A comparative study of formality in academic papers, conference papers, TED talks and popular science texts
ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, David Chan, for his splendid advice, his never-ending support and his patience. I am grateful for having had the opportunity to work with such an enthusiastic and wise professor.

Furthermore, I would like to thank my fellow students, friends and family for their moral support and advice. I could not have wished for a more supporting environment.
**Table of contents**

ACKNOWLEDGEMENTS ................................................................................................. 3

1. ABSTRACT .............................................................................................................. 7

2. INTRODUCTION .................................................................................................. 8

3. THEORY .................................................................................................................. 11

   3.1. Theory on formality ....................................................................................... 11

      3.1.1. Context ..................................................................................................... 13

      3.1.2. Deep/surface formality ........................................................................... 13

   3.2. Theory on genre ............................................................................................. 17

      3.2.1. The academic paper ................................................................................ 17

      3.2.2. The popular science article ..................................................................... 18

      3.2.3. The conference paper ............................................................................. 19

      3.2.4. The popular science speech .................................................................... 21

4. RESEARCH QUESTIONS ....................................................................................... 23

5. METHODOLOGY ................................................................................................... 25

   5.1. Corpus ............................................................................................................. 25

   5.2. Method ............................................................................................................. 29

6. RESULTS AND DISCUSSION ............................................................................. 31

   6.1. The scores ....................................................................................................... 31

   6.2. Results ............................................................................................................. 32

      6.2.1. What is the level of formal language use for each genre? ....................... 32

      6.2.2. Do the written genres score higher than the spoken genres? ................... 35

      6.2.3. Do the more popular genres score lower than the more specialised genres?........ 36

   6.3. Discussion ....................................................................................................... 37

      6.3.1. Register ..................................................................................................... 37

      6.3.2. Nominalisation ......................................................................................... 37

   6.4. Concluding remark ......................................................................................... 38

7. CONCLUSION ....................................................................................................... 39

8. BIBLIOGRAPHY .................................................................................................... 40
8.1 Primary sources ................................................................................................................. 40
  8.1.1. Popular science speech .............................................................................................. 40
  8.1.2. Popular science article .............................................................................................. 40
  8.1.3. Conference paper ..................................................................................................... 40
  8.1.4. Academic papers ...................................................................................................... 41
8.2 Secondary sources .............................................................................................................. 41
9.  APPENDICES ..................................................................................................................... 43
1. ABSTRACT
This paper researches the formality level in four different genres of spoken and written discourse: academic papers, conference papers, popular science articles and popular science speeches. For the study a corpus of twenty samples was comprised. The samples were run through a POS-tagger. As programme the “The multilingual LT3 linguistic pre-processing toolkit” by Van de Kauter et al. (2013) was used. Then the amount per relevant word classes was counted manually. The word classes that were counted were: nouns, adjectives, prepositions, articles, pronouns, verbs, adverbs and interjections. The level of formality was then calculated by applying the F-formula by Heylighen & Dewaele (2002). The results were then ranked from high to low and were compared on two bases: written versus spoken discourse and more popular genres versus more specialised genres. It can be concluded that the results are overall rather high for all of the genres. Especially the conference papers scored higher than expected. A possible explanation for the high results might be that register cannot be detected in the F-formula, resulting in similar scores per genre but a different perception by the recipient. Furthermore, the study concluded that written genres score higher than spoken genres, albeit with a minimal difference. The difference between more popular and more specialised genres is more outspoken, with the more popular genres scoring lower.
2. **INTRODUCTION**

This study aims to calculate and compare the formality level of four genres: academic papers, conference papers, popular science articles and popular science speeches. These four genres will be divided into two types of categories: written and spoken genres and popular and specialised genres.

The starting point for this study is the bachelor paper done by the present author Keppens (2014), which analyses the level of formal language use in TED talks by means of Heylighen & Dewaele’s (2002) F-formula and comparison to the baseline figures provided by Heylighen & Dewaele (2002). However, it needs to be remarked that there were several limitations to the study concerning corpus size and baseline figures.

Firstly, the corpus was too small in size to draw any firm conclusions. Furthermore, the results of the study relied on comparison with baseline figures provided by Heylighen & Dewaele. This might have had an influence on the conclusions, as the compared material was created in different settings. The material studied by Heylighen & Dewaele was provided by students who were asked to write a text in a certain genre or have a conversation consistent with a particular situation. The material studied By Keppens originated from a different setting, consisting of experts who give a speech on a certain topic aimed at a live audience of mostly non-experts.

The present study broadens the scope of that original study by investigating the level of formality of several written and spoken genres instead of focussing on one genre and comparing it to baseline figures. Additionally, the figures will be provided by applying the F-formula to corpus material selected by the researcher. The results of the research will be ranked from high to low and will be compared on two bases: written versus spoken genres and popular versus less popular genres. The first basis investigates whether or not written genres score higher than spoken genres. The second basis observes the difference in score between popular and specialised genres.

Academic papers are used as a standard for formal language use in written genres and have therefore been researched extensively. There have been numerous studies on research articles from a variety of viewpoints (Swales, 1981; Cooper, 1985; Dillon, 1991). Research articles can
be considered a prime example of writing produced in an academic community by experts for experts, thus lending itself well to this study as one of four types of language uses. More details on the academic paper can be found in the theory section on genre.

Popular science articles can be seen as the counterpart of the research paper, but aimed at a non-specialist audience. Popular science articles are articles that are published by newspapers or magazines. These articles are usually written by journalists or, at times, scientists and report on rather scientific topics. It can be noted that this particular genre has not been examined in great detail yet (Garces-Conejos & Sanchez-Macarro, 1998; Parkinson & Adendorff, 2004). More detail on this particular genre can be found in theory section on genre.

Conference papers can be generally considered to be the oral counterpart of the research article. However, unlike their written equivalent, conference papers have not yet received detailed examination. As several researchers (Holmes, 1997; Carter-Thomas & Rowley-Jolivet, 2001) confirm, more attention is given to the research article as it is written by experts to an audience of other experts. Details on the characteristics of this genre and its role in the scientific process can be found in the theory section on genre.

The last genre that will be included in this research is the popular science speech. This genre primarily consists of programmes or speeches that are given by experts and are aimed towards a non-expert audience. This genre can be seen as the spoken counterpart of the popular science article. This study will use TED talks to represent the popular science speech.

TED is a non-profit, nonpartisan organisation which started in 1984. Its aim is to spread novel and inspiring ideas to an audience of non-experts, mostly by speeches that are unlimited in length or topic (www.ted.com). Today, TED talks cover a broad spectrum of topics— from science to business to global issues. It is this increasing popularity among the public which makes TED an appealing subject of study from a linguistic point of view¹. Further information can be found on theory section on genre.

¹ Few papers address the linguistic characteristics of TED talks. The results of the study by Keppens (2014) suggest that the language is even more formal than in academic writing. The present study may confirm the results or put the results in perspective.
The four aforementioned genres (research articles, conference papers, popular science articles and TED talks) will be used to compile a corpus. The titles for the corpus are selected by year, type and topic. The topic in this study is smartphones, as this topic has been widely discussed, both in the scientific world and in daily life. Furthermore, the topic of smartphones contains language use that can both be technical as well as plain, which makes it an interesting subject from a linguistic point of view.

This paper will examine two main research questions:

- What is the level of formal language (calculated using Heylighen & Dewaele’s F-formula) use for each genre?
- How do the figures of the genres compare to each other?
  - Do the written genres score higher than the spoken genres?
  - Do the popular genres score lower than less popular versions?

This leads to the hypothesis that the research article and the conference paper will score higher on the F-formula than their popular counterparts. Furthermore, the written forms will score higher than the oral forms. The expected ranking of the scores will be:

1. Academic papers
2. Conference papers
3. Popular science articles
4. Popular science speeches

Further information concerning the hypotheses can be found in theory sections on formality and genre.

This paper will make use of POS-tagging and the F-formula created by Heylighen & Dewaele (2002) to analyse the self-compiled corpus. This formula will establish both the figures for this study as well as give an idea how the genres relate to each other in terms of score. More information on the employed method can be found in the methodology section.
3. THEORY

3.1. Theory on formality
The concept of formality can be defined in-depth or in a more clear-cut manner, depending on the situation the definition will be used in. In situations that do not require an in-depth definition, formality and informality are seen as clear-cut concepts that possess a distinctive boundary. However, in scientific texts that address the theory of formality, this definition is deepened. In such texts, formality and informality are defined as two opposite ends of a rather broad spectrum, between which there are many possibilities that include elements of both sides of the spectrum. There is, however, a distinction between definitions for spoken discourse and definitions for written discourse.

In the Longman dictionary the noun formality is clearly defined as “careful attention to polite behaviour and language in formal situations”. The adjective formal is defined as “used in official or serious situations”. While these definitions do explain the general concept, they do it in a rather simplified manner by explaining the context in which formal or informal language is used. They do not attend to the entire spectrum of (in)formality, in which many situations occur that possess elements from both formal as well as informal contexts.

The accents in the formality definitions differ in spoken and written discourse. The emphases in the definitions are influenced by factors that characterise the medium which is used to convey the message. In spoken discourse, there will be a focus on the circumstances in which formality is used and what language is appropriate to use in such circumstances. In written discourse, the definition of formality will rely more on specific word choice and what type of information is added in the text to make the text more understandable to the reader.

In academic papers concerning spoken discourse the following definitions of (in)formality can be found:

Sifianou (2013, p.3) argues that many factors contribute to the concept of formality such as familiarity, seriousness and politeness. According to her, this means that formality is rather difficult to define. She feels that formality is regularly seen as the equivalent of politeness.
Politeness describes the context in which the speaker adjusts his or her style according to the place the recipient has on the hierarchy.

Brooke et al. (2010, p. 90) see formality as “a continuous property”. The researchers state that formality is “related to the appropriateness of a word in a given context”. The appropriateness of a word in a given context can be seen as a form of politeness, which is a concept referred to by Sifianou (2013, p.3). However, while the concept of appropriateness does overlap, in part, with the definition of politeness, these two terms cannot be seen as equivalents.

Andrén et al. (2010, p.224) state that any utterance that does not follow the natural flow of conversation can be considered formal. Additionally, they seem to adopt the definition of Atkinson (1982). Atkinson defines formality as “non-conversational” and feels formal language use is rather functional.

The concept of context is of great importance when discussing formality and will be discussed further in this chapter. It should also be noted that, although there is a slight difference of emphasis, these definitions could be used for written discourse as well.

In written discourse the following definitions of the “formality” concept can be found:

Lahiri et al. (2011, p. 447) state that “As a document becomes more formal, it starts introducing more context (...) So the document usually becomes longer, with more intricate sentence structure”. This illustrates the importance of context once more.

Deforest et al. (2001) employ the historic origin of a word to distinguish formal from less formal language. The researchers use a definition by C.A Ferguson (1959) to argue that Latinate words are more common in the language use of higher social classes while Germanic words are seen as more present in the vocabulary of the lower classes. From this, she derives that Latinate words will be perceived as more formal as they are associated with the higher class, whereas Germanic words will be perceived as less formal.
3.1.1. Context
Both the definitions for written and spoken discourse have context, or appropriateness, as a common denominator. However, in spoken discourse, the focus lies on the outlying context in which formality is used. Written context focuses more what context is explicitly stated inside of the text. Hence the difference in terminology: in spoken discourse the term appropriateness is used, whereas in written discourse the term context is used.

This study makes use of two types of context. Firstly, there is the context of situation (Halliday & Hasan, 1985). This concept refers to all contextual information that is available outside of the text. These include but are not limited to: the knowledge of the recipient, the knowledge of the speaker, the function of the discourse and the power relations between the speaker and hearer. As this type of context influences the results, the results and discussion section will employ this type of context to put the results into perspective.

Secondly there is the context within the text (referred to as text as a metaphysical construct in Halliday & Hasan, 1985). This refers to all linguistic choices a speaker or writer makes to adjust to the context of situation, including but not limited to: sentence structure, register, choice in word class etc. This type of context is important, as it plays an important role in the employed method. The use in word class is one of the ways in which context within the text can be explained. The use of word class will be analysed for this study with a formula created by Heylighen & Dewaele (1999). Further information on this formula can be found in the deep/surface formality section.

These concepts of context are an important part of this study as the different genres can make use of different contexts of situation as contexts in text. The amount of context in text might influence the use of language, which in turn might influence the results of this study.

3.1.2. Deep/surface formality
The most useful definition of formality for this study is found in Heylighen & Dewaele (1999), as previously mentioned. In their study, the authors define formal language and the word classes that are typically found in formal as well as less formal language, in contrast to other authors who
define the context in which formal or less formal language is used. The definition by Heylighen & Dewaele shows what linguistic choices the writer or speaker makes when discoursing in a particular genre, which equals to one of the elements that determine the context in text, as can be read above. The definition is accompanied by a formula for measuring the level of formality in a text as well, which is an important part of the methodology of this study.

The researchers distinguish two types of formality: “Deep formality” and “Surface formality”. The first type is used for explicit and direct communication and is not related to any standard formality use in a language, such as the honorifics system used in the Japanese language. An example of deep formality is the sentence “Mrs. Smith has arrived at the house” instead of “She is here” (own examples). By using “Mrs. Smith” instead of “she” and “at the house” instead of “here” the sender removes words that can only be understood when the recipient is aware of the context.

The second type of formality, surface formality, adheres to formality use in standard language. This type of formality is used by the speaker to adjust the register to a certain context (Heylighen & Dewaele, 1999, p.3). An example of surface formality can be found in the honorifics system in the Japanese language, as mentioned above. A person who interacts with someone they view as superior addresses the person with their name and adds the suffix sama to give verbal expression to the power relations between speaker and hearer. This is a language standard in Japanese and cannot be freely chosen according to style. The only factor of influence is the setting in which the discourse takes place, to determine which suffix will be used (Okamoto, 1999).

Examples of both types of formality can be found in the corpus as well. Deep formality can be found in the structure found in conference paper CP_2, of which further information can be found in the section on the corpus material. The author of CP_2 uses “Furthermore” and “Thus” to structure the presentation so the listener is able to follow the speech. These words do not refer to an object or an idea. Instead they clarify the structure of the text by showing how two sentences or more relate to each other. An example of surface formality can be seen in the following sentence from a popular science article (PSA_4): “The justices don their robes, stroke their chins and lob their questions”. This sentence is not particularly direct or clear, since it makes use of
descriptive and figurative language. However, this type of colourful writing is typical in the context of popular science articles and thus does not confuse the reader. It is similar to the honorifics system as this writing style adheres to a certain language standard. Here the language standard is the journalistic style which is used in a popular science article. Both examples are not employed to make the text especially clear, but rather to maintain the register that is already present.

Despite this detailed definition by Heylighen & Dewaele, they state that not every utterance can be clearly labelled as one type. In most sentences and texts the level of formality can be determined provided more context outside of the text is given, however, there are exceptions. There are expressions that cannot fully be understood, even when all available context is provided. Such expressions are termed fuzzy expressions (Heylighen & Dewaele, 1999) and appear when crucial contextual information is not available or open to interpretation. Examples include concepts that can be interpreted differently according to the individual, such as the concept of hope or love.

To be fully able to determine the level of formality in an expression, a detailed context needs to be available. Heylighen & Dewaele (1999) have done research on contexts and have distinguished two categories. First, there are low-context situations. In such situations all necessary contextual information external to the text is explicitly provided in the text itself. The recipient does not need extra contextual information to understand the text. Second, there are high-context situations. In this type of situation contextual information from outside the text is not given explicitly. The speaker or writer will refer to the context implicitly or not at all because it is assumed that the recipient is fully aware of the context.

The level of formality of an expression is not only determined by context outside of the text. Heylighen & Dewaele have studied how formality manifests itself inside the text, particularly focussing on word class. They have concluded that in deep formality other word classes are dominant than in surface formality. Each has four dominant word classes.
In deeply formal expressions, the researchers found nouns, adjectives, articles and prepositions to be more prominent. These word classes mostly contain self-explanatory words that do not need any further context, since they refer directly to the object or idea of conversation. An example of deeply formal word classes can be found in a sentence found in CP_4: “Traditionally, data on social networks have been difficult to collect”. The noun phrase “social networks” and the adjective “difficult” can both be understood without further information since these word classes communicate the object or idea directly. This illustrates the conclusion made by Heylighen & Dewaele (2002), which states that nouns and adjectives can be considered deeply formal word classes.

The researchers have also concluded that pronouns, adverbs, verbs and interjections tend to appear more often in expressions that use surface formality. The reason for this is because these classes are more likely to contain words that refer to an object or situation indirectly and can therefore not be understood without contextual information. An example that demonstrates the use of deictic word classes in expressions that use surface formality is in CP_4. In this conference paper the sentence “With these data we can define the ‘adoption network’” can be found. Without more information it is impossible to know what the pronoun “we” refers to, which illustrates the contextual nature of the pronoun as a word class.

In conclusion it can be noted that the level of formality and the amount of context in text are not correlated. The two characteristics can be present at the same time in a text; however, a text might possess only one of the two characteristics as well. A text can possess a high level of surface formality and still possess much context in text, for example a journal article where a new technology is explained. There is a high level of surface formality, as the journalistic style is not meant to make the text clearer, only to adhere to the language standard of a journal article. It has much context in text because the article explains all context surrounding the new technology, so the reader may be able to follow. Similarly, a text can be deeply formal, yet have little context in text. For example, a conference paper on a certain scientific topic. The text is deeply formal, as the process needs to be clearly explained. However, there might be little context in text, as the conference paper reflects the given speech, which is held before a live audience, who are aware of the context.
3.2. **Theory on genre**

This part of the paper describes in detail the different genres that are used in the corpus for the study and detail the reasons for their hypothesised scores. This section also explains why theses genres are relevant for this research. The genres are divided into a written and a spoken category. From the chapter on formality, it can be hypothesised that the written category will score higher on the test than the spoken category, as the written category will have more context in text. First the written category will be discussed. The spoken category will be discussed afterwards.

3.2.1. **The academic paper**

An academic paper, also known as a research article, is a scientific paper written by experts for an audience of other experts. These papers are mostly published in scientific journals. This type of paper conveys new theories, ideas or experiments to an audience of experts and constitutes an important step in the scientific process.

Hyland (2006) claims:

> The fact that academic writers do not simply produce texts that discuss social or natural realities but use language to acknowledge, construct and negotiate social relations means that effectively controlling interpersonal features becomes central to building a convincing argument and creating an effective text.

There has been ample research on the language used in academic papers (Biber, 1991; Carter-Thomas & Rowley-Jolivet, 2003; Holmes, 1997). The language is considered by the entire academic community to be highly formal, to the point of academic writing being generally considered the most formal form of communication. The reason for this assumption is the context in which the paper is published: a scientific context where every theory and experiment needs to be explained completely without ambiguity. This results in more use of non-deictic words, such as nouns and articles to maximize the available context in text, as can be read in the previous chapter on formality. This supports the hypothesis that academic papers will score the highest on the test.

In Heyligen & Dewaele (2002)'s F-formula test, informational writing was ranked highest, indicating a very formal use in language. However, it is not entirely clear what kinds of specific
texts constitute the category informational writing, as a definition of the term is not provided on that study. This uncertainty limits the study made by Keppens (2014) and will be addressed in this thesis.

3.2.2. The popular science article
A popular science article is a written text that reports on scientific news to a target audience of non-experts. Such articles typically appear in widely spread newspapers and magazines. The journalist primarily bases his information on published academic papers.

However, several studies have found that several aspects of the popular science article do not match with a research article.

Firstly, there is a difference in language use. Parkinson & Adendorff (2004) have found that the language used in popular science articles is very unlike the language used in a research article. The researchers argue that research articles avoid addressing the reader or themselves, which influences the writing style. Parkinson & Adendorff (2004, p.388) claim that “(...) academic texts focus on theories (research articles and textbooks) and methods (research articles) while popular articles focus on people and what they say and think.” The focus of the popular science article on the researchers and their opinions will reflect in the use of formality in the text. More pronouns will be used as they might refer to the person speaking. This might lower the score.

Furthermore, Parkinson & Adendorff argue that the difference in which objectivity is established differently in research articles and popular science articles influences the language use in another way as well. According to the researchers, this means that popular texts will have less nominalisation and passivisation as a result of their focus on people. This results in a higher concentration of verbs, as nouns are verbalised. Additionally, the more popular text might make use of a different register, such as humour, to engage with the readers.

A second aspect that differentiates popular science articles and research articles is the different structure. Some researchers consider popular science articles a simplification of the scientific paper. Dimopoulos et al. (2003, p. 189) mention that articles in popular science magazines re-
contextualize the information from the academic paper and organise the information to match with the organisation principles. This means the new and old information is placed differently so the reader can follow certain logic in the text.

Researchers call the process from research article to a popular science article the popularization of science. There have been numerous theories as to whether or not this popularization equals simplification of the scientific facts and consequently can miss some nuances. Hilgartner (1990, p.528) states that “The boundary between real science and popularized science can be drawn at various points depending on what criteria one adopts, and these ambiguities leave some flexibility about what to label ‘popularization’”.

3.2.3. The conference paper
The conference paper is the oral presentation of academic research before an audience of experts and can therefore be considered as the spoken form of the academic paper. However it has not been studied as thoroughly as the research article.

Shalom (1993) defines a conference paper as “an oral presentation of recent or ongoing work, generally carried out by a team of researchers, but presented by one member of the team” (quoted in Rowley-Jolivet, 2002, p. 20). Normally, a conference paper is first written down then delivered orally and can in some cases be included as an article in a journal or conference proceedings publication.

One of the conclusions Rowley-Jolivet & Carter-Thomas (2005) have drawn, is that while the setting of a conference paper is slightly more informal, it uses more deictic word classes such as personal pronouns and adverbs of place. However, the language can still be considered formal. The researchers argue that conference presentations are dissimilar to everyday conversations and show a better likeness to institutionalised genres of academic speech.

Furthermore, Rowley-Jolivet & Carter-Thomas (2005, p.49-50) have pointed out a few significant differences between written and presented conference papers.
Firstly, they state that conference presentations focus on the novelties of their research. This will influence the manner of presenting a certain topic, requiring a different structure.

Secondly, conference papers are presented before a live audience, which influences the speech style of the expert, since the speaker needs to be certain the fellow experts are able to follow. The lecturer may also need to change their style as a reaction to the audience.

Additionally, presentations happen in real time. This means the expert is required to alter the information distribution. Conference papers as a rule are filled with information and this needs to be adjusted so the audience does not become overwhelmed.

Lastly, conference presentations are commonly accompanied by visual aids. This allows the speaker to show rather than explain some points in the presentation.

Rowley-Jolivet & Carter-Thomas (2005) believe that the conference paper is rather difficult for new researchers to engage with because of a significant scarcity of material for deconstruction and genre-based ways of teaching the characteristics. The researchers believe this is mainly due to the lack of interest in this topic. This insinuates that researchers may not be aware of the characteristics of a conference paper in terms of formality and writing style and may therefore produce conference papers that do not uphold the language standards of a conference paper. This has implications for the use of formality and may affect the results of the study for this genre.

They continue by claiming that conference papers are mostly accepted by the conference committee not by looking at “the appropriacy of the oral presentation itself to the contextual constraints of the genre or on the communicative proficiency of the speaker” (Rowley-Jolivet & Carter-Thomas, 2005, p.47), but merely by looking at the content of the presentation. This may implicate inconsistencies in the use of formal language use. The researchers feel this is a matter that needs to be looked into.
3.2.4. The popular science speech

A popular science speech is the oral delivery of scientific news given by experts and aimed at an audience of mostly non-experts. It can be seen as the oral equivalent of the popular science article as it shares some of its characteristics:

- Both genres are aimed at an audience of non-experts and are required to adapt the language to the recipients.
- Both genres are created by experts or reflect the findings of an expert.
- Both genres want to communicate scientific news specifically.

Since popular science speeches are aimed at non-experts (Parkinson & Adendorff, 2004) it can be argued that the language use can be theorized as being less formal than a conference paper. In the speech more deictic word classes, such as adverbs of place and time, verbs and personal pronouns, may be used to make the presentation more accessible to the audience of non-experts.

The format of a popular science speech varies. It ranges from television networks which want to distribute scientific news to an audience of non-experts to school presentations aimed at students. This study, however, will use the format of a TED talk.

TED, as mentioned in the introduction, is a non-profit organization. It mainly wants to provide a platform for speakers to present an invention, idea or insight to a broad audience.

TED started out in 1984 as a conference where Technology, Entertainment and Design (which form the initials TED) were discussed and presented. Today, TED talks cover a broad scale of topics— from science to business to global issues. The organization now hosts talks and events in more than 100 languages and hosts a website where most talks can be viewed online. There are over 1800 talks available in the archive, with online video views reaching one billion in 2012.

A TED talk is a speech or presentation that is given by an expert before an audience of mostly non-experts. This has implications for the use in language. The speaker needs to use a different register to assure the listeners are able to understand the vocabulary in and the structure of the presentation.
Not many papers have addressed the linguistic characteristics of TED talks. The results of the research done by present author Keppens (2014) suggest that the language used in TED talks is even more formal than in academic writing. She suggests this might be due to the nature of a TED talk. This type of presentation is given by an expert, who might use vocabulary he or she deems “self-explanatory”, but are not perceived as such by the audience. The results of this paper may further confirm or alter the perception of these results by comparing the register in TED talks to three other genres. The hypotheses for this thesis differ from the conclusions made by the previous study, as that particular study has several limitations which may have had an influence on the perception of the results. For one, the baseline figures from Heylighen & Dewaele were retrieved in different circumstances than the study. Secondly, there are no clear definitions of the categories for which the figures by Heylighen & Dewaele were calculated, as was previously mentioned. This might have influenced the comparability of the baseline figures with the figures Keppens calculated in her study. Taking these limitations into account, the hypotheses formulated in this study are preferred as they are based on consulted literature and a corpus that covers the full range of genres analysed.
### 4. RESEARCH QUESTIONS

The goal of this study is to compare four genres of text in terms of formal language use. The main research question for this study is:

What is the level of formal language (calculated using Heylighen & Dewaele’s F-formula) use for each genre?

Based on existing studies on formality in written and spoken genres, the following hypotheses were formulated:

- Academic papers will score the highest, as this genre is considered to use deep formality the most by all the top literature.
- Popular science articles and popular science speeches will be the lowest scoring genres respectively.

The logic behind this hypothesis is the following: conference papers are still highly scientific in nature, which means the study needs to be explained in a specific way to ensure other scientists are able to recreate the study if necessary. However, since conference papers are presented orally, the language standards are different, since the setting allows for more contextual language without deterring from the understanding.

While the published studies support these hypotheses, the study from Keppens (2014) does suggest that the popular science speech might score higher than the score calculated for informational writing by Heylighen & Dewaele (2002). However, in the latter, the definition of what constitutes informational writing is not entirely clear, which might mean that informational writing does not equal academic papers. Additionally, the material for the informational writing was retrieved in a different setting than the TED talks, which might have influenced the results of the study.

The second and third research question compares the figures of the genres on two different levels:
• Do the written genres score higher than the spoken genres?
• Do the popular genres score lower than more specialised genres?

These two research questions are of interest to this study since it further shows how context in text, as mentioned in the theory section on formality, influences the word choices of the speaker or writer. It might therefore be interesting to observe any differences in choice of language according to the setting (specialist audience versus non-specialist audience) and the medium (written discourse versus spoken discourse).

Based on the scientific literature surrounding this topic, the following hypotheses were formulated:

• The written genres will score higher than the spoken genres.
• The more specialised genres will score higher on the test than the more popular genres

The rationale behind these hypotheses is that the written genres need to explain more of the context in text. This influences the choice of word classes, which is being studied in this research and might result in a higher score. In the spoken genres, the audience is present as the speech is being delivered, which influences the amount of context that needs to be explicitly mentioned in text. This will influence the choice of word classes as well and will consequently result in a lower score.

The results of this study could confirm or refute the results of an earlier study made by Keppens (2014). Although previous results suggest that popular science speeches score the highest on the deep formality test, the corpus size was too small to generate results that can be generalised. Furthermore, this study had its limitations, as was mentioned previously. This study wants to widen the scope of the study so that the results can be considered general.
5. METHODOLOGY

5.1. Corpus
The corpus for this study consists of twenty manually selected pieces. Per genre five examples have been selected to assure an accurate representation of the language use in the genre.

The broad topic common to all of the selected corpus material is smartphones. This subject was chosen because of its popularity in everyday life, which constitutes a need for re-contextualisation of scientific news surrounding the subject. Furthermore, the scientific field of smartphones is developing at a solid pace, which means a rather constant stream of scientific developments in this area. This constant stream assures ample material on this subject.

The selection criteria for the corpus were divided into criteria that applied to all genres and criteria per genre. For all genres, the criteria were the subject and language use. There needed to be terms that could be phrased either scientifically or in a more casual manner. This way, it could be examined what choice the sender of the message had made.

Per genre the criteria were the following:
For the selection of the academic papers the main criteria were the length and the year of publication. The papers were to have a minimum length of 6000 words to ensure a detailed description of the study, a criterion that is shared by most academic papers. Furthermore, the papers were selected by year, to ensure a selection of contemporary studies, which document the recent developments in the area of cell phones. This ensures a good comparison can be made between the language use in the more scientific genres and the more popular genres. The more popular genres will select subjects that are relevant to the recipients at that moment and these subjects need to be linguistically compared with subjects that are equally as relevant to the public at that time. The papers were mainly found via Google Scholar and Science Direct. The articles were found using this list of search items:

- Smartphone
- Cell phone
- Smartphone user
- Smartphone use
The selection of the popular science articles for this corpus is based on the amount of readers the magazine has and on the focus on topics\(^2\). The selected magazines Time and The Economist both have obtained a steady, vast reader base, from which it can be derived that these magazines are highly popular. Furthermore, the magazines are known for being high quality magazines, which ensures a certain standard for language. These magazines consequently both have experience in popularising formal topics as well as adjusting the register to the reader. They can as a result be seen as reliable sources for popular science articles.

The selection of conference papers, aside from topic, was based on length of the document, to ensure a difference in information selection between academic papers and conference papers. As a conference paper is not as lengthy as an academic paper, the maker of the conference paper needs to restructure the information to guarantee the paper can be followed still. The selected conference papers had a limit of 3500 words in length. Additionally, the papers were selected by year, including each year from 2010 to 2015.

The criteria for the selection of TED talks were the year in which the speech was given and the category in which the speeches were found. The transcriptions of the TED talks were found on the TED website. For this study, speeches were selected from the search categories technology and smartphones on the TED website. The topic, as previously mentioned, was narrowed down to smartphones in particular. The selection of the talks would ideally have talks per year from 2010 to 2015, however, this could not be achieved to the full extend. There are, however, talks from 2007 and 2009, which might give an equally interesting view of potential evolutions in formality use over the course of five years.

All selected works have been coded to facilitate references in the rest of the study. Below a list of all material, their code and the year of publishing can be found.

\(^2\) There was an attempt to select the articles by year; however, due to limitations in viewing rights for the free articles it was impossible to achieve an ideal division of the articles per year.
### Popular science speech

<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eric Topol</td>
<td>The wireless future of medicine</td>
<td>2009</td>
</tr>
<tr>
<td>Fabian Hemmert</td>
<td>The shape-shifting future of the mobile phone</td>
<td>2010</td>
</tr>
<tr>
<td>Jan Chipchase</td>
<td>The anthropology of mobile phones</td>
<td>2007</td>
</tr>
<tr>
<td>Jorge Soto</td>
<td>The future of early cancer detection?</td>
<td>2014</td>
</tr>
<tr>
<td>Malte Spitz</td>
<td>Your phone company is watching</td>
<td>2012</td>
</tr>
</tbody>
</table>

### Popular science articles

<table>
<thead>
<tr>
<th>Source</th>
<th>Title</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIME Magazine</td>
<td>Why Your Smartphone Will Be Your Next PC</td>
<td>2013</td>
</tr>
<tr>
<td>Study: Fewer than 50% of Smartphone Users Make Calls</td>
<td>PSA_2</td>
<td>2011</td>
</tr>
<tr>
<td>Peak Battery: Why Smartphone Battery Life Still Stinks, and Will for Years</td>
<td>PSA_3</td>
<td>2013</td>
</tr>
<tr>
<td>The Economist</td>
<td>There’s no app for that</td>
<td>2014</td>
</tr>
<tr>
<td>How speech-recognition software got so good</td>
<td>PSA_5</td>
<td>2014</td>
</tr>
</tbody>
</table>

### Conference Papers

<table>
<thead>
<tr>
<th>Authors</th>
<th>Title</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kramer, K. M. et al.</td>
<td>Smartphone Based Face Recognition Tool for the Blind</td>
<td>2010</td>
</tr>
<tr>
<td>of Dedicated Devices</td>
<td>Yamada, M. - &quot;Talking is the new typing&quot;: Challenging smartphone users to dictate instead of type in order to enrich the mobile learning experience</td>
<td>CP_3</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Link, M. W. et al. - Not So Fun? The Challenges of Applying Gamification to Smartphone Measurement</td>
<td></td>
<td>CP_5</td>
</tr>
</tbody>
</table>

**Academic papers**

| Lee, H. et al. - A new dimension of the digital divide: Exploring the relationship between broadband connection, smartphone use and communication competence | AP_1 | 2015 |
| Mylonas, A. et al. - Delegate the smartphone user? Security awareness in smartphone platforms                                                                                                                   | AP_2 | 2013 |
| Wu, X. et al. - Analysis of smartphone user mobility traces for opportunistic data collection in wireless sensor networks                                                                                     | AP_3 | 2013 |
| Verkasalo, H. et al. - Analysis of users and non-users of smartphone applications                                                                                                                           | AP_4 | 2010 |
| Pitt, L.F. et al - Integrating the smartphone into a sound environmental information systems strategy: Principles, practices and a research agenda                                                               | AP_5 | 2011 |

The full transcriptions of the material can be found in the appendices.
5.2. Method
This section will explain the method that was used for this study.

Firstly, the transcription of all the selected fragments was processed by a POS-tagger. POS-tagging means the identification and tagging of the word class of each word in a text. POS-tagging is usually processed with the help of a computer programme which uses an algorithm to calculate what word class is most likely to be in a certain position in a certain sentence. There are several programmes available for the process, all of which visually label the word classes for the user's convenience.

For this study the POS-Tagger “The multilingual LT3 linguistic pre-processing toolkit” by the Van de Kauter et al. (2013) was used.

Note: the POS-tagger automatically generates its results, so a margin of error needs to be taken into account. The accuracy of the POS-tagger ranges from 93.38% to 97.57% (Van de Kauter et al., 2013).

For this research, all twenty fragments processed by the POS-tagger. The programme analysed the texts and labelled and coloured the words according to word class. The results of these fragments were transferred into a Microsoft excel document. Then, the number of words belonging to each word class was counted manually by filtering the document by word class and looking at the number of results.

The second stage was to select the results for the word classes which are relevant for the Heylighen & Dewaele formula and can also be found in the theory section on formality. Those word classes are: nouns, adjectives, prepositions, articles, pronouns, verbs, adverbs and interjections. The nouns, adjectives, prepositions and articles are more present in deep formal texts. Pronouns, verbs, adverbs and interjections are more frequent in texts that make use of surface formality. Word classes that are not relevant to this research were not analysed.

In the third stage, these results were used in the F-formula to calculate the final outcome.
As mentioned in the theory section on formality, this paper made use of the F-formula, created by Heylighen & Dewaele (2002).

Heylighen & Dewaele (2002) have found that these aforementioned word classes can be utilised as a way to measure formality. To calculate the formality level of a text, the frequencies of words that indicate deep formality are added. The frequency will be calculated by dividing the quantity of a certain word class by the total number of words in the text. This then will be multiplied by 100, as to get a percentage. Then all words that indicate surface formality are subtracted, also using frequency percentages. 100 is added to this result and then the outcome is divided by two. The outcome is the score of the transcript.

The “F-formula” (cf Heylighen & Dewaele, 2002) can be found below:

\[ F = \frac{\text{noun frequency} + \text{adjective freq.} + \text{preposition freq.} + \text{article freq.} - \text{pronoun freq.} - \text{verb freq.} - \text{adverb freq.} - \text{interjection freq.} + 100}{2} \]

The resulting score lies between 0 and 100. If the end result is at the higher end of the scale, it means it is a deep formal fragment. If the result leans towards the lower end of the scale, the fragment is more contextual. Contextual fragments use more surface formality, which means that the formal language is used to conform to a certain language standard and not to make the explanation more clear for the recipient.

The results were ranked from highest to lowest score and compared on two axes: written versus spoken versions and the more academic versions versus the more popular versions.
6. RESULTS AND DISCUSSION

6.1. The scores
In this section the results of the study will be calculated. As mentioned in the methodology section, the score determines the level of formality. Results leaning towards 100 will be considered deeply formal. Texts using deep formality aim to communicate all available context in the text itself.

If the result of the test is rather low the result will be interpreted as contextual, which means that the text uses mostly surface formality. Surface formality is employed to comply with a formality standard in a language. This type of formality does not communicate all available context in the text. It is assumed that the recipient is already fully aware of all the contextual elements outside of the text.

Per genre a general average score and an average score per word class will be calculated. The scores per sample can be found in the appendices.

**Academic papers**

<table>
<thead>
<tr>
<th>Score on test</th>
<th>Deep formal</th>
<th>Surface formal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nouns</td>
<td>32.32</td>
<td>12.87</td>
</tr>
<tr>
<td>Adjectives</td>
<td>9.62</td>
<td>8.05</td>
</tr>
<tr>
<td>Prepositions</td>
<td>8.05</td>
<td>2.68</td>
</tr>
<tr>
<td>Articles</td>
<td>15.44</td>
<td>4.69</td>
</tr>
<tr>
<td>Pronouns</td>
<td>4.69</td>
<td>0</td>
</tr>
</tbody>
</table>

**Conference papers**

<table>
<thead>
<tr>
<th>Score on test</th>
<th>Deep formal</th>
<th>Surface formal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nouns</td>
<td>23.28</td>
<td>9.25</td>
</tr>
<tr>
<td>Adjectives</td>
<td>8.89</td>
<td>3.12</td>
</tr>
<tr>
<td>Prepositions</td>
<td>12.75</td>
<td>16.78</td>
</tr>
<tr>
<td>Articles</td>
<td>9.25</td>
<td>4.69</td>
</tr>
<tr>
<td>Pronouns</td>
<td>4.69</td>
<td>0</td>
</tr>
</tbody>
</table>
### Popular science article

<table>
<thead>
<tr>
<th>Score on test</th>
<th>Deep formal</th>
<th>Surface formal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nouns</td>
<td>adjectives</td>
<td>prepositions</td>
</tr>
<tr>
<td>62.80</td>
<td>28.52</td>
<td>12.37</td>
</tr>
</tbody>
</table>

### The popular science speech

<table>
<thead>
<tr>
<th>Score on test</th>
<th>Deep formal</th>
<th>Surface formal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nouns</td>
<td>adjectives</td>
<td>prepositions</td>
</tr>
<tr>
<td>52.02</td>
<td>20.81</td>
<td>10.97</td>
</tr>
</tbody>
</table>

### 6.2. Results

This chapter will discuss the outcome of this research.

#### 6.2.1. What is the level of formal language use for each genre?

The average score per genre can be found in the section on the scores above:

A visual representation of these results can be found in the chart below.

![Level of formality](chart.png)

**Chart 1: general scores per genre**
The results show that there is little variation in the levels of formality per genre. Most genres possess a relatively high F-score, indicating deeply formal texts. The high scores are explained by the large amount of nouns that was found in all genres, as well as the large amount of prepositions found. A possible reason for the remarkably high level of prepositions might be that prepositions help to nominalise a sentence. This will be further clarified in the discussion section.

These results indicate that all of these genres want to communicate all available context of situation in the text.

Academic papers score the highest, as was mentioned in the hypothesis. It should be noted, however, that the score is higher than Heylighen & Dewaele’s (2002, p.316) average score of 66 for scientific texts (cf Heylighen & Dewaele 2002, 311), which can be viewed as a standard for deeply formal texts. This might be due to Heylighen & Dewaele’s chosen samples, which are different from the samples selected for this corpus. Additionally, Heylighen & Dewaele’s corpus contains more material than the corpus of this study, which may influence the average results obtained from the test.

Conference papers score remarkably high on the test, with an average score of 69.29. This suggests that conference papers resemble academic papers more than they resemble popular science speeches. This is a remarkable result, as it was hypothesised that the conference papers would score significantly lower than an academic paper due to the live setting, which would suggest a different approach in explicitly stating some of the context in text and would consequently significantly influence the choice of word classes.

What does differ from the academic paper is the decreased amount of nouns (academic papers: 32.32, conference papers: 23.28) and the increased amount of verbs (academic papers: 15.44, conference papers: 16.78). This does correspond with a text that employs more surface formality; however, the difference between the scores of the conference papers and the academic papers is minimal. An explanation for the scores for nouns and verbs might lie in the purpose of a conference paper. Initially, a conference paper gives an overview of the speech that has been delivered on a conference, which implies a writing style that reflects a speech, thus possessing
elements of a less formal style. The difference is minimal since it is still a written text and will thus possess elements of a deeply formally written text as well.

One of the most remarkable results is CP_4, which has the second highest score overall, 72.25. This score and the score of all individual samples can be found in the appendices. This particular score is due to a higher amount of prepositions (with a score of 15.95) and a lower amount of verbs (with a score of 13.50). This is illustrated in the following example sentence: “Since random adoption gives a flat $P_x(k)$, the positive slope of the results in Figure 3 are again taken as evidence for social effects (of some kind) in adoption— for both products.” Several things can be observed from this example sentence:

- The higher amount of prepositions and the lower amount of verbs is confirmed, resulting in a rather nominalised sentence. Further details on nominalisation can be found in the discussion.
- The presence of scientific references and formulas such as “$P_x(k)$” does not show how the researcher would refer to this with words, indicates this sentence does not resemble a speech transcription. As mentioned above, an actual conference speech transcription would perhaps have altered the score.

Popular science articles score higher than expected as well. The average score is 62.80, which suggests that popular science articles are deeply formal texts. There are a few plausible explanations for this result:

- The subject of smartphones might include terminology that cannot be replaced by more contextual words. Examples of this include nouns such as “speech-recognition software” (PSA_1) and noun phrases such as “graphite-based anodes” (PSA_3).
- Because of the similar verb to noun ratio, the articles score as high as other categories. A possible difference in perception of formality use might be in the register used in the articles. The scores from the F-formula do not change with any alteration of register as the

---

3 If transcriptions of audio files had been available for research, a different result might have been observed.
F-formula does not take register into account.

Another explanation for the difference in perception and score might be the journalistic style in which journal articles are written. The writing style of a journalistic article needs to be completely clear, to allow the reader to fully understand the article. This might have implications for the sentence structure. It might be that the need for clarity influences the type of word classes being used, resulting in a decrease of pronouns and an increase of nouns and verbs. The end result is a score that is close to that of an academic paper, with an increased amount of adverbs to adhere to the specific writing style of a journal article.

Popular science speeches score the lowest as expected. The lower score is mainly due to the heightened use of pronouns, which can be explained by the setting of these presentations. The presentations are held before a live audience, thus allowing for more contextual use of speech, since the audience is present and is fully aware of all contextual factors outside the text. Register might play a role in the more popular perception of this genre as well, however, changes in register cannot be perceived by the F-formula. Register will be further considered in the discussion section. Nominalisation seems to be less present in this genre, as the more deeply formal nouns and prepositions are replaced with verbs and pronouns. Nominalisation will be discussed further in the discussion section. The popular science speech is also the only genre where interjections were observed. This might be due to the fact that popular science speeches were analysed using word for word transcriptions of the speech, rather than a written summary of the speech, as with the conference papers.

6.2.2. Do the written genres score higher than the spoken genres?
The scores confirm a difference in written and spoken genres. However, when ranking the scores from high to low, the ranking is as follows:

1. Academic papers
2. Conference papers
3. Popular science articles
4. Popular science speeches
While the written academic papers are ranked highest, the second highest score is a spoken genre: the conference paper.

This is a remarkable result; however, this result can be explained, as was mentioned in the discussion of the first research question.

The last two genres of the ranking are – as predicted- popular science articles and popular science speeches.

Popular science articles might rank lower than academic paper because of the audience it is aimed at: non-experts. Since the audience is assumed to not have any background knowledge, formulas and abstract theories need to be avoided and need to be replaced by a logical, interesting story that can attract any reader and will leave the reader without further questions. As the style can be more storytelling, more pronouns and verbs can be used to ensure a flowing story that invites the reader to read.

The popular science speeches especially fulfil the expectations with a rather low score. Popular science presentations are held before a live audience, which allows for more use of contextual language such as pronouns. Additionally, these presentations are aimed at an audience of non-experts, causing the speaker to adjust the used language. This may include using more verbs to ease the flow of the presentation.

6.2.3. Do the more popular genres score lower than the more specialised genres?
This study confirms the hypothesis that the more popular version per category scores the lowest. When calculating the difference in score between the less popular and more popular genres of the written category, the result is 7.15 points. This result is remarkably low and might be due to the use of different register per genre, which impacts the way the reader feels about the text but which cannot be observed in the produced scores.

In the spoken category it can be remarked that the difference in score is much higher. The difference is 17.27. However, the score of the conference papers might be influenced by the
absence of the register factor in the formula by Heylighen & Dewaele, as mentioned in the discussion of the first research question.

Considering scores as they are it can be noted that the order of the higher and lower scoring genre is as predicted in the hypothesis.

Lastly, it needs to be remarked that the scores of three of the four genres are relatively close to each other. This influences the conclusion of this research question, as the difference in scores are, at times, minimal and might give an altered perception of the truth.

6.3. Discussion

6.3.1. Register
The influence of register on the results of this study has already been mentioned in the results section. The influence of register is an important part of the research, especially in the genre of the popular science article, as the results suggest a minimal difference in the writing style of an academic paper and a popular science article. The perception of the reader, however, would suggest otherwise. This means there is a gap in what the test labels as deeply formal and what the reader perceives as deeply formal. The reason for the similar scores is the similar amount of nouns and verbs being used in both genres. The reason for the difference in perception might be allotted to register. A difference in register does not necessarily mean a difference in the usage of certain word classes. For example, a journalist could read a sentence in an academic paper and copy the whole sentence, while only adjusting the words themselves, not the word classes. The journalist might, for example, change a technical noun to a more popular noun, or exchange a rather formal sounding verb for a more suitable verb. This would maintain the amount of a certain word class, while still making a difference in perception.

6.3.2. Nominalisation
As mentioned previously, a possible explanation for the high amount of nouns and prepositions overall might be nominalisation, which means converting word classes that are not nouns into nouns. As a result of the increase in nouns, an increased amount of prepositions can be observed as a means to link the words together into a comprehensible sentence. Nominalisations are more
common in written scientific texts, as the writer usually aims to sound more deeply formal by avoiding verbs, which are commonly used in expressions using surface formality, as mentioned in the chapter on formality. It is therefore quite remarkable to find characteristics of nominalisations in the language used in a conference paper. As previously mentioned, conference papers do possess a higher level of verbs and a lower level of nouns than is the case with academic papers, the difference is, however, remarkably small. Nominalisations might be an explanation for this result, as conference papers are not, in fact, word by word transcriptions of a conference speech, as can be observed above. Instead, they seem to be written texts that reflect the content of the speech. This might possibly influence the writer and cause the writer to use nominalisations more frequently than the author would in a live speech before an audience.

6.4. Concluding remark
The remarkably high results of this study show that the F-formula by Heylighen & Dewaele does not integrate all elements necessary to fully assess the formality of a text as they are perceived. This study shows that including the chosen register is of paramount importance when fully investigating the language use in several genres. As this matter was not picked up by Heylighen & Dewaele, additional research might be useful in this regard.

The results of this study also suggest that the genre of conference papers may lean closer to written discourse than to spoken discourse. While the genre does share characteristics with spoken discourse, there are elements that suggest an influence of written discourse as well.

It should also be noted that this research does possess some limitations, the first and foremost being the size of the corpus, which might not be sufficient to draw generalizing conclusions from this research.
7. CONCLUSION

This paper investigated the formality level of four spoken and written genres: academic papers, conference papers, popular science articles and popular science speeches. The level of formality was determined by analysing a corpus of twenty samples, selected by criteria of length, source and time. Academic papers and conference papers were selected on the basis of topic, length and year of publication. The popular science articles were selected on the basis of topic, magazine of publication and year. The popular science speeches were selected from the smartphone and technology search categories on the official TED website and were selected on the basis of topic and year of publication. The corpus was analysed by a POS-tagger. The programme used in this study is “The multilingual LT3 linguistic pre-processing toolkit” by the Van de Kauter & Lefever (2013). The programme labelled all samples so that all words were assigned to a certain word class. Then the amount of nouns, adjectives, prepositions, articles, pronouns, verbs, adverbs and interjections was observed and employed to calculate the F-formula created by Heylighen & Dewaele. The results of the F-formula were then examined and a ranking was made on the basis of highest score to lowest. The ranking from highest to lowest score is as follows: academic papers, conference papers, popular science articles and popular science presentations. Additionally, a comparison was made between the written and spoken categories and the popular and more specialised categories. It can be concluded that all genres scored relatively high on the test. Especially the conference papers scored higher than previously assumed. It can also be concluded that the written category scores higher than the spoken category and that the popular category scores less than the more academic category, although it needs to be remarked that the exceptional result of the conference paper genre might have an influence on the comparability of the results.

It is especially interesting that all genres score rather high on the test. This might be due to the fact that changes in register cannot be recorded by the F-formula. This might then be an interesting subject for another study, which uses a larger corpus and different means of observation.
8. BIBLIOGRAPHY

8.1 Primary sources

8.1.1. Popular science speech

8.1.2. Popular science article

8.1.3. Conference paper
recognition tool for the blind. In Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE (pp. 4538-4541). IEEE.


Yamada, M. (2013, April). "Talking is the new typing": Challenging smartphone users to dictate instead of type in order to enrich the mobile learning experience. TCC Conference.

8.1.4. Academic papers


8.2 Secondary sources


9. APPENDICES

See CD attached to this thesis.