight.

1.1.1.3 1960s

In the 1960s more emphasis was made to include 'common sense' world knowledge (Jones, 2001). In order to do so, some pragmatics had to be incorporated in the way language was processed by mechanics. Some of the tries carried out achieved new inventions, as Weizenbaum (1964) demonstrated with his computer program ELIZA. The most well-known script of ELIZA is probably DOCTOR, which mimics a psychotherapist (Bassett, 2019). In order to do so, pattern matching was used. This is the checking of a certain string of tokens for some pattern by the program. Some of ELIZA's users assigned human-like emotions to the script (Weizenbaum, 1976: p. 3). It should be noted that ELIZA did not process any part of the conversation through more complex recognition such as parsing or by applying a neural network. When a neural network is applied in natural language processing, linguistic features (such as word meaning) are aimed to abstract from natural language by using and labelling the latter as input and that would result in computer recognition of the features – in other words, a computer is 'taught' features of language by feeding it very small parts of the language that it needs to recognise over and over. One important difference between parsing and using neural networks in natural language processing, is that the first one requires little input since a researcher needs to come up with rules that, for example, 'put' linguistic elements in their right grammatical category – in neural networks, a lot of data is required). Furthermore, ELIZA did not notice any pragmatic differences of an utterance. However, this was not so much of an issue since the meanings of the text entered by users was very simple and as a result not differentiated by pragmatics. Machine translation research was almost terminated in 1966 by the Automatic Language Processing Advisory Committee, established in 1964 which claimed the absence of "immediate or predictable prospect of useful machine translation" (ALPAC, 1966: p. 32). However, they did carry out some successful attempts in the area of syntax (Jones, 2001).

1.1.1.4 1970s

Still, natural language processing research did prevail as more tangible results demonstrate 1970s with the rise of more automated conversational agents. However, in contrast to ELIZA, these conversational agents used conceptual ontologies (Schank et al., 1973; Meehan, 1976; Lehnert, 1977), which are labellings of categories and relations between concepts. Thus, computer scientists had to include some pragmatics as well as this may decide upon the referent (and, therefore, meaning) of a conceptualisation which was a.o. published on by Montague (1976). Nevertheless, the rules that had to be composed in order to carry out the (pragmatic) distinctions that made up the ontologies were created manually. It seemed that the non-literal meaning of language is especially hard to grasp by a machine. Therefore, reveals the importance of this thesis' research in which it is attempted to abstract one of these non-literal, 'soft power' aspects of linguistic meaning.

1.1.1.5 1980s

In the late 1980s a revolution in natural language processing came about (Jones, 2001) caused by the increase of computational power and the use of machine learning algorithms presenting output based on statistics instead of rule-based systems developed in the 1970s (Foote, 2019). Sometimes, this use of natural language processing was labelled 'Good Old-Fashioned AI' regarding the heavy reliance of grammar rules for grammaticality judgement (Jackson & Moulinier, 2008: p. 7) – still (in, for example assigning grammatical categories to language) words had to be manually labelled somehow, whether this would be a result of parsing or by have a human judging the words. The improved network from the 1970s receives an assigned feature of language and uses it in order to judge other words on grammar. No specific attention was paid to pragmatics since the words were statistically treated in isolation, i.e., it would not matter for processing whether they were uttered in a context of another register for their

meaning. As an illustration one could imagine the word *neger* ("nigger") in a Facebook post by a private caucasian person versus retrieval of the word from a school book of history for 12-years-old children. Nevertheless, it is described in the influential article "Minds, Brains and Programs" by language philosopher Searle (1980) whether a machine would grasp some of the non-literal meaning of words. In this paper, a computer processes language and it is questioned whether the machine understands or merely simulates communication. Most noteworthy for this thesis is the part in which a person replaces the computer in the thought experiment mentioned in 1950 by Turing. The computer receives Chinese messages and translates them to English with the help of a written version of the program that was run on the computer. In other words, the program that processes the language still would not *understand* language; a machine gets an input of tokens, takes action following rules by a user and provides output accordingly - there is no actual grasp of the language (Searle, 1980). This suggests that according to Searle, pragmatics (being the invisible part of communication as physical context also makes up for it) is not graspable by a machine. What's more, if the latter would be a necessary ingredient for language understanding, this entails that a computer programme would not be capable of language understanding. This is confirmed in the terminology framework set up by Jackson & Moulinier (2001: p. 3), who make a distinction between NLP and NLU (*natural language understanding*).

It is noteworthy that the statistical processing of language is very closely related to the field of distributional semantics. Given the relevance for the experiment discussed in chapter 4 of this thesis, this topic will be introduced more elaborately.

1.1.1.6 Distributional semantics

1.1.1.6.1 Introduction

The earliest the term "distributional semantics" was used was in 1954 by Harris (Harris, 1954), who explained it as "an environment of an element A is an existing array of its co-occurrents, i.e. the other elements, each in a particular position, with which A occurs to yield an utterance".

The hypothesis was brought under wider attention (Bartsch & Evert, 2014) by Firth (1957, p. 11), who illustrated it as "you shall know a word by the company it keeps". Another widely cited definition of the distributional semantics theory comes from Rubenstein & Goodenough (1965), who argued that "words with similar meanings will occur with similar neighbours if enough text material is available".

However, due to influence from generative linguistics, the distributional semantics hypothesis was pushed aside (Lenci, 2008). After all, according to influential publications such as Chomsky's *Syntactic Structures* (1957) meaning is obtained from what is internalised in an ideal speaker. Adherence of meaning (through, from Chomsky's point of view, grammar) is something a speaker is 'born' with. Abstracting this from the way words are distributed would is, contrastingly, a bottom-up view. Langacker (1998) also challenged the hypothesis of distributional semantics by publishing in his widely-cited paper that internalised semantics come about through "not only abstract or intellectual "concepts" but also sensory, motor and emotive experience;". Obviously, this is one pitfall of the distributional semantics hypothesis, connected to the symbol grounding problem as coined by Harnad (1990) – speakers do obtain some meaning of words through sensory experience. After all, speakers for example come up with new words that sound like the things that they refer to.

In research using corpus methods however this caveat is cast aside. This is due to the fact it is within this field that corpora of language are used as the material experimented on, and that correlations are regarded as arguments for linguistic processes outside of the corpora.

1.1.1.6.2 Two versions

According to Lenci (2008), two types of the distributional semantics hypothesis exist: a weak and a strong version. If the strong distributional semantics hypothesis is assumed, meaning of words is **caused** by their surroundings. If the weak version is adopted, it is assumed that words do not obtain meaning from their surroundings initially; however, a correlation between a frequent appearing surrounding of a certain word can indicate meaning of the surrounding transferred to this element, but this is not necessarily so.

Currently, the distributional semantic hypothesis has caused a wave of development of distributional semantic models (DSMs). In these models, meaning of a word is represented as a vector that oscillates between two or more words that are approximated as axes on a graph. According to Sahlgren (2006), there are two types of distributional models: syntagmatic and paradigmatic. Syntagmatic models account for relations between words that co-occur next to each other within the

same text region. Paradigmatic models concern words that do not themselves co-occur but are surrounded by words that are often the same. According to Bruni, Tran & Baroni (2014), distributional semantics is a very good tool to harvest effective meaning representations on a large scale. However, they do not touch upon issues addressed previously by other researchers involving indication of meaning by distributional semantic models. In line with this, Nerbonne (2005) underlines the possibility of using these models to not only explain meaning of words, but about where the words came about (for example, between cultures in language contact) as well.

1.1.1.6.3 Other scientific disciplines

The distributional semantics hypothesis can be spotted in research fields outside of linguistics and computer science. For example, the psychologist Charles (2000) explores in a frequently cited paper the overlap between cognitive and lexicographic processing of synonyms by using their respective contexts. Homayouni et al. (2005) used DSMs to identify gene relationships to contribute to the field of biomedical science. Boleda & Herbelot (2017) investigated DSMs from a philosophical point of view by introducing a framework of Formal Distributional Semantics to improve DSMs.

1.1.1.7 1990s

In the early 1990s, researchers started experimenting inspired by the computational grammar theory connected to logics for meaning representation in order to deal with the speaker's beliefs and intentions, combined with discourse features like emphasis and theme (Jones, 2001). Remarkably enough, computer science seemed to influence linguistic theory by implementing this grammatico-logical approach to a more general, "informal" use of predicate, logic meaning representations. In its turn, this led to a shift in the meaning of pragmatics (Jones, 2001). Statistical processing of language as developed in the 1980s was extended further as strong AI was developed. For example, in strong AI, a machine started to be capable of teaching itself about language by 'updating' the hidden layer in which the features of a linguistic unit – that assign meaning to it – over and over. This resulted in better conversational agents since they now would be capable of treating the material they received from users as input to learn from.

Aside from this, the lexicalist perspective from the 1980s has become increasingly influential. The lexicon has taken over some role of syntax (Jones, 2001), which means that it became more accepted among academic researchers that words contain meaning of their own without requiring a relation with other words. For example, the word "green" would have a meaning without standing in relation to another word that it would modify (say, "tree") even though the latter probably refers to a more concrete entity.

Besides the theoretical influence, Hirschenberg & Manning (2015) note that during this decade more technical related changes occur. First of all, computing power increases even more. Secondly, more and more linguistic data becomes available due to the increased accessibility of internet. Thirdly, the further development of "highly successful machine learning (ML) methods". Fourthly, the importance of how human language is structured and the role of social context in interaction is discovered.

Jackson & Moulinier (2008: p. 7) note that in the 1990s, the approach of statistical analysis of language was increasingly implemented. Sometimes, it is characterised as 'empirical' because it includes the process of extracting language data from extensive digital corpora such as news feeds and web pages. By doing so, the empirical approach includes a designation that acquired some derogatory connotations in the arena of twentieth-century scholarship.

1.1.1.8 2000s until now

According to Jones (2001), in the beginning of 2000s, natural language processing is "flourishing", and it will continue to do so in a way within research that combines formal theories and statistical data (Jones, 2000). Up until now, this seems to be confirmed, as natural language processing applications are invested in to analyse historical texts. Examples of these are OpenSoNaR, a database that contains more than 500 million words that are lemmatised or grammatically labelled, and TextGrid, in which researchers can assign images to text. Initiatives like the Common Language Resources and Technology Infrastructure (CLARIN) and Digital Research Infrastructure for the Arts and Humanities (DARIAH), which are platforms that seek to connect these applications so that they are available to researchers worldwide.. Usually, these projects do not receive any commercial injections; rather, they originate from the humanities as most of them are aimed to exploit the opportunities offered by digital data for humanities research, as Piotrowski (2008: p. 6) points out. Interestingly enough, commercial computer

science seems to draw increasingly from linguistics to make use of the large amount of data to systematically retrieve information from texts, including the linguistic branches of syntax (for the way sentences are grammatically structured) and discourse analysis (for coreference structure across discourses) (Grishman, 2005).

Unfortunately enough, due to the dividedness of the different groups (commercial and non-profit) working on the mechanism of natural language processing, little common ground on frameworks or material, such as data quality, seems to exist. For example, De Roeck (2005) points out that "there is no agreed common framework for approaching the systematic understanding of the role of data".
According to Lamel & Guavain (2008), major advances in spoken language technologies have been achieved. This might have paved the road for the advancements noticed by Hirschberg & Manning (2015), which refers to the extension of the 1980s research focusing on machine translation and text analysis to develop applications such as dialogue systems, text mining on social media and sentiment analysis for products and services. Sometimes, the data itself is applied on the data on their own, as Nicolov & Salvetti (2005) show in their successful spam classifier by making use of URL weblog data. However, it should be noted that Jones (2001) argues that language engineering is different from computational linguistics and natural language processing (Jones, 2001). In Jackson & Moulinier (2002) it is described as the "function of software or hardware components in a computer system which analyze or synthesize spoken or written language". Therefore, it is debatable whether the trend noticed by Hirschberg & Manning (2015) is one of natural language processing or more technical language engineering. The fields seem to overlap, too. Therefore it is relevant to notice here that the latter two predict "a further substantial progress in NLP".

More importantly, they signalise the augmented realisation within the language engineering field that NLP "is not simply about solving engineering problems". It would be due to this trend that "many areas of linguistics are becoming more quantitative".

Lately, most academic research on natural language processing seems to have focused on data collection for medical (Pons et al., 2016; Azab et al., 2019) or humanitory (Hinrichs et al., 2005; Sporleder & Lascarides, 2005) purposes.

1.1.2 Natural language processing - types of algorithm models

Naturally, due to increased calculation performance of computers, ways in which natural language processing is carried out have become numerous. Jurafsky (2001) claims that there are two types of algorithms for speech act interpretation: the Belief, Desire & Intention (BDI) models and cue-based models. In the first, knowledge of speaker and hearer's intentions, actions, knowledge and beliefs are assigned to linguistic input . Abstracting these categories is done by axiomatising (i.e., by reducing reasoning to a stepwise procedure) to actions and planning of speakers in order for them to be understood by hearers. According to Jurafsky (2001: p. 7) every axiomatisation has the following parameters:

1. Preconditions: conditions that must be true before performing the action
2. Effects: conditions that become true as a result of successfully performing the action
3. Body: a set of partially ordered goal states that must be achieved in performing the action.

In cue-based algorithms, outcomes rely on the lexical, syntactic, prosodic and discourse factors provided as input that are assigned to some speech or dialogue act. Examples of the cue-based algorithm type are Nagata & Morimoto (1994), Suhm & Waribel (1994), Warake et al. (1997), Stoleke et al. (1998), Taylor et al. (1998) and Chu-Caroll (1998).

As all models for mimicking natural language, none are perfect. The advantage of a BDI model is that it comes with the symbolic, structural and philosophical paradigms of linguistics because these frameworks arise from the circumstances and entities mentioned in Jurafsy's list above. For example, pragmatics is also a product of a condition that is true before the speech act is carried out (i.e., does a word stand in a school book or an advertisement). However, the disadvantage of this is that in order for a BDI model to work is that each utterance used as an input is treated as having a single literal meaning since there are so many factors that play a role. This meaning, through plan inference rules produces a non-literal and also pragmatic interpretation (Jurafsky, 2001: p. 8). This makes the BDI model unsuitable if a one-sided meaning is desired as an output.

The advantage of a cue-based algorithm is that it selects for quantifiable characteristics, so that it can be applied to a broader spectrum of speech acts. For example, if the intonation of a certain speech

act like a promise always comes with the same referent, this intonation can be applied to other speech acts as well to find out if they are a promise. However, due to not taking unquantifiable aspects into account – such as the emotions that a person is feeling when uttering a sentence –a part of meaning is missed. After all, it could be the case that a person feels restricted from showing these emotions.

1.1.3 Algorithmisation of Gricean theory

In Jurafsky (2001, p. 4), it is pointed out that in order for an automated conversational agent to come across as a human being and – therefore – is capable of "complex language behaviour" (Jurafsky, 2001: p. 3) it requires knowledge of pragmatics since this refers to "knowledge of the relationship of meaning to the goals and intentions of the speaker" (Jurafsky, 2001: p. 4). This also closely resembles what Grice aimed to explain in 1975 as a conceptualisation of communication. Consequently, it is inevitable to take some pragmatics into account when simulating interaction.

Attempts to incorporate the theory by Grice in algorithms have been made since the 1980s. Goodman (1986) looked for cases where faulty referring expressions are generated by humans and put forward a 'Find What I Mean' model for determining what the speaker had intended to refer to. However, in his study no attempt is made to construct utterances by a machine. What's more, the latter did not describe a referring expression that could be applied or would be grammatical in several situations. Instead, it was investigated what characterises a nonsense referring expression. This resulted in a narrower understanding of communication when compared to Grice (1975), as the latter came up with a description of what does make referring expressions, or any interaction, successful.

Dale & Reiter (1994) aimed to create an algorithm that satisfies the maxim of quantity. The latter did so by creating regular expressions that would provide the hearer with just about enough information so that it would be clear what the algorithm-generated expression was referring to. They used only one particular, simple kind of referring expression (along the line of "black dog", in which "black" is stored as a colour in order to differentiate it from any other kind of dog that is also stored). The researchers do mention Grice in their paper, but this is only to illustrate that there is more to communication than semantics or syntax. It is specifically pointed out that this one simple referring expression is used so that "pragmatics can be investigated". It can be deduced from this that at the time pragmatics was a complex issue that does not thrive as a research subject by big data. What's more, they note about the Gricean maxim to be brief is "very expensive to implement".

Grim (2011) made an attempt to simulate the establishment of maxims by Grice through computer simulated models. He does so through making use of a computerised version of the game theory. In his research, the hypothesis that communication emerges as a result of environmental pressure on the basis of individual gains. In other words, communication arises organically not out of empathy (i.e. a hearer **wanting** the speaker to tell her or his story, or because the hearer is interested in the speaker's life) but because a certain way of uttering is simply quicker. The models they use show that over generations, the agents in their game model do 'obey' the maxims of quantity, quality and relevance because if they do not they would be not capable of communicating. Therefore, so latter claims, these Gricean maxims "come for free" when incorporated in such a model. However, there was no actual language generated in their research; instead, binary 'codes' for language were used. It was the agent's goal to produce as little different codes as possible, so that communication would require the least amount of energy. The actual content of the codes was extremely simple, which makes the research by Grim (2011) an unsuitable pillar for this thesis to stand on.

Shahed Sorower et al. (2011) used Grice's maxims as a guideline to simulate information. They 'translated' the maxim of quantity into "domain knowledge", and argued that a speaker would not utter more information than was already in the domain knowledge of the hearer.

It seems that increasing attention is paid to humanities in computer science research. However, incorporating the theory on human language that was established in this area is still not fully demonstrating itself in the applied versions of IT research. After all, and as pointed out in the introduction, Siri would not notice a difference of meaning in the same sentences if this was caused by the context. Before describing the attempt to make this meaning more accessible as carried out in this thesis, it should be pointed out what is understood by the contextual (and therefore pragmatic) feature of meaning. This forms the content of the upcoming chapter.

2 Pragmatics and maxims

2.1 Introduction

Now that it is clear what gap the research in this paper aims to fill, the required content on pragmatics is to be digested. Up until this day, no agreement on defining pragmatics unambiguously has been reached by linguists (see also Ariel (2012: p.24)). Naturally, it is impossible to reach a one dimensional view of the topic because of its philosophical nature. To illustrate this, several handbook definitions on pragmatics will be presented. These are followed by the classification of viewpoints on pragmatics in Ariel's (2012) *Cambridge Handbook of Pragmatics* (p. 24). These viewpoints – or paradigms – have arisen since publications by leading theorists such as Grice and Chomsky, which makes these a reliable categorisation. This inspires this thesis author's definition of pragmatics, in which it is aimed to complement the dense and summarised handbook definitions a little.

2.2 Handbook definitions of pragmatics

One definition of pragmatics can be found in the *The Stanford Encyclopedia of Philosophy* (2015), in which it is stated that pragmatics is "characterized [to be] dealing with the effects of context (...) and utterances". Explicitly mentioned in this work is that this is opposed to logic and semantics, that "traditionally deal with properties of *types* of expressions, and not with properties that differ from use to use".

The vast *Oxford Handbook of Pragmatics* (2017) writes about the concept to be initially broadly defined as "the study of language use in context" (p. 1). What is worth mentioning is that according to the latter two different schools of pragmatics are taught in universities: the so-called Anglo-American component view and the European Continental traditions. From an Anglo-American point of view, pragmatics is just another core component placed in the same set within a linguistics theory. Within European Continental frameworks, pragmatics should be included in the study of linguistics as a functional aspect of all linguistic behaviour (p. 3).

In the handbook *Foundations of Pragmatics* (2011) pragmatics is described as "being fundamentally concerned with communicative action and its felicity in context, investigating action with respect to the questions of what action is, what may count as action, what action is composed of, what conditions need to be satisfied for action to be felicitous and how action is related to context". In short, from this perspective pragmatics is a mechanism that always needs to be defined within a certain context, while it decides upon the context of communication at the same time. The author notes that pragmatics has a "multifaceted nature".

In order to add to these definitions while combining them at the same time, information is drawn from the frameworks by Ariel (2012), which are presented now.

2.2.1 Form/function pragmatics

The first framework mentioned is that of form/function pragmatics. From this point of view, naturally occurring examples are motive for establishing a small subset of pragmatic meanings associated with these example constructions and discourse markers. According to Ariel (1983), Noam Chomsky's generative syntax would be the leading theory behind the ideas of this paradigm.

A research example fitting within this paradigm was provided by Ariel (1983), who analysed a specific apositive construction of Hebrew. She pointed out that this construction was exclusively used to introduce very important people. Along the same line, Prince (1978) found that a so-called *it-cleft* (which reveals itself in a sentence such as "It was a Ferrari 458 that broke down") represents systematically different information than a *wh-cleft* (exposed in "What broke down was a Ferrari 458"): the former represents information that a speaker assumes the hearer is thinking about, whereas the latter represents information the hearer may definitely not know.

These examples are used to prove that grammar is at least partly geared towards communication (Ariel, 2012: p. 30) since pragmatists argued from early on that truth conditions distinguish semantics from pragmatics (Brinton, 2008), and this condition would reveal itself through syntactic constructions. This makes pragmatics inherent to syntax, and, therefore grammar.

2.2.2 Historical/typological pragmatics

The second framework is the historical/typological framework. If the latter is followed, it needs to be assumed that any current grammar of a language is pragmatically motivated. In other words, it is assumed within this framework that the current meaning of an utterance came about through the accumulated influence of a certain recurring context. Then, over time, a conversational meaning became

conventional because it was structurally used mostly in specific situations. Historical and typological pragmatists argue that pragmatics is the driving force behind conventionalisation.

To illustrate a historical/typological interpretation of pragmatics, one could imagine the following situation: two persons are walking in each other's pathways in opposite directions. Both have to sidestep to let the other pass and to have to continue their path. The first time, a random side is chosen to step to; if both unhappily chose another direction (i.e., one chooses "right" and the other one "left"), they both cannot pass. However, if they by coincidence **both** decide to step "right" as the first time to choose the same direction, it would make sense to silently choose to step "right" the next time they have to let the other pass and continue one's own way. Then, a convention is born: it becomes the norm to move right. A more theoretical linguistic example of this would be of the coming about of the meaning of the French affix *se*, which is a so-called reflexive. This is an element that only occurs in combination with a verb that makes the content of the verb reflect back on the argument of the verb. For example, *se* in combination with *laver* - "wash" - means that the argument that is subject of the verb is committed to undergoes the event of washing. Over time, it has become a convention to automatically mention *laver* in combination with *se* as a full infinitive since it is so common to refer to an activity of a reflexive of washing instead of 'just' washing.

2.2.3 Inferential pragmatics

Finally, the framework of inferential pragmatics is put forward. From this perspective, inference is the leading mechanism for conveying a speaker's intended meaning, and, therefore, the pragmatics of the meaning. The inferential pragmatician's view coincides with a Gricean (1989) view on pragmatics (Ariel, 2012: p. 25).

As this thesis treats one of Grice's maxims as a starting point, it is necessary to explain his view in detail. According to Grice, a sharp decision between *natural* and *nonnatural* meaning exists. Pragmatics is the one ingredient that would turn a natural into a nonnatural meaning. For example, if the following sentence would be uttered:

(1) Can you hand me the white bread?

context can decide upon the inference, and, therefore, meaning of the sentence. For example, if the sentence would be spoken by someone to a commensal during breakfast when there are two types of bread available, say white and wheat, it is immediately clear what is to be inferred. However, if this would be uttered on the merchant's side of a bakery's counter to a fellow employee where several types of white bread are sold, what needs to be inferred (most possibly, from the author's own experience as an employee of the type of inn mentioned here) is that "white bread" refers to the bread closest to the employee being addressed. Otherwise, excessive walking behind a counter would take place, which is understandably undesirable regarding limited space. This shows that (physical) discourse situations can decide upon what is meant, i.e. determine inferences made based on an utterance (and thus: determine the meaning). According to inferential pragmatists, this is what can be encapsulated by pragmatics.

2.2.4 This thesis' definition of pragmatics

Now that the different perspectives from which pragmatics are approached have been addressed, it can be pointed out why the handbook definitions on pragmatics are not conclusive. First of all, as it has been mentioned by *The Stanford Encyclopedia of Philosophy* (2015), pragmatics is not something to be regarded as a "type of utterance". However, it is not taken into account that a type of utterance can be very binary-like be established by non-pragmatic factors (syntax) as research within the form/function paradigm argues. What is also unfortunate is that the *Oxford Handbook of Pragmatics* do not unify the Anglo-American and Continental European frameworks. If it is assumed that pragmatics is just another functional aspect of all linguistic behaviour – following the European view – it comes about through all linguistic aspects such as phonology, syntax, etcetera. This clashes with the Anglo-American view a little, which regards pragmatics as a way of assigning meaning to utterances alike syntax or semantics. However, according to the form/function paradigm pragmatic meaning can be a result of syntax as well. Aside from that, if the inferential paradigm is adopted, the speaker's intention forms the biggest contribution to pragmatic value. That way, some psychological factors on the speaker need to be included as well since everyone expresses themselves differently.

Still, what would unify the definitions – in combination with the frameworks - is that the most defining characteristic of pragmatics seems to be everything that is activated except for convention and

literal meaning of words on their own in order to have communication come about. From a historical (pragmatics) framework this seems to be an awkward definition – after all, it is possible for conventions to be influenced by the context. In that sense, it becomes a question when an expression exists as a convention. For that reason, it is assumed in the first line that when the convention is still being influenced by the context it is not a full convention yet. What follows from this is that pragmatics would still be at work. What's more, in that sense, pragmatics can be broadly regarded as the 'invisible force' that contributes to meaning. More literally speaking, one could argue that pragmatics is not about what is being said - what words are used and what they refer to in a dictionary way - but what is meant. Therefore, literal messages, stripped of their conventional force would have no pragmatic meaning at all.

To conclude, "pragmatics" refers in this thesis to the processes that take place to derive meanings from an utterance that are all but a "literal meaning" as the result of the utterance, but without taking into account 'newly invented' communicative styles that vary from person to person. After all, communication is a knife cutting at two sides and it does not work when the speaker's conversational style is not grasped by the hearer.

As the Gricean side of pragmatics – that is, pragmatics being seen as something that is established through the intention of the speaker – plays a main role in the methodology of this thesis, Gricean language theory is explained in the following chapter.

3 Gricean language theory

3.1 The establishment of Gricean pragmatics

Fitting into the newly emerged paradigm of inferential pragmatics, Herbert Paul Grice was one of the pioneers to analytically describe the difference between the literal meaning of an utterance and the concept it refers to (Grandy & Warner, 2017). Grice (1957, 1969, 1975, 1982) argued that there are certain fixed principles that meaning is governed by. Pre-Gricean believes roughly speaking went along the line that language can be described by truth values in logic, as Frege (Zalta, 2019) and Russell (Irvine, 2019) claimed, or that the meaning of linguistic units is entirely found in their use, as Wittgenstein (1985: p. 43) states.

Grice published his pragmatics breakthrough (Chapman, 2005: p. 63) in the paper *Meaning* (Grice, 1957). In this paper, a distinction is made between non-natural and natural meaning. Natural meaning entails the truth value of what is meant. In *Meaning* this is illustrated with the word "mean", which can be illustrated by the following example:

(1) A knock on your door at 2 a.m. means your days are counted.

In the case of (1), "means" can refer to "a knock on your door" being a direct equivalent of "your days are counted". This is what Grice (1957) calls the natural meaning and is exhausted by its truth value alone.

Intuitively, however, a fluent speaker of English senses that in (1) there is more to "mean" than this latter reference – since it not literally addresses "being the equivalent of", it opens up the reference of "might". A proficient user of English can cleverly use non-literal addressing to 'soften' the expression as a wholeGrice explains this by stating that a certain utterance does not have meaning *because* it is used in communication (Grice, 1957). Rather, the added, intuitively sensed meaning of "mean" in (1) refers to is what Grice (1957) calls "nonnatural" meaning. This would be realised through the intention of the speaker.

Still, nonnatural meaning as Grice states could not come about without the speaker also intending for the hearer to recognise the speaker's intention. To illustrate this one could imagine the audience asking for an utterance's meaning; according to Grice, this would be "ask[ing] for a specification of the intended effect". See also Strawson's (1964) analytical explanation of nonnatural meaning:

> "*S* nonnaturally means something by an utterance *x* if S intends (*i*,) to produce by uttering *x* a certain response (*r*) in an audience A and intends (*i*,) that A shall recognize S's intention (*i*,) and intends (*i*,) that this recognition on the part of A of S's intention (*i*,) shall function as A's reason, or a part of his reason, for his response *r*."
> (Strawson, 1964: p. 446)

Few years later, Grice revised his theory by pointing out that nonnatural meaning cannot be addressed by using a calculative approach like the paraphrased bit above (Grice & Alan, 1961). He illustrated this with the declarative sentence (3):

> (3) That box looks red to me.

The one uttering (3) is unsure whether the books looks red in every context possible (in some philosophical theories on possible worlds "context" can also be replaced with "world"). Grice (1961: p. 451) stated that the doubtfulness of the expression does not only result from the words "looks red to me" only taking their semantic value into account. Instead, this uncertainty is inherent to "a general feature" (Grice, 1961: p. 457) which he names as e "one should not make a weaker statement rather than a stronger one unless there is a good reason for so doing" (Grice & Allan, 1961: p. 469). Another example of this is given is (4):

> (4) Rudy is either in *x* or *y*.

In this, just like in (3), the uncertainty of the speaker is inherent to the use of "either...or". However, this is not caused by semantics alone; if this would be the case, Grice (1961: p. 459) says that it would be unnecessary to mention "either...or" because it entails that there is an equal chance that Rudy is in *x* compared to being in *y*. Still, the expression is used, and (4) is weakened. It must be for another reason than a state-of-the-art expression: namely, to express uncertainty.

3.1.1 Grice's maxims and conversational implicatures

All of the above was compiled in the book chapter *Logic and Conversation* in 1975 (Grice, 1975). In this chapter, it is stated that the symbols used in formal logic (i.e. ¬, ∧, ∨ or ∀) do not suffice to account for all expressions made in natural language. For example, if one would utter (5)

> (5) You can take the bus or the train to Budapest.

this would mean that it is possible for the hearer to take a train to Budapest, but that a bus is also counted as a possibility. Therefore, the two are not mutually exclusive, as the symbol used in formal logic ∨ entails.

For that reason, Grice (1975) refutes the concept of an "ideal" language and states that "there must be a place for an unsimplified, and so more or less unsystematic, logic of the natural counterparts of these [formal logic] devices".

This particular logic is governed by the cooperative principle, which encapsulates that whenever an utterance is made, it is made "such as is required, at the stage at which it occurs, by the accepted purpose of the talk exchange". More generally speaking, it entails that whenever a language is produced, the producer tries to make it understandable and the hearer tries to understand. Underlying to this principle are four maxims, labelled as follows (Grice, 1975: p. 47):

1. Quantity
   Do not make your utterance more or less informative than required
2. Quality
   Do not produce an utterance which you are conscious of it being untrue
3. Relevance
   Do not produce irrelevant contributions
4. Manner
   Do not produce obscure and ambiguous - do produce brief and orderly

Whenever a maxim is flaunted, Grice (1975: p. 49) argues that this happens because of the following reasons:

i. It is the speaker's intention to lie
ii. The speaker "opts out" of the cooperative principle (e.g. by saying "My lips are sealed")
iii. It is not possible to fulfil two maxims at the same time
iv. The speaker overtly flouts a maxim

The case of 4. might seem impossible as it may seem that a speaker would never overtly flout a maxim without sticking to the cooperative principle. After all, if the cooperative principle includes that communication arises because the hearer wants to understand and the speaker wants to be understood, overtly flouting a maxim makes communication less opaque. If the maxim is not flouted overtly enough, it would leave a hearer confused about the speaker's intention. Still, the cooperation principle can hold because it can be the case that context or background knowledge would support the implicature to such an extent that the hearer can abstract the meaning of a sentence – the latter just must try a bit harder. See also (6), in which the speaker seems to overtly flout a maxim by not providing enough information:

(6) *Imagine a situation in which two people - a child of 12 years old (S1) and her caretaker (S2). It is the "work goal" of the child for that day to finish her Dutch homework. The caretaker promised the child that she could have two portions of ice cream after she finished her homework. The child just had ice cream. The caretaker knows that the child has not finished the mandatory homework yet.*

(7a) S1: I want more ice cream.
(7b) S2: You didn't do your homework yet.

Literally speaking, (7b) can seem as completely irrelevant information independent of (6a). The meaningfulness of sentence (7b) arises through the conversational implicature. According to Grice (1975), the hearer does this as follows:

1. Process the conventional meaning of the utterance
2. Combine the previous with knowledge on the cooperative principle and the maxims
3. Take into account the (linguistic) context
4. Add knowledge on "other items of background knowledge"
5. Acknowledge that 1-4 are available to both participants and that both participants know or assume this to be the case.

As for (7b), this would work as follows:

1. Process the conventional meaning of the utterance
   S2 declares to S1 that S1 did not do the homework yet
2. Combine the previous with knowledge on the cooperative principle and the maxims
   There is no reason to assume that S2 would be ignoring the cooperative principle. After all, S2 wants to get his message to be understood by S1. It would also be faulty to assume that the maxim would be flouted by S2 covertly, either, since there is no reason for S2 to mislead S1
3. Take into account the (linguistic) context
   Both know that S1 had one portion of ice cream and that it was agreed upon that S1 could have two portions if she finishes her Dutch language homework
4. Add knowledge on "other items of background knowledge"
   Both know that it is up to S2 to decide whether or not S1 can have more ice cream
5. Acknowledge that 1-4 are available to both participants and that both participants know or assume this to be the case.
   Since an agreement had been made beforehand between S1 and S2 that S2 could only have the ice cream if the homework was finished, S1 is able to abstract the conversational implicature from S2 (S2 cannot have the ice cream) when S2 is denying a certain situation.

Now that it is demonstrated how the Grice's theory – intention is leading for the pragmatic value of an utterance – works in practice, language processing by machines is touched upon.

4 Methodology

4.1 Introduction

Having taken notice of the relevant background information, the content of the experiment can be touched upon. As mentioned in the introduction**,** in this thesis, an attempt is made to simulate Grice's maxim of relevance by slightly modifying a sentence pair he used in his article from 1975 to illustrate that particular entity. Therefore, the research question is:

> *Can distributional semantics be used to quantify the qualitative concept of the Gricean maxim of relevance?*

A schematised version of the hypothesis, also mentioned in the introduction, is:

> *H_0 : Distribution of two words does not play a role of there being a Gricean relation of relevance between them*
> *H_1: Distribution of two words does play a role of there being a Gricean relevance relation between them; namely, that if the words appear close to each other, this does lead to the latter relation.*

Of course, it is impossible to measure distribution of all words available in all natural speech (thus, natural speech occurring in all situations thinkable) compared to one another that might possibly stand in a Gricean relation of relevance. A solution for this was found in using a corpus that would match natural speech from one language as closely as possible.

As mentioned in 2.2 (for a more precise explanation another look at (7) is advised), a Gricean relation of relevance can make at-first-glance nonsense utterances 'make sense' by involving the situation into it.[1] A relation between the activity of getting ice cream and doing homework are in completely different semantic domains. However, they become relevant to each other through the background knowledge that two speakers have of the situation.

Thus, it became a challenge to simulate the Gricean relevance relation by having (in this case minimalised to) two different words, abstracted by statistical methods from a corpus, presented in a manually composed dialogue. In order to find out how reliable this way of abstracting a Gricean relevance relation actually would be, this dialogue needed to be presented to human participants. After all, then it would be revealed how accurate this way to retract pragmatics by distributional methods actually would be. Aside from that, dialogues would be included in the judgment, including words that were returned after the corpus mining experiment to be completely irrelevant to one another. If these would get a significant lower rating than the dialogues including words that were found in the corpus to be relevant to one another, this would demonstrate that it does make sense to state that two words standing in each other's nearness have a higher chance to be in a Gricean relevance relation.

The method section is divided in 5 subsections. First of all, the corpus that serves as test material is introduced. Succedingly, it is explained how the corpus is 'mined', i.e., how the words that are supposed to stand in a Gricean relevance relation are found. This is followed a section on how the survey is composed in which participants are able to judge the word pairs that are supposed to be in a Gricean relevance relation (and those that are not).

[1] This sentence should not be interpreted as if it is argued that the situation allows for the Gricean relation to occur; rather, the relation arises because of the situation but it is only because this is acknowledged by the speakers.

4.1.1 The corpus

In order for the test data to match natural Dutch speech as close as possible, the *Corpus Gesproken Nederlands* (CGN (Version 2.0.3), (2014)) is used. The corpus, completed in 2004, was created to provide users with the most accurate sample of Dutch speech as possible. It consists of +/- 9 million words, constructing the 13 000 speech fragments. This makes the CGN the largest Dutch speech corpus . All of this makes up for the fifteen different categories the corpus is composed of (Van Eerten, 2007):

1. Spontaneous conversation (face-to-face)
2. Interviews with teachers of the Dutch language
3. Telephone conversations recorded from the telephone interoffice
4. Telephone conversations recorded from a minidisk
5. Business negotiations
6. Interviews and discussions
7. Discussions, debates, meetings (mostly political)
8. Lessons
9. Spontaneous commentaries (a.o. sports), broadcasted on radio and television
10. Current affairs sections and reports broadcasted by radio and television
11. News bulletins broadcasted by radio and television
12. Observations and commentaries broadcasted on radio and television
13. Readings, public speeches
14. Lectures, solemn speeches
15. Text read out loud

In this thesis, all of the categories are used except for the fifteenth. This choice has been made because in written text the language user can make more thoughtful choices. What's more, these choices may have been made on basis of poetic or stylistic grounds. It is expected that these choices influence the sample to that extent that they would spoil the evidential value of the research.

4.1.1.1 'Mining' the corpus

From the CGN words were selected in order that the maxim of relevance by Grice (1975) could be simulated. In order to do so, the sentence pair by Grice (1975) to illustrate his maxim of relevance was used as a reference point. This is the following:

(8) A) I am out of petrol.
B) There is a garage round the corner.

Even though these sentences could stand on their own, they still make sense when put in their respective slots for a dialogue form. This can be explained according to Gricean theory as follows:

1. Process the conventional meaning of the utterance
   For both A) and B) the meaning of the sentences can be taken literal.
2. Combine the previous with knowledge on the cooperative principle and the maxims
   It would be odd for A) to be uttered as a matter-of-fact. After all, if it would be common to do so people would utter all kind of declarative (i.e. factual) sentences all the time to each other. However, the cooperative principle being in effect always, the speaker wants to be understood and (in this situation especially) the listener would want to understand. The listener understands that A) is not uttered just as a matter-of-fact because the sentence is too simple to contain too much information, is too declarative to be false and is very understandably produced. The relevance principle is not flouted because it would make sense to state a matter-of-factly sentence if it adds information that contrasts with the status quo.
3. Take into account the (linguistic) context
   It is important for A) to be brief and clear, because it is addressed at a stranger. The dialogue contributors have no common ground and may differ in physical background such as age, gender, native language, etc. It would be very odd for A) to cause a maxim to be

flouted. This also goes for B), which connects to "petrol" because a garage is a place which is worldwidely known as a gas containing occasion. Just like A), B) is not flounting

4. Add knowledge on "other items of background knowledge"
   A) is uttered when the speaker of the latter is standing on the side of the road next to car. It is easy to conclude that A) needs help because broken things need to be fixed.
5. Acknowledge that 1-4 are available to both participants and that both participants know or assume this to be the case.
   The situation in which A) and B) are uttered is very basic. Grice does not mention any details concerning the latter. However, he does mention that both are speakers of English and that they are strangers to each other. Therefore, there is no reason to assume that either the utterer of A) or the utterer of B) have different interpretations of knowledge of 1-4. The most important would be that A) knows that a garage is a place where one can get petrol.

In the survey, the sentence pair had to be translated to Dutch because the CGN was used. This resulted in (9):

(9)	A)	Ik	heb	geen	benzine		meer.	
		I	have	no	petrol		anymore.	
	B)	Er	is	een	garage	om	de	hoek.
		There	is	a	garage	around	the	corner

After that, the object of the first sentence ("petrol" in A)) and the subject of the second sentence ("a garage" in B)) were omitted. What's more, to 'generalise' the sentence in B), "om de hoek" was deleted as well and replaced by other common propositions such as *vlakbij* ("nearby") and *daar* ("over there"). This was done so that a container not taking up a lot of space, such as *koffiepot* ("coffee pot") would come across as natural in this context. After all, following Grice, the meaning of the sentence would be derived from the intention of the speaker. As a result, the intention of A) being "requesting for (a place to find) *x*" and the intention of B) "being helpful by providing information in order to find *x*" there would be no loss of natural meaning by executing these acts. This, in its turn, resulted in 4) :

(10)	A)	Ik	heb	geen	____	meer.
		I	have	no	____	anymore.
	B)	Er	is	een	____	∅.
		There	is	a	____	∅.

Occasionally, the construction *I have* ("ik heb") was replaced by *er zijn* ("there are") in A) to create sentences as natural as possible without loss of intention as the object in A) was sometimes plural.

In order to create accurate test data, the gap in A) needed to be filled with a word referring to a quantifiable, physical entity. Therefore, the CGN was searched within by using regular expressions that were particularly made up to select words that are physical and quantifiable. All the regular expressions can be found in Appendix A, for both the gap filling words in A) and B).

The expressions for words in A) resulted in 61,254 hits. In order to obtain the largest result from the sample, the fifty most frequent words were used. After all, this way the chance that a word would co-occur with another word would be the highest, increasing the possibility that the hypothesis would be falsified.

Unfortunately, many of the results turned out to be words referring to a non-physical entity. For example, 1297 times the words *keer* ("occasion") was abstracted, making up for most of the results. Therefore, the fifty most frequent words were compared to the distinction made in nouns by *Genootschap Onze Taal* (Our Language Society), a Dutch association that curates Dutch language. This distinction entails that two types of nouns exist: those referring to concrete, and those referring to abstract nouns. Furthermore, the concrete nouns are made up from people, animals and things. The abstract nouns are made up from "feelings, period of times, characteristics, events and imaginary entities" (E-ANS, 1997; Onze Taal, 2012). As concrete nouns are per definition quantifiable (after all, they refer to physical entities) all the words that refer to other concepts than concrete as defined by Onze

Taal (2012) were removed from the list. This resulted in a new top-fifty list for gap filling words in A). These can be found in Appendix B.

Equal to the composing of the word list filling the gap in A), the fifty most frequent words were used to create a largest possibility of the hypothesis being falsified. The words this resulted in can be found in Appendix C.

With the words collected, it was necessary to check their distribution throughout the CGN. This was done by using the most recent version of CORpus EXploitation (COREX). In order to do so, descriptive research as mentioned in the handbook by Leary (2001) "describe characteristics of behaviours of a given population" (p. 104) had to be carried out. First, all the search queries had to be created and inserted manually in the program. After that, the results were saved to the hard drive. This resulted in 50 csv files that were in their turn loaded in Excel. With all the word co-occurrences collected, an average per word was calculated. This can be found in Appendix F, together with the frequency per word.

Finally, an example of a survey entry would look as follows:

A)	Ik	heb	geen	*gulden*	meer.
	I	have	no	*florin*	anymore.
B)	Er	is	een	*stad*	*vlakbij*.
	There	is	a	*city*	*nearby*.

The word *gulden* ("florin") was found 1692 times in the corpus with the regular expressions used. The word *stad* ("city") was found 823 times. Close to each other, they were found 20 times. The word *vlakbij* ("nearby") was added so that the sentence would come across as natural. An example of an entry composed of words that did not occur close to each other in the corpus is the following:

A)	Er	zijn	geen	*mensen*	meer.
	I	have	no	*people*	anymore.
B)	Er	is	een	*krant*	*daar*.
	There	is	a	*newspaper*	*over there*.

Both the words *mensen* ("people") and *krant* ("newspaper") did occur frequently in the corpus – 1352 and 474 times respectively. However, they did not occur close to each other in the corpus. The pairs were presented to participants in a survey and required being rated on a 7-point Likert scale. The points were labelled as follows (translation is written in italics and was not visible to the participants):

1 Ondenkbaar antwoord *Unthinkable answer*	2 Heel ver-gezocht ant-woord *Very far-fetched answer*	3 Denkbaar, maar vergezocht *Imaginable, but far-fetched*	4 Ongeveer vergelijkbare gangbaarheid met tankstation *About as common as gas station*	5 Vergelijk-bare gangbaar-heid met tankstation *Comparable commonness to gas station*	6 Bijna net zo gangbaar als tank-station *About as common as gas station*	7 Exact zo gangbaar als tankstation *Exactly as common as gas station*

By comparing the word pair presented in the survey to the pair *benzine-tankstation* ("petrol" - "gas station"), it is attempted to get as close to a measurement of the maxim of relevance as possible.

4.1.2 Compiling the survey

4.1.2.1 Survey length

It should be noted that if the prevalence for each word combination was tested, this would have resulted in a survey with 50 x 50 = 2500 points of judgement. Therefore, participants were not asked to rate all the words. After all, this can possibly result in monotonic responses (i.e. receiving the same response from participants over and over) as mentioned by Rolstad, Adler & Rydén (2011). According to the latter, in a subject-centered scale survey research there are several factors at play that should decide upon the length of a survey. First of all, there is the homogeneity of the respondents. More homogeneity

would require a larger sample. However, as in this research the information abstracted requires a full, native proficiency of Dutch language, not much variance is possible in the participant group.

Secondly, the researchers mention response alternatives as a factor that influences survey length. Increasing this value will also increase the amount of time it takes to complete the survey and therefore respondent's fatigue. In this research, there are no response alternatives per research question as respondents answer a question per 1 item.

Thirdly, the number of scale items plays a role. Naturally, choosing among more scale points would increase the time it takes for a participant to express their judgement. In the survey, a 7-point Likert scale was used as according to the widely cited chapter by Krosnick & Presser (2009, p. 20) this is an optimal amount of points.

4.1.2.2 Composing the sentences

In order to carry out complete research, word pairs were selected that satisfied the following criteria:

1. Occurring more than four times the average in the corpus
2. Not co-occuring close enough to each other as pairs in the corpus

The first category concerning word pairs occurring frequently[2] resulted in 64 questions (marked green in the survey, to be found in the appendix).The second category considered word pairs that did not co-occur close enough as pairs in the corpus (marked red in the survey to be found in the appendix). These are used to be able to falsify the hypothesis. The content of the survey was created while keeping guidelines set up by Lawrence Neuman (2014, p. 326) in mind.

Furthermore, points on the scale were named as according to Krosnick & Presser (2009, p. 20) this would improve the reliability of the research. They were, therefore, named accordingly to the Dutch standards. The survey as a whole is presented in Appendix G. The word pairs that occurred more than four times the average each are marked green in the survey. Words that did not occur together at all are marked in red.

4.1.3 Carry-out of survey

There were no criteria set for the research participants other than that they had to be native Dutch speakers. Participants signed a consent form, included in the survey on the first page. Participants were surveyed themselves by the thesis author to make sure that the questionnaire had their full attention. After completing the survey participants were offered a small gift. On average, participants took about 16,5 minute to complete the questionnaire.

5 Results

51 participants took part in the research. 26 of them were male, 25 of them were female. The average age was 25 (mode=25, median=25).

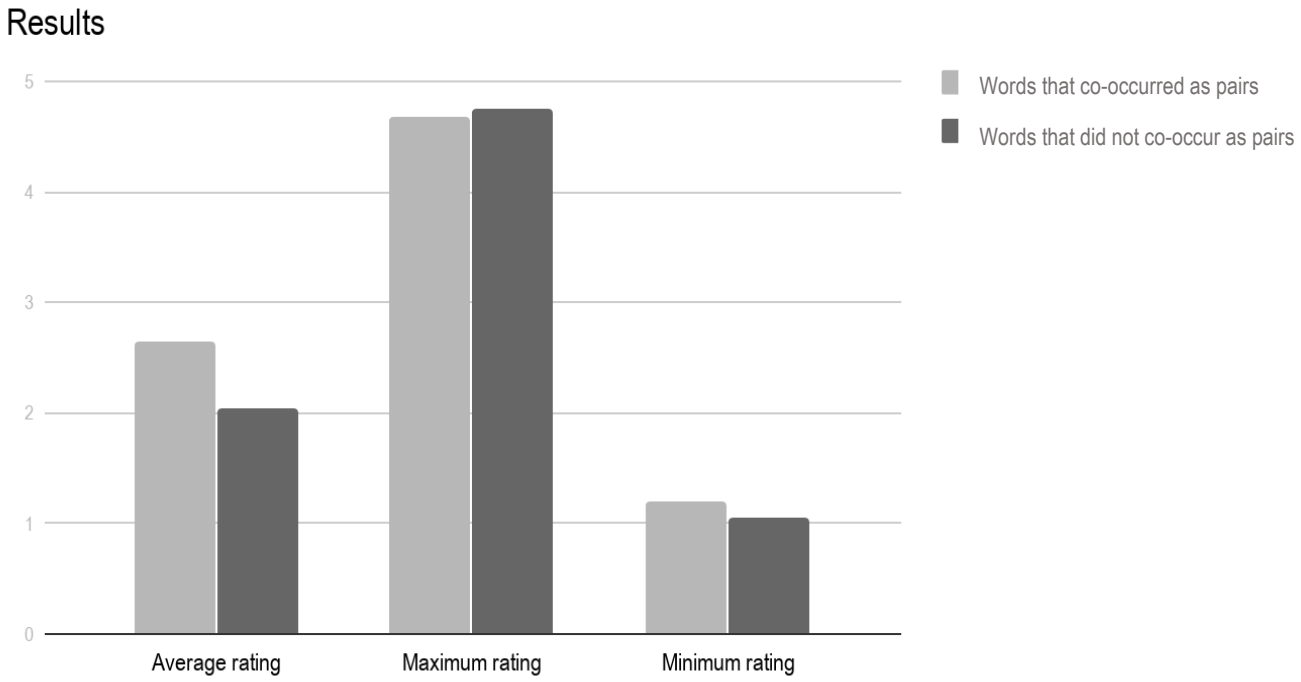
For an indication of inter-annotator agreement, and by using previous research making use of annotation for computational linguistics (Reisinger & Mooney, 2010; Silberer & Lapata, 2014; Hill, Reichart & Korhonen, 2016) the average of pairwise Spearman ρ correlations was calculated between the ratings of all respondents. Overall agreement turned out to be $\rho = 0{,}866$. This suggests that participants were able to understand the characterisation of acceptability as presented in the questionnaire and apply it to the concepts of diverse ratings consistently.

The average rating for word pairs that were found to co-occur within 20 words distance from each other was 2,658. This entails that the words falling into the first category mentioned in 4.1.2.2 were roughly speaking considered to stand in a "very far-fetched" and "imaginable, but far fetched" relation to each other compared to the perfect relevance relation that the words *benzine* ("petrol") and *tankstation* ("gas station") have as mentioned in (9).

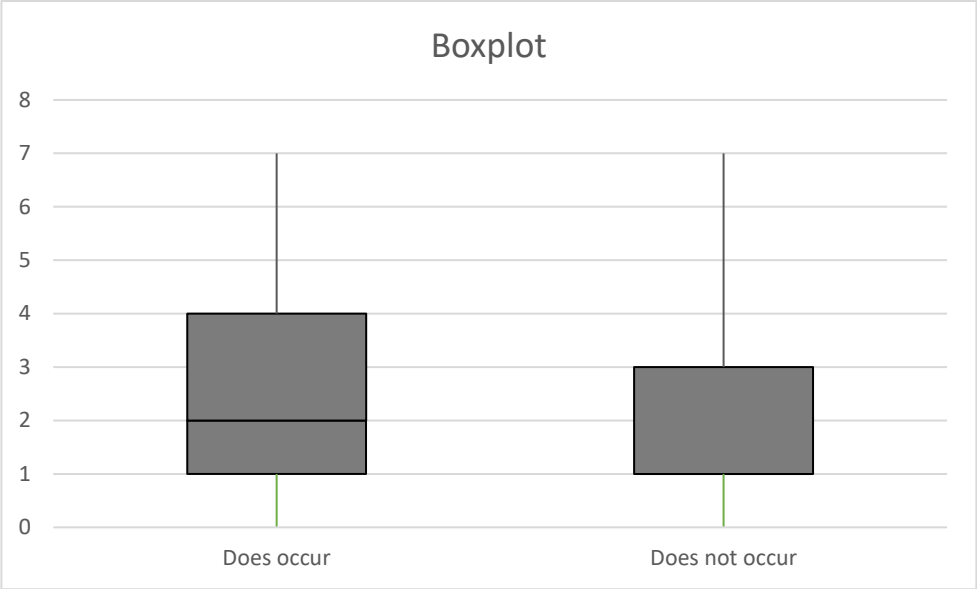
Contrastingly, the rating for word pairs that did not occur in the corpus was 2,036. This tentatively points towards a confirmation of the hypothesis. After applying a paired t-test comparing per

[2] 'Word pairs occurring frequently' refers to the words that made it into being gap filling in A) and in B). Therefore, this means that they occurred in a top fifty of results found by the regular expressions mentioned in 4.2 and that they were in a top 41 of having a word distance of 20 words of one another. Of these top fifties found for A) and B), 41 pairs were found that stood in a distance of 20 words of one another.

participant (*n*=51) the mean the judgement for both categories (following Sachs 1982: p. 309, Hinton et al. 2014, and Pandis 2015) the two sided p value is 6.04E-15. Being smaller than the common significance border of the two sided p value of 0.05, it is safe to state that the difference of judgement of sentences with words that do not co-occur as pairs in the corpus versus judgement of sentences created using words that are found to co-occur as pairs in the corpus, is statistically significant. The null hypothesis can be refuted.



However, the maximum rating of all word pairs in the survey receive a rating of 4,755. The minimum rating for word pairs that were 'objectively' in a Gricean relevance relation is 1,251, whereas the rating for word pairs that were 'objectively' not in this relation is 1,055. It can be deduced from this that the sentences composed from words that were 'mined' from the corpus and thus considered to be in a Gricean relation already did not have a great rating. As mentioned before, on average, participants judged these latter conversations between 'Very far-fetched' and 'Imaginable, but far-fetched'. It is clear after the t-test that the difference of *average* rating between conversations made up from words that were found in the corpus to be in a Gricean relevance relation and those that were not, is significant. However, it would be undesirable to argue that just because words appear close to each other in a certain context, it can be logically deduced from this that they will stand in a Gricean relevance relation. In the boxplot in which all the ratings are processed, this detail becomes even more clear:



For the words occurring in the corpus, the quarter lowest ratings for words that were not found in the corpus are that low that the score obtained in this quarter coincides with the second quartile. Both word pairs that objectively did and did not have a Gricean relevance obtained the lowest and highest rating as minimum and maximum rating respectively.

Still, for the words that did objectively stand in a Gricean relevance relation, a larger portion (notice the third quartile of both boxplots) did get a rating between 2 and 4. For words of the second category (that were not found in the corpus to be in a Gricean relation), this turned out to be 1 to 3. As can be observed in the boxplot, the first quartile for this category was 1.. In review: even though both categories did not receive a high judgement, the words that – as found by the distributional method – stand in a Gricean relation did get higher ratings.

A summary of the most valuable results are included in Appendix G.

6 Conclusion & Discussion

In review, if participants would rate frequently occurring word-pairs equally as relevant as infrequently co-occurring word pairs this would confirm the null hypothesis. However, the judgement of the two types of sentences was significantly different, so the null hypothesis can be refuted. Therefore, it is safe to state that the alternative hypothesis (words that suffice the distributional criterium are more likely to be in a Gricean relationship than words that do not suffice this criterium) needs to be adopted.

It should be noted that there are many ways in which this research can be improved. First of all, it is found in my sample that there is a positive correlation between the occurrence of two words and a Gricean relation (i.e. the closer and more frequent two words appear in each other's neighbourhood, the higher the chance they are in a Gricean relation). However, this relation would have be cross-examined in follow-up research with larger and more diverse samples. It is very advisable to regard this thesis research as a stepping stone, an introductory study towards what should be a much more extensive research in order to find out to what extent quantifiable methods can be used in order to extract a qualitative phenomenon from texts.

Secondly, the regular expressions themselves might not have been sufficient to abstract all the words that stand in a Gricean relevance relation. It is advisable to increase the amount of regular expressions so that more words are 'mined' from the corpus. This will also strengthen the conclusion it will result in; if more pairs are found the strength of their relation will also be different. Ideally, this will be recognised by the research participants so that they are able to judge more accurately. One suggestion realise this is to leave the regular expressions as they are, and to modify them so that more words are retrieved.

Thirdly, when the participants are considered, to reduce error of estimation it is advisable to increase the number of participants. In this research, economic sampling was what was moved towards instead of maximising of the sample which does not necessarily lead to spoiled results as is stated in

Leary (2001, p. 110). However, error of estimation can still be large since the sample size does certainly not match all native Dutch speakers. At the same time, since the participant group is quite homogenous, this can counterbalance the error of estimation. Furthermore, as Leary (2001, p. 116) points out, "probability samples are virtually never used in experimental research". Then again, all experimental research involving participants can always be increased by making use of a larger sample.

Fourthly, inter-annotator agreement quantification points out that participants were able to understand the categories. However, compared to experiments by Reisinger & Mooney (2010), Bruni et al. (2010) and Silberer & Lapata (2014) the inter-annotator agreement is quite low as the latter obtained an inter-annotator agreement of respectively 0.7, 0.84 and 0.76. As an idea for further research it is suggested to use a different - perhaps smaller - scale. However, this should also be approached cautiously as Bruni et al. (2014) did use a 7-point Likert scale.

Fifthly, it is assumed that the relevance judgement by participants in the survey would unmistakably refer to relevance as explained by Grice (1975). However, even though inter annotator agreement pointed out that the survey was understood by the participants, it is possible that the interpretation of 'relevance' did differ from one participant to another. What's more, it could be that it is not in line with the conceptualisation by Grice.

As a sixth and somewhat inevitable point of critique, it should be noted that the sentence pair by Grice that was used as a mole is English. In spite of the obvious one-to-one translation in Dutch, the possibility exists that the English implication is not perfectly transferrable. It would be worth repeating the experiment but with using an English corpus comparable to the CGN (such as enTenTen2012, an English corpus made up from "linguistically valuable" (Sketch Engine, 2015) texts on the internet) and English native speakers.

Finally, of course, the judgement of different types of sentence pairs differed significantly. However, the judgement of the sentence pairs that contained Gricean relevant word pairs, were on average still not assigned a score than 2,658. Generously rounded as 3, this refers to "imaginable but far-fetched". Therefore, it is highly questionable whether the sentence pairs resulted in this score, and most importantly the words responsible for being supposedly Gricean relevant to each other, do actually stand in a Gricean maxim relationship of relevance.

Disregarding these suggestions for improvements, that underline the infancy status of this research, the overall aim – connecting qualitative concept of relevance to a measure that be reached by a computer algorithm – has been reached. It has been pointed out that there is a lot of information that humans subtract from an utterance when in interaction that is most often not taken notice of by machines built for being master of the whole event of being in (spoken) interaction. This information has been described by the linguist Grice, whose specialism can be categorised as 'pragmatic language philosophy'. According to his theory, pragmatic information is used in order to determine how relevant one utterance is to another. Some computer science researchers do take notice of pragmatics when developing algorithms simulating human language; however, if the relationship in this thesis can be developed further, this could be implemented in future applications irrespective of future complexity.
In this research, a different approach has been taken by using a corpus supposed to model Dutch natural speech as closely as possible and abstracting certain words according to an example sentence presented by Grice (1975) to illustrate his relevance principle. It was demonstrated that it is not preferred to use a word space of 20 words to have one sentence be relevant to another when following the corpus method used in this thesis to simulate natural speech. However, it has been shown that it does make some difference if the words that are used are found in this distance from one another compared to when they are not within 20 words of distance to one another. Concluding, it can play a role in order to determine when one utterance makes sentence as a reply to another in interaction when modelling pragmatics. Consequently, this can contribute to taking pragmatics into account when building a conversational agent for human interaction. This, in its turn can ultimately serve to find out how natural the speech comes across. In the end, that is among the key features in linguistics interactions whether human or machine.

References

ALPAC (1966). *Languages and Machines. Computers in Translation and Linguistics. A report by the ALPAC*. Washington, USA: Printing and Publishing Office National Academy of Sciences.

Ariel, M. (1983). Linguistic marking of social prominence: The Hebrew *mi še* introducer. *Journal of Pragmatics,* 7 (4), pp. 389-409.

Ariel, M. (2012). Research paradigms in pragmatics. In Allan (Ed.), *The Cambridge Handbook of Pragmatics.* Cambridge, UK: Cambridge University Press.

Bartsch, S. & Evert, S. (2014). Towards a Firthian Notion of Collocation. *OPAL,* 2, pp. 48-61.

Bassett, Caroline (2019). The computational therapeutic: exploring Weizenbaum's ELIZA as a history of the present. *AI & Society,* 34 (4), pp. 803-812.

Bloomfield, Leonard (1933). *Language.* New York, USA: Henry Holt.

Boleda & Herbelot (2017). Formal Distributional Semantics: Introduction to the Special Issue.*Computational Linguistics,* 42 (4), pp. 619-627.

Brinton, L. J. (2008). *The Comment Clause in English*. Cambridge, UK: Cambridge University Press.

Bruni, Tran & Baroni (2014). Multimodal Distributional Semantics. *Journal of Artificial Intelligence Research,* 48, pp. 1-47.

Chang et al. (2005). Multilingual Dependency Parsing: A Pipeline Approach. In Nicolov et al. (Eds), *Recent Advances in Natural Language Processing IV*. Amsterdam, The Netherlands: John Benjamins Publishing Company.

Chapman, Siobhan (2005). *Paul Grice, Philosopher and Linguist.* New York, USA: Palgrave Macmillan.

Charles, Walter (2000). Contextual correlates of meaning. *Applied Psycholinguistics,* 21, pp. 505-524.

Chomsky, Noam (1957). *Syntactic structures.* The Hague, The Netherlands: Mouton.

Chu-Caroll, J. (1998). A statistical model for discourse act recognition in dialogue interactions. In Chu-Caroll, J. & Green, N. (Eds.), *Applying Machine Learning to Discourse Processing. Papers from the 1998 AAAI Spring Symposium.* Tech. rep. SS-98-01, pp. 12-17. Menlo ParK, USA: AAAI Press.

Corpus Gesproken Nederlands - CGN (Version 2.0.3) (2014) [Data set]. Available at the Dutch Language Institute: http://hdl.handle.net/10032/tm-a2-k6.

Dale, R. & Reiter, E. (1994). Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions. Paper presented at the Fourteenth International Conference on Computational Linguistics (Coling-92).

De Roeck, Anne (2005). The Role of Data in NLP: The Case for Dataset Profiling. In Nicolov et al. (Eds.), *Recent Advances in Natural Language Processing IV*. Amsterdam, The Netherlands: John Benjamins Publishing Company.

Eerten, L van. (2007). Over het Corpus Gesproken Nederlands. *Nederlandse taalkunde,* 12 (3), pp. 194-215.

E-ANS. Following *Algemene Nederlandse spraakkunst*. Groningen, The Netherlands: Wolters-Noordhoff. Retrieved from http://ans.ruhosting.nl/e-ans/index.html on November 29, 2019.

enTenTen: Corpus of the English Web. (2015, May 15). Retrieved from <https://www.sketchengine.eu/ententen-english-corpus/> on July 30, 2019.

Foote, Keith (2019, May 22). *A Brief History of Natural Language Processing (NLP)*. Retrieved from < https://www.dataversity.net/a-brief-history-of-natural-language-processing-nlp/# > on November 1, 2019.

Goodman, Bradley (1986). *Miscommunication and Plan Recognition*. Unpublished paper from UM86, the International Workshop on User Modeling, Maria Laach, West Germany.

Grandy, Richard E. & Warner, Richard (2017). Paul Grice. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition),URL=<https://plato.stanford.edu/archives/win2017/entries/grice/> on May 14, 2019.

Grice, Paul (1957). Meaning. *The Philosophical Review,* 66 (3), pp. 377-388.

Grice, Paul & White, Alan (1961). The Causal Theory of Perception. *Aristotelian Society,* supp. volume 35, pp. 121-152. In Swartz, R. (Ed.), *Perceiving, Sensing and Knowing,* pp. 438-72, New York, USA: Garden City.

Grice, Paul (1969). Utterer's Meaning and Intentions. *The Philosophical Review*, 68, pp. 147–77.
Grice, Paul (1975). Logic and Conversation. In Cole, P. & Morgan, J. (Eds.), *Syntax and Semantics*, 3, pp. 41-58. New York, USA: Academic Press.
Grice, Paul (1982). Meaning Revisited. In Smith, N.V. (Ed.), *Mutual Knowledge*. New York, USA: Academic Press, pp. 223–243.
Grice, Paul (1989). *Studies in the Way of Word.* Cambridge, USA: Harvard University Press.
Grim, Patrick (2011). Simulating Grice. Emergent Pragmatics in Spatialized Game Theory. In Benz et al. (Eds.), *Language, Games and Evolution. Trends in Current Research on Language and Game Theory.* Berlin, Germany: Springer Nature.
Grishman, Ralph. NLP: An Information Extraction Perspective. In Nicolov et al. (2005), *Recent Advances in Natural Language Processing IV*. Amsterdam, The Netherlands: John Benjamins Publishing Company.
Frederking, R. (1996). Grice's Maxims: "Do the Right Thing". Unpublished manuscript, Center for Machine Translation, Carnegie Mellon University, Pittsburgh (US).
Harnad, S. (1990). The symbol grounding problem. Physica D: Nonlinear Phenomena, 42 (1-3), pp. 335-346.
Harris, Z. (1954). Distributional structure. *Word,* 10 (23), pp. 146-162.
Huang, Y. (2017). Introduction: What Is Pragmatics? In Huang (Ed.), *The Oxford Handbook of Pragmatics.* Oxford, UK: Oxford University Press.
Irvine, A.D. (2019). Bertrand Russell. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), forthcoming URL = <https://plato.stanford.edu/archives/sum2019/entries/russell/>.
Jackson, P. & Moulinier, I. (2002). *Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization.* Amsterdam, The Netherlands: John Benjamins Co.
Jones, Karen S. (2001). Natural language processing: a historical review. Unpublished manuscript, Computer Laboratory, University of Cambrige, Cambridge, UK.
Jurafsky, Dan (2001). Pragmatics and Computational Linguistics. Ostman, J. & Verschuren, J. (Eds.), *Handbook of Pragmatics*. John Benjamins.
Hinrichs et al. (2005). A Data-driven Approach to Pronominal Anaphora Resolution for German. In Nicolov et al. (2005), *Recent Advances in Natural Language Processing IV*. Amsterdam, The Netherlands: John Benjamins Publishing Company.
Korta, Kepa & Perry (2015). Pragmatics. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2015 Edition), URL = <https://plato.stanford.edu/archives/win2015/entries/pragmatics/>.
Kumar, Ritesh [Microsoft Research]. (2017, September 11). *Computational Pragmatics.* Retrieved from <https://www.youtube.com/watch?v=UkBhelk_XpI&t=348s>.
Krahmer, E. J. (2010). What computational linguists can learn from psychologists (and vice versa). *Computational Linguistics*, 36(2), 285-294.
Hill, Reichart & Korhonen (2016). SimLex-999: Evaluating Semantic Models With (Genuine) Similarity Estimation. *Computational Linguistics,* 41 (4), pp. 665-695.
Hinton et al. (2014). *SPSS Explained. Second edition.*
Homayouni et al. (2005). Gene clustering by Latent Semantic Indexing of MEDLINE abstracts. *Bioinformatics,* 21 (1), pp. 104-115.
Labbe, Mark (December 12, 2019). AI trends 2020 outlook: More automation, explainability. *Techtarget,* retrieved from <https://searchenterpriseai.techtarget.com/feature/AI-trends-2020- outlook-More-automation-explainability> on December 18, 2019.
Langacker, Ronald (1998). Conceptualization, symbolization and grammar. In Tomasello, Michael (ed.), *The New Psychology of Language. Cognitive and Functional Approaches to Language Structure.* Mahwah, NJ, Lawrence Erlbaum, pp. 1-39.
Lawrence Neuman, W. (2014). *Social Research Methods: Qualitative and Quantitative Approaches.* Edinburgh, UK: Pearson.
Lamel, L. & Gauvain, J. (2008). Speech Processing for Audio Indexing. In Nordström & Ranta (Eds.), *Natural Language Processing. 6th International Conference (GoTAL 2008).* Gothenburg, Sweden: Springer.

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Rivista di Linguistica,* 20 (1), pp. 1-31.
Lehnert, Wendy (1978). A Conceptual Theory of Question Answering. *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, 1, pp. 158-164.
Lijst van eenheden van tijd (July 27th, 2018). In *Wikipedia.* Retrieved June 15, 2018, from s https://nl.wikipedia.org/wiki/Lijst_van_eenheden_van_tijd.
Jackson, P., & Moulinier, I. (2002). *Natural Language Processing for Online Applications : Text Retrieval, Extraction, and Categorization.* Amsterdam: John Benjamins Publishing Co.
Krosnick, J. & Presser, S. (2009). Question and Questionnaire Design. In Wright, J. & Marsden, P. (Eds.), *Handbook of Survey Research (2nd Edition).* San Diego, USA: Elsevier.
Leary, M. (2001). *Introduction to Behavioral Research Methods.* Needham Heights, USA: Allyn & Bacon.
M. Jaszczolt & Allan (2012). Introduction. In Allan (Ed.), *The Cambridge Handbook of Pragmatics.* Cambridge, UK: Cambridge University Press.
Meehan, James (1976). TALE-SPIN, an Interactive Program That Writes Stories. *Natural Language,* 5, pp. 91-98.
Miller, A. & Charles, W. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes,* 6, pp. 1-28.
Mitchell, J. & Lapata M. (2010). Composition in distributional models of semantics. *Cognitive science*, 34, pp. 1388-1429.
Montague, R. (1974). English as a formal language. In Thomason, R. (Ed.), Formal Philosophy. New Haven: Yale University Press.
Nagata, M. & Marimoto, T. (1994). First steps toward statistical modeling of dialogue to predict the speech act type of the next utterance. *Speech Communication,* 15, pp. 193-203.
Nerbonne, John (2005). Linguistic challenges for computationalists. In Nicolov et al. (Eds.), *Recent Advances in Natural Language Processing IV.* Amsterdam, The Netherlands: John Benjamins Publishing.
Nicolov, N. & Salvetti, F. (2005). Efficient Spam Analysis for Weblogs through URL Segmentation. In Nicolov et al. (Eds.), *Recent Advances in Natural Language Processing IV.* Amsterdam, The Netherlands: John Benjamins Publishing.
Onze Taal (September 28, 2012). Wat is een zelfstandig naamwoord? Retrieved from https://onzetaal.nl/taaladvies/zelfstandig-naamwoord on June 28, 2019.
Pandis, Nikolaos (2015). Comparison of 2 means (independent *z* test or independent *t* test). *American Journal of Orthodontics and Dentofacial Orthopedics,* 148, pp. 350-1.
Phelan, Mark (2014). Experimental Pragmatics: An Introduction for Philosophers. *Philosophy Compass,* 9 (1), pp. 66-79.
Piotrowski, Michael (2008). *Natural Language Processing for Historical Texts.* Retrieved from <ebookcentral.proquest.com> by Morgan & Clayp Publishers.
Pons et al., (2016). Natural Language Processing in Radiology: A Systematic Review. *Radiology,* 297 (2), pp. 329-343.
Prince, Ellen (1978). A Comparison of Wh-Clefts and it-Clefts in Discourse. *Language,* 54 (4).
Rapp, Reinhard (2003). Word sense discovery based on sense descriptor dissimilarity. In *Proceedings of the Ninth Machine Translation Summit*, pp. 315–322.
Reisinger & Mooney (2010). A mixture model with sharing for lexical semantics. In *Proceedings of EMNLP*, 1173–1182.
Revilla & Ochoa (2017). Ideal and Maximum Length for a Web Survey. *International Journal of Market Research*, 59 (5), pp. 557-565.
Rubenstein & Goodenough (1965). Contextual Correlates of Synonymy. Communications of the ACM, 8 (10), pp. 627-633.
Sahlgren, Magnus (2008). The distributional hypothesis. *Rivista di Linguistica,* 20 (1), pp. 33-53.
Sachs, Lothar (1982). *Applied Statistics. A Handbook of Techniques.* New York, USA: Springer Verlag.

Schebella, Callan. (December 16, 2019). How Conversational AI Will Boost The 2020s Economy. *AiThority,* retrieved from <https://www.aithority.com/guest-authors/how-conversational-ai-will- boost-the-2020s-economy/> on December 18, 2019.

Searle, John (1969). *Speech acts. An essay in the philosophy of language.* Cambridge, UK: Cambridge University Press.

Searle, John (1975). Indirect speech acts. In Cole, P. & Morgan , J.L. (Eds.), *Speech Acts: Syntax and Semantics*, 3, pp. 59-82. New York, USA: Academic Press.

Shahed Sorower et al. (2011). Inverting Grice's Maxims to Learn Rules from Natural Language Extractions. *NIPS*, pp. 1-9.

Shank et al. (1973). Margie: Memory Analysis Response Generation and Inference on English. *Natural Language Systems,* pp. 255-261.

Silberer, C. & Lapata, M. (2014). Learning grounded meaning representations with autoencoders. In Proceedings of ACL, Sofia.

Sporleder, C. & Lascarides, A. (2005). Exploiting Linguistic Cues to Classify Rhetorical Relations. In Nicolov et al. (Eds.), *Recent Advances in Natural Language Processing IV.* Amsterdam, The Netherlands: John Benjamins Publishing.

Stevenson, Charles L. (1944). *Ethics and Language.* New Haven, USA: Yale University Press.

Strawson, Peter F. (1964). Intention and convention in speech acts. *The Philosophical Review LXXIII.* Reprinted in Peter Strawson (1971): pp. 170-89.

Stolcke et al. (1998). Dialog act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics,* 26 (3), pp. 339-371.

Swade, Doron (2002). *The Difference Engine. Charles Babbage and the quest to build the first computer.* New York, USA:

Venislav, Georgiev (2018). (August 31, 2018). Chatbot Statistics 2018. *Chatbots Magazine*, retrieved from <https://chatbotsmagazine.com/chatbot-statistics-2018-68168987ab7e> on December 18, 2019.

Weaver, Warren (1949). The Mathematics of Communication. *Scientific American,* 181 (1), pp. 11-15.

Weizenbaum, Joseph (1966). ELIZA - a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9, pp. 36-45.

Weizenbaum, Joseph (1976). *Computer Power and Human Reason: From Judgment to Calculation.* New York, USA: Morgan Kaufmann Publishers.

Wentworth et al. (2012). *How to Think Like a Computer Scientist: Learning with Python 3 Documentation, Release 3rd Edition.* Free Software Foundation, Inc.

Wittgenstein, Ludwig (1958). In Anscombe, G.E.M. (Ed.), *Philosophical Investigations.* Oxford, UK: Basil Blackwell.

Zalta, E. (2019). Gottlob Frege. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition), forthcoming URL = <https://plato.stanford.edu/archives/sum2019/entries/frege/>.

Appendix A

Regular expressions used for filler in A):

1. [pos_vwtype = "pers"] [word = "heb" | word = "heeft" | word = "hebben" | word = "had" | word = "hadden"] [word = "geen"] [pos_head = "n"]
2. [pos_vwtype = "pers"] [word = "heb" | word = "heeft" | word = "hebben" | word = "had" | word = "hadden"] [word = "nog"] [pos_head = "n"]
3. [pos_vwtype = "pers"] [word = "heb" | word = "heeft" | word = "hebben" | word = "had" | word = "hadden"] [word = "meer"] [pos_head = "n"]
4. [pos_vwtype = "pers"] [word = "heb" | word = "heeft" | word = "hebben" | word = "had" | word = "hadden"] [word = "minder"] [pos_head = "n"]
5. [pos_head = "tw"] [pos_head = "n"]

Regular expression used for filler in B):

[pos_head = "vz"] [pos_head = "lid"] [pos_head = "n"]

Appendix B

Top fifty of words to be filled in in A) used in the corpus together with their frequencies

1.	gulden *florin*	1692
2.	mensen *people*	1352
3.	euro *euro*	990
4.	frank *franc*	945
5.	man *man*	931
6.	kinderen *children*	907
7.	punten *points*	891
8.	punt *point*	880
9.	dingen *things*	874
10.	ding *thing*	852
11.	mannen *men*	843
12.	personen *persons*	836
13.	boeken *books*	831
14.	klas *class*	825
15.	dollar *dollar*	819
16.	leerlingen *students*	707
17.	piek *zenith*	700
18.	jongens *boys*	687
19.	cent *cent*	672
20.	vrouwen *women*	663
21.	bladzijden *pages*	644
22.	geld *money*	639
23.	verdieping *floor*	638
24.	stuk *piece*	634
25.	Nederlanders *Dutch*	631
26.	auto's *cars*	629
27.	huizen *houses*	626
28.	boek *book*	622
29.	pagina's *pages*	619
30.	bezoekers *visitors*	605
31.	meisjes *girls*	597
32.	persoon *person*	595
33.	paal *pole*	450
34.	stap *step*	387
35.	lijn *line*	356
36.	kaartjes *tickets*	340
37.	inwoners *inhabitants*	265
38.	zetels *seats*	241
39.	kaarten *cards*	204
40.	schapen *sheep*	141
41.	slaapkamers *bedrooms*	134
42.	vrouw *woman*	127
43.	gezicht *face*	115
44.	fiets *bike*	110
45.	huis *house*	81
46.	broek *pants*	72
47.	hond *dog*	61
48.	collega's *colleagues*	58
49.	tenten *tents*	30
50.	erts *ore*	11

Appendix C

Top fifty of words filling gaps in B) together with their frequencies

1.	auto *car*	1066
2.	hoofd *head*	942
3.	buurt *neighbourhood*	902
4.	stad *city*	823
5.	kamer *room*	738
6.	huis *house*	695
7.	klas *class*	689
8.	grond *ground*	663
9.	trein *train*	562
10.	wereld *world*	518
11.	boek *book*	513
12.	ogen *eyes*	500
13.	krant *newspaper*	474
14.	weekend *weekend*	470
15.	toekomst *future*	445
16.	ziekenhuis *hospital*	441
17.	tuin *garden*	431
18.	buitenland *abroad*	422
19.	bus *bus*	412
20.	gang *hallway*	406
21.	kant *side*	401
22.	weg *road*	398
23.	telefoon *phone*	390
24.	water *water*	385
25.	kerk *church*	381
26.	computer *computer*	379
27.	bank *couch*	378
28.	plaats *place*	366
29.	groep *group*	356
30.	zomer *Summer*	344
31.	keuken *kitchen*	344
32.	veld *field*	340
33.	markt *market*	340
34.	handen *hands*	326
35.	verleden *past*	324
36.	middag *noon*	317
37.	mond *mouth*	298
38.	gemeente *municipality*	296
39.	universiteit *university*	294
40.	dorp *town*	285
41.	station *station*	280
42.	lucht *air*	278
43.	kop *head*	276
44.	tafel *table*	271
45.	winkel *store*	268

46. bedrijf *company*	266	49. voeten *feet*	260
47. midden *middle*	263	50. oog *eye*	260
48. centrum *centre*	260		

Appendix D

Co-occurrence of A) and B) fillers and their average co-occurrence

1. gulden-stad
   florin-city
   20
2. inwoners-stad
   inhabitants-city
   42
3. schapen-stad
   sheep-city
   25
4. slaapkamers-stad
   bedrooms-city
   21
5. huizen-stad
   houses-city
   11
6. dingen-kamer
   things-room
   10
7. huis-kamer
   house-room
   11
8. huizen-kamer
   houses-room
   8
9. huis-stad
   house-city
   11
10. gulden-huis
    florin-house
    10
11. dingen-huis
    things-house
    10
12. geld-huis
    money-house
    4
13. verdieping-huis
    floor-house
    4
14. mensen-klas
    people-class
    13
15. boek-boek
    book-book
    2368
16. gulden-kamer
    florin-room
    9
17. mensen-weg
    people-road
    21
18. kinderen-weg
    children-road
    11
19. dingen-weg
    things-road
    11
20. mensen-kerk
    people-church
    13
21. man-groep
    man-group
    26
22. dollar-groep
    dollar-group
    11
23. handen-huis
    hands-house
    11
24. man-ogen
    man-eyes
    16
25. dingen-ogen
    things-eyes
    9
26. bladzijden-handen
    pages-hands
    8
27. vrouwen-ogen
    women-eyes
    8
28. gezicht-ogen
    face-eyes
    5
29. mensen-wereld
    men-world
    21
30. kinderen-wereld
    children-world
    11
31. dingen-wereld
    things-world

11
32. inwoners-buurt
inhabitants-neighbourhood
42
33. schapen-buurt
sheep-neighbourhood
25
34. slaapkamers-buurt
bedrooms-neighbourhood
21
35. geld-winkel
money-shop
3
36. vrouw-winkel
woman-shop
4
37. gulden-bedrijf
florin-company
6
38. mensen-bedrijf
people-company
15
39. geld-bedrijf
money-company
10
40. mensen-midden
people-middle
10
41. klas-oog
class-eye
5

Appendix E

Survey

Phrases that were compiled with the words listed in Appendix E are highlighted green, word pairs that did not occur were marked red. This was not visible to participants.

Media Technology scriptieonderzoek

Media Technology thesis research

TOESTEMMINGSVERKLARING
Hierbij verklaart de deelnemer aan dit afstudeeronderzoek, dat ongeveer 15 min. duurt, toestemming te verlenen om de door haar of hem ingevulde gegevens op te slaan. Het onderzoek betreft verwerking van taal door een computerprogramma en mensen. De deelnemer begrijpt dat het ieder moment mogelijk is om deelname te stoppen. De verzamelde gegevens zullen vertrouwelijk worden behandeld. Bij het kiezen van "akkoord" stemt deelnemer als uniek persoon in met het bovenstaande.
CONSENT DECLARATION
Hereby the participant of this thesis research, that will take about 15 minutes, declares to give permission to store the data supplied by him or her. The research concerns the processing of language by a computer program and people. The participant understands that it is possible every moment to stop participation. The collected results will be treated confidentially. By choosing "agree" the participant agrees as a private person with the above.

- Akkoord
- *Agree*

Start van blok: Standaard vragenblok

Q15 Heel erg fijn dat je wilt meehelpen aan dit scriptieonderzoek. Vul alsjeblieft je persoonsgegevens in:
Thank you very much for assisting in this thesis research. Please fill in your personal details:

Geslacht:
Gender:
- Vrouw
- *Female*
- Man
- *Male*
- Niet definieerbaar als één van bovenstaande
  Not definable as any of the above

Q16 Leeftijd:
Age:

_____

Pagina-einde

Q17 Stel je de volgende situatie voor.Een autorijder staat langs de kant van de weg met auto. De auto staat zonder benzine. De autorijder loopt op een voorbijganger af en zegt:"Mijn auto heeft geen benzine meer."De voorbijganger antwoordt:"Er is een tankstation om de hoek."
Imagine the following situation. A car driver stands on the side of the road with the car. The car is without gas. The car driver walks towars a passenger and says: "My car does not have gas anymore." The passer-by replies: "There is a gas station around the corner."

Q18 Je zou de woorden "benzine" en "tankstation" ook door iets anders kunnen vervangen, bijvoorbeeld "brood" en "bakker". Er zijn echter nog veel meer combinaties mogelijk. Hoe vind je de volgende combinaties klinken?
You could replace the words "gas" and "gas station" also by something else, for example "bread" and "baker". There are in fact many more combinations possible. How do the following combinations sound to you?

Pagina-einde

Q14 Hoe gangbaar is het **antwoord** (beginnend met "ga naar") in de volgende gesprekken beginnend met "ik zoek" in vergelijking tot **eerdergenoemd voorbeeld** (zie onderaan pagina)?
*How common is the **response** (starting with "go to") in the following conversations starting with "I am looking for" in comparison to **previously mentioned example** (see bottom page)?*

	1 Ondenkbaar antwoord *Unthinkable answer*	2 Heel ver-gezocht ant-woord *Very far-fetched answer*	3 Denkbaar, maar vergezocht *Imaginable, but far-fetched*	4 Ongeveer vergelijkbare gangbaarheid met tankstation *About as common as gas station*	5 Vergelijk-bare gangbaar-heid met tankstation *Comparable commonness to gas station*	6 Bijna net zo gangbaar als tank-station *About as common as gas station*	7 Exact zo gangbaar als tankstation *Exactly as common as gas station*
Ik heb geen gulden meer. Antw.: Er is een stad verderop. *I don't have a florin anymore. There is a town nearby.*	○	○	○	○	○	○	○
Er zijn geen inwoners meer. Antw.: Er is een stad verderop. *There are no inhabitants anymore.*	○	○	○	○	○	○	○
Ik heb geen schapen meer. Antw.: Er is een stad verderop. *I don't have sheep anymore. There is a town nearby.*	○	○	○	○	○	○	○
Ik heb geen slaapkamers meer. Antw.: Er is een stad verderop. *I don't have bedrooms anymore. There is a town nearby.*	○	○	○	○	○	○	○
Ik heb geen huizen meer. Antw.: Er is een stad verderop. *I don't have houses anymore.* There is a town nearby.	○	○	○	○	○	○	○
Ik heb geen dingen meer. *I don't have things anymore.* Antw.: Er is een kamer verderop. *There is a room nearby.*	○	○	○	○	○	○	○
Ik heb geen huis meer. Antw.: Er is een kamer verderop. *I don't have a house anymore.*	○	○	○	○	○	○	○

There is a room nearby.
Ik heb geen huizen meer.
Antw.: Er is een kamer verderop.
I don't have houses anymore.
There is a room nearby.
Ik heb geen huis meer.
Antw.:Er is een stad verderop.
I don't have houses anymore.
There is a city nearby.

Q6

Ik heb geen benzine meer.
Er is een tankstation om de hoek.

"I don't have gas anymore."
"There is a gas station around the corner."

Einde blok: 8 ja

Start van blok: 7 ja

Q13 Hoe gangbaar is het **antwoord** (beginnend met "ga naar") in de volgende gesprekken beginnend met "ik zoek" in vergelijking tot **eerdergenoemd voorbeeld** (zie onderaan pagina)?
*How common is the **response** (starting with "go to") in the following conversations starting with "I am looking for" in comparison to **previously mentioned example** (see bottom page)?*

	1 Ondenkbaar antwoord *Unthinkable answer*	2 Heel ver-gezocht ant-woord *Very far-fetched answer*	3 Denkbaar, maar vergezocht *Imaginable, but far-fetched*	4 Ongeveer vergelijkbare gangbaarheid met tankstation *About as common as gas station*	5 Vergelijk-bare gangbaar-heid met tankstation *Comparable commonness to gas station*	6 Bijna net zo gangbaar als tank-station *About as common as gas station*	7 Exact zo gangbaar als tankstation *Exactly as common as gas station*
Ik heb geen gulden meer. Antw.: Er is een huis verderop. *I don't have a florin anymore. There is a house around the corner.*	○	○	○	○	○	○	○
Ik heb geen dingen meer. Antw.: Er is een huis verderop. *I don't have things anymore. There is a house nearby.*	○	○	○	○	○	○	○
Ik heb geen geld meer. Antw.: Er is een huis verderop. *I don't have money anymore. There is a house nearby.*	○	○	○	○	○	○	○
Er is geen verdieping meer. Antw.: Er is een huis verderop. *There is no floor anymore. There is a house nearby.*	○	○	○	○	○	○	○
Er zijn geen mensen meer. Antw.: Er is een klas verderop. *There are no people anymore. There is a class nearby.*	○	○	○	○	○	○	○
Ik heb geen boek meer. Antw.: Er ligt een boek verderop. *I don't have a book anymore. There lays a book over there.*	○	○	○	○	○	○	○
Ik heb geen gulden meer. Antw.: Er is een kamer verderop. *I don't have a florin anymore. There is a room nearby.*	○	○	○	○	○	○	○

Q59

"I don't have gas anymore."

"There is a gas station around the corner."

Q12 Hoe gangbaar is het **antwoord** (beginnend met "ga naar") in de volgende gesprekken beginnend met "ik zoek" in vergelijking tot **eerdergenoemd voorbeeld** (zie onderaan pagina)?
*How common is the **response** (starting with "go to") in the following conversations starting with "I am looking for" in comparison to **previously mentioned example** (see bottom page)?*

	1 Ondenkbaar antwoord *Unthinkable answer*	2 Heel ver-gezocht ant-woord *Very far-fetched answer*	3 Denkbaar, maar vergezocht *Imaginable, but far-fetched*	4 Ongeveer vergelijkbare gangbaarheid met tankstation *About as common as gas station*	5 Vergelijk-bare gangbaar-heid met tankstation *Comparable commonness to gas station*	6 Bijna net zo gangbaar als tank-station *About as common as gas station*	7 Exact zo gangbaar als tankstation *Exactly as common as gas station*
Er zijn geen mensen meer. Antw.: Er is een krant daar. *There are no people anymore. There is a newspaper over there.*	○	○	○	○	○	○	○
Ik heb geen ding meer. Antw.:Er is een krant daar. *I don't have a thing anymore. There is a newspaper over there.*	○	○	○	○	○	○	○
Ik heb geen personen meer. Antw.: Er is een krant daar. *I don't have persons anymore. There is a newspaper over there.*	○	○	○	○	○	○	○
Er zijn geen vrouwen meer. Antw.: Er is een krant daar. *There are no women anymore. There is a newspaper over there.*	○	○	○	○	○	○	○
Ik heb geen geld meer. Antw.: Het is zo weekend. *I don't have money anymore. It will be weekend soon.,*	○	○	○	○	○	○	○
Er zijn geen personen meer. Antw.: Het is zo weekend. *There are no persons anymore. It will be weekend soon.*	○	○	○	○	○	○	○
Er zijn geen mensen meer. Antw.: Er is een weg verderop. *There are no people anymore. There is a road over there.*	○	○	○	○	○	○	○
Er zijn geen kinderen meer.	○	○	○	○	○	○	○

Antw.: Er is een weg verderop.
There are no children anymore.
There is a road over there.
Er zijn geen dingen meer.
Er is een weg verderop.
There are no things anymore.
There is a road over there.
Er zijn geen mensen meer.
Antw.: Er is een kerk verderop.
There are no people anymore.
There is a church nearby.

Q60

Ik heb geen benzine meer.
Er is een tankstation om de hoek.

"I don't have gas anymore."
"There is a gas station around the corner."

Einde blok: Eerste 6 nee, andere 4 ja

Start van blok: Eenerlaatste 2 nee, andere 6 ja

Q11 Hoe gangbaar is het **antwoord** (beginnend met "ga naar") in de volgende gesprekken beginnend met "ik zoek" in vergelijking tot **eerdergenoemd voorbeeld** (zie onderaan pagina)?
*How common is the **response** (starting with "go to") in the following conversations starting with "I am looking for" in comparison to **previously mentioned example** (see bottom page)?*

	1 Ondenkbaar antwoord *Unthinkable answer*	2 Heel ver-gezocht ant-woord *Very far-fetched answer*	3 Denkbaar, maar vergezocht *Imaginable, but far-fetched*	4 Ongeveer vergelijkbare gangbaarheid met tankstation *About as common as gas station*	5 Vergelijk-bare gangbaar-heid met tankstation *Comparable commonness to gas station*	6 Bijna net zo gangbaar als tank-station *About as common as gas station*	7 Exact zo gangbaar als tankstation *Exactly as common as gas station*
Er is geen man meer. Antw.: Er is een groep vlakbij.	○	○	○	○	○	○	○
Er is geen dollar meer. Antw.: Er is een groep vlakbij. *There is no man anymore. There is a group nearby.*	○	○	○	○	○	○	○
Er zijn geen handen meer. Antw.: Er is een huis vlakbij. *There are no hands anymore.* There is a house nearby.	○	○	○	○	○	○	○
Ik heb geen man meer. Antw.: Er zijn ogen in de buurt. *I don't have a man anymore. There are eyes nearby.*	○	○	○	○	○	○	○
Er zijn geen dingen meer. Antw.: Er zijn ogen in de buurt. *There are no things anymore. There are eyes nearby.*	○	○	○	○	○	○	○
Ik heb geen geld meer. Antw.: Er is een veld verderop. *I don't have money anymore. There is a field nearby.*	○	○	○	○	○	○	○
Er zijn geen inwoners meer.Antw.: Er is een veld verderop. *There are no inhabitants anymore. There is a field nearby.*	○	○	○	○	○	○	○
Ik heb geen bladzijden meer. Antw.: Er zijn handen daar. *I don't have pages anymore. There are hands over there.*	○	○	○	○	○	○	○

Q61

Ik heb geen benzine meer.
Er is een tankstation om de hoek.

"I don't have gas anymore."
"There is a gas station around the corner."

Einde blok: Eenerlaatste 2 nee, andere 6 ja

Start van blok: 8 ja

Q10 Hoe gangbaar is het **antwoord** (beginnend met "ga naar") in de volgende gesprekken beginnend met "ik zoek" in vergelijking tot **eerdergenoemd voorbeeld** (zie onderaan pagina)?
*How common is the **response** (starting with "go to") in the following conversations starting with "I am looking for" in comparison to **previously mentioned example** (see bottom page)?*

	1 Ondenkbaar antwoord *Unthinkable answer*	2 Heel ver-gezocht ant-woord *Very far-fetched answer*	3 Denkbaar, maar vergezocht *Imaginable, but far-fetched*	4 Ongeveer vergelijkbare gangbaarheid met tankstation *About as common as gas station*	5 Vergelijk-bare gangbaar-heid met tankstation *Comparable commonness to gas station*	6 Bijna net zo gangbaar als tank-station *About as common as gas station*	7 Exact zo gangbaar als tankstation *Exactly as common as gas station*
Er zijn geen vrouwen meer. Antw.: Er zijn ogen in de buurt. *There are no women anymore. There are eyes nearby.*	○	○	○	○	○	○	○
Er is geen gezicht meer.Antw.: Er zijn ogen in de buurt. *There is no face anymore. There are eyes nearby.*	○	○	○	○	○	○	○
Er zijn geen mensen meer.Antw.: Er is een wereld. *There are no people anymore. There is a world.*	○	○	○	○	○	○	○
Er zijn geen kinderen meer.Antw.: Er is een wereld. *There are no children anymore. There is a world.*	○	○	○	○	○	○	○
Ik heb geen dingen meer. Antw.: Er is een wereld. *I don't have children anymore. There is a world.*	○	○	○	○	○	○	○
Er zijn geen inwoners meer.Antw.: Er is een buurt om de hoek. *There are no inhabitants anymore. There is a neighbourhood around the corner.*	○	○	○	○	○	○	○

Er zijn geen schapen meer.Antw.: Er is een buurt om de hoek.
There are no sheep anymore.
There is a neigbourhood around the corner.
Er zijn geen slaapkamers meer.Antw.: Er is een buurt om de hoek.
(15)
There are no bedrooms anymore.
There is a neighbourhood around the corner.

Q62

Ik heb geen benzine meer.
Er is een tankstation om de hoek.

"I don't have gas anymore."
"There is a gas station around the corner."

Q9 Hoe gangbaar is het **antwoord** (beginnend met "ga naar") in de volgende gesprekken beginnend met "ik zoek" in vergelijking tot **eerdergenoemd voorbeeld** (zie onderaan pagina)?
*How common is the **response** (starting with "go to") in the following conversations starting with "I am looking for" in comparison to **previously mentioned example** (see bottom page)?*

	1 Ondenkbaar antwoord *Unthinkable answer*	2 Heel ver-gezocht ant-woord *Very far-fetched answer*	3 Denkbaar, maar vergezocht *Imaginable, but far-fetched*	4 Ongeveer vergelijkbare gangbaarheid met tankstation *About as common as gas station*	5 Vergelijk-bare gangbaar-heid met tankstation *Comparable commonness to gas station*	6 Bijna net zo gangbaar als tank-station *About as common as gas station*	7 Exact zo gangbaar als tankstation *Exactly as common as gas station*
Ik heb geen frank meer. Antw.: Er is een kop daar. *I don't have a franc anymore. There is a head over there.*	○	○	○	○	○	○	○
Ik heb geen geld meer. Antw.: Er is een kop daar. *I don't have money anymore. There is a head over there.*	○	○	○	○	○	○	○
Ik heb geen geld meer. Antw.: Er is een winkel verderop. *I don't have money anymore. There is a store over there.*	○	○	○	○	○	○	○
Ik heb geen vrouw meer. Antw.: Er is een winkel verderop. *I don't have a woman anymore. There is a store around the corner.*	○	○	○	○	○	○	○
Ik heb geen gulden meer. Antw.: Er is een bedrijf verderop. *I don't have a florin anymore. There is a company nearby.*	○	○	○	○	○	○	○
Ik heb geen mensen meer. Antw.: Er is een bedrijf verderop. *I don't have people anymore. There is a company nearby.*	○	○	○	○	○	○	○
Ik heb geen geld meer. Antw.: Er is een bedrijf verderop. *I don't have money anymore. There is a company around the corner.*	○	○	○	○	○	○	○
Ik heb geen mensen meer.Antw.: Er is een midden verderop.	○	○	○	○	○	○	○

I don't have people anymore.
There is a middle around the corner.
Ik heb geen klas meer.
Antw.: Er is een oog vlakbij.
I don't have a class anymore.
There is an eye nearby.

Q63

"I don't have gas anymore."
"There is a gas station around the corner."

Einde blok: eerste 2 nee, andere 7 ja

Start van blok: 9 nee

Q8 Hoe gangbaar is het **antwoord** (beginnend met "ga naar") in de volgende gesprekken beginnend met "ik zoek" in vergelijking tot **eerdergenoemd voorbeeld** (zie onderaan pagina)?
*How common is the **response** (starting with "go to") in the following conversations starting with "I am looking for" in comparison to **previously mentioned example** (see bottom page)?*

	1 Ondenkbaar antwoord *Unthinkable answer*	2 Heel ver-gezocht ant-woord *Very far-fetched answer*	3 Denkbaar, maar vergezocht *Imaginable, but far-fetched*	4 Ongeveer vergelijkbare gangbaarheid met tankstation *About as common as gas station*	5 Vergelijk-bare gangbaar-heid met tankstation *Comparable commonness to gas station*	6 Bijna net zo gangbaar als tank-station *About as common as gas station*	7 Exact zo gangbaar als tankstation *Exactly as common as gas station*
Ik heb geen gulden meer. Antw.: Er is een toekomst. *I don't have a florin anymore. There is a future.*	○	○	○	○	○	○	○
Ik heb geen dollar meer. Antw.: Er is een toekomst. *I don't have a dollar anymore. There is a future.*	○	○	○	○	○	○	○
Er zijn geen leerlingen meer. Antw.: Er is een toekomst. *There are no students anymore. There is a future.*	○	○	○	○	○	○	○
Er is geen klas meer.Antw.: Er is een toekomst. *There is no class. There is a future.*	○	○	○	○	○	○	○
Er zijn geen auto's meer.Antw.: Er is een ziekenhuis verderop. *There are no cars anymore. There is a hospital nearby.*	○	○	○	○	○	○	○
Er zijn geen huizen meer.Antw.: Er is een ziekenhuis verderop. *There are no houses. There is a hospital nearby.*	○	○	○	○	○	○	○
Ik heb geen boek meer.Antw.: Er is een ziekenhuis verderop. *I don't have a book anymore. There is a hopsital nearby.*	○	○	○	○	○	○	○
Er zijn geen kaartjes meer. Antw.: Er is een tuin verderop. *There are no tickets anymore. There is a garden nearby.*	○	○	○	○	○	○	○
Er zijn geen inwoners meer. Antw.: Er is een tuin verderop.	○	○	○	○	○	○	○

There are no inhabitants anymore. There is a garden close by.

Q64

Ik heb geen benzine meer.
Er is een tankstation om de hoek.

"I don't have gas anymore."
"There is a gas station around the corner."

Einde blok: 9 nee

Start van blok: 9 nee

Q19 Hoe gangbaar is het **antwoord** (beginnend met "ga naar") in de volgende gesprekken beginnend met "ik zoek" in vergelijking tot **eerdergenoemd voorbeeld** (zie onderaan pagina)?
*How common is the **response** (starting with "go to") in the following conversations starting with "I am looking for" in comparison to **previously mentioned example** (see bottom page)?*

	1 Ondenkbaar antwoord *Unthinkable answer*	2 Heel ver-gezocht ant-woord *Very far-fetched answer*	3 Denkbaar, maar vergezocht *Imaginable, but far-fetched*	4 Ongeveer vergelijkbare gangbaarheid met tankstation *About as common as gas station*	5 Vergelijk-bare gangbaar-heid met tankstation *Comparable commonness to gas station*	6 Bijna net zo gangbaar als tank-station *About as common as gas station*	7 Exact zo gangbaar als tankstation *Exactly as common as gas station*
Er zijn geen zetels meer. Antw.: Er is een tuin verderop. *There are no seats anymore. There is a garden close by.*	○	○	○	○	○	○	○
Ik heb geen dollar meer.Antw.: Er is een lucht boven. *I don't have a dollar anymore. There is a sky above.*	○	○	○	○	○	○	○
Er zijn geen leerlingen meer.Antw.: Er is een lucht boven. *There are no students anymore. There is a sky above.*	○	○	○	○	○	○	○
Er zijn geen leerlingen meer.Antw.: Er is een lucht boven. *There are no students anymore. There is a sky above.*	○	○	○	○	○	○	○
Er zijn geen jongens meer.Antw.: Er is een lucht boven. *There are no boys anymore. Therer is a sky above.*	○	○	○	○	○	○	○

Er zijn geen Nederlanders meer.Antw.: Er is een mond daar. *There are no Dutch anymore.* *There is a mouth over there.*	○	○	○	○	○	○	○
Er zijn geen auto's meer.Antw.: Er is een mond daar. *There are no cars anymore.* *There is a mouth over there.*	○	○	○	○	○	○	○
Er zijn geen huizen meer.Antw.: Er is een mond daar. *There are no houses anymore.* *There is a mouth over there.*	○	○	○	○	○	○	○
Ik heb geen boek meer. Antw.: Er is een mond daar. *I don't have a book anymore.* *There is a mouth over there.*	○	○	○	○	○	○	○

Q65

Ik heb geen benzine meer.
Er is een tankstation om de hoek.

"I don't have gas anymore."
"There is a gas station around the corner."

Einde blok: 9 nee

Start van blok: 7 nee

Q21 Hoe gangbaar is het **antwoord** (beginnend met "ga naar") in de volgende gesprekken beginnend met "ik zoek" in vergelijking tot **eerdergenoemd voorbeeld** (zie onderaan pagina)?
*How common is the **response** (starting with "go to") in the following conversations starting with "I am looking for" in comparison to **previously mentioned example** (see bottom page)?*

	1 Ondenkbaar antwoord *Unthinkable answer*	2 Heel ver-gezocht ant-woord *Very far-fetched answer*	3 Denkbaar, maar vergezocht *Imaginable, but far-fetched*	4 Ongeveer vergelijkbare gangbaarheid met tankstation *About as common as gas station*	5 Vergelijk-bare gangbaar-heid met tankstation *Comparable commonness to gas station*	6 Bijna net zo gangbaar als tank-station *About as common as gas station*	7 Exact zo gangbaar als tankstation *Exactly as common as gas station*
Er zijn geen pagina's meer. Antw.: Er is een mond. *There are no pages anymore.*	○	○	○	○	○	○	○

There is a mouth.
Er zijn geen bezoekers meer.Antw.: Er is een mond daar. *There are no visitors.*
There is a mouth.
Er is geen stuk meer.Antw.: Er zijn voeten daar.
There is no piece anymore.
There are feet over there.
Er zijn geen Nederlanders meer.Antw.: Er zijn voeten daar.
There are no Dutch anymore.
There are feet over there.
Er zijn geen auto's meer.Antw.: Er zijn voeten daar.
There are no cars anymore.
There are feet over there.
Er zijn geen huizen meer.Antw.: Er zijn voeten daar.
There are no houses anymore.
There are feet over there.
Er zijn geen bezoekers meer.Antw.: Er zijn voeten daar.
There are no visitors anymore.
There are feet over there.

Q66

Ik heb geen benzine meer.
Er is een tankstation om de hoek.

Einde blok: 7 nee

Start van blok: 6 nee

Q22 Hoe gangbaar is het **antwoord** (beginnend met "ga naar") in de volgende gesprekken beginnend met "ik zoek" in vergelijking tot **eerdergenoemd voorbeeld** (zie onderaan pagina)?
*How common is the **response** (starting with "go to") in the following conversations starting with "I am looking for" in comparison to **previously mentioned example** (see bottom page)?*

1 Ondenkbaar antwoord *Unthinkable answer*	2 Heel ver-gezocht ant-woord *Very far-fetched answer*	3 Denkbaar, maar vergezocht *Imaginable, but far-fetched*	4 Ongeveer vergelijkbare gangbaarheid met tankstation *About as common as gas station*	5 Vergelijk-bare gangbaar-heid met tankstation *Comparable commonness to gas station*	6 Bijna net zo gangbaar als tank-station *About as common as gas station*	7 Exact zo gangbaar als tankstation *Exactly as common as gas station*	1 Ondenkbaar antwoord *Unthinkable answer*
Ik heb geen geld meer. Antw.: Het is zo weekend. *I don't have money anymore. It will be weekend soon.*	○	○	○	○	○	○	○
Er zijn geen jongens meer. Antw.: Er staat een trein om de hoek. *There are no boys anymore. There is a train around the corner.*	○	○	○	○	○	○	○
Er zijn geen zetels meer. Antw.: Er staat een trein om de hoek. *There are no seats anymore. There is a train around the corner.*	○	○	○	○	○	○	○
Er zijn geen auto's meer. Antw.: Er is een dorp vlakbij. *There are no cars anymore. There is a village nearby.*	○	○	○	○	○	○	○
Er is geen boek meer. Antw.: Er is een dorp vlakbij. *There is no book anymore. There is a village nearby.*	○	○	○	○	○	○	○
Er zijn geen schapen meer.Antw.: Er is een grond. *There are no sheep anymore. There is a floor.*	○	○	○	○	○	○	○

Q67

Ik heb geen benzine meer.
Er is een tankstation om de hoek.

"I don't have gas anymore."
"There is a gas station around the corner."

Appendix F

Results summarised

t-Test: Paired Two Sample for Means

	Words occurred	Words did not occur
Mean	2,653754185	2,036346246
Variance	0,644640562	0,472537122
Observations	51	51
Pearson Correlation	0,866346535	
Hypothesized Mean Difference	0	
df	50	
t Stat	10,99314775	
P(T<=t) one-tail	3,01972E-15	
t Critical one-tail	1,675905025	
P(T<=t) two-tail	6,03944E-15	
t Critical two-tail	2,008559112	

	Words occurred	Words did not occur
Mean	2,653754185	2,036346
Standard Error	0,112427804	0,096257
Median	2,658536585	2,04878
Mode	3,317073171	2,04878
Standard Deviation	0,802895113	0,687413
Sample Variance	0,644640562	0,472537
Kurtosis	-0,40632611	3,536865
Skewness	0,371130361	1,363095
Range	3,487804878	3,707317
Minimum	1,195121951	1,04878
Maximum	4,682926829	4,756098
Sum	135,3414634	103,8537
Count	51	51