Perfect use in dialogue contexts from *Harry Potter and the Philosopher's Stone*

Is the Perfect tense use in dialogues written in *Harry Potter and the Philosopher's Stone* comparable with the Perfect tense use in Spontaneous speech?

Isa Buwalda (5901340)

Universiteit Utrecht

Bachelor Artificial Intelligence
Utrecht University
Supervisor: Henriëtte de Swart
Second corrector: Sjoerd Stuit
January 2020

Table of Contents

1   Introduction

Within the field of linguistics a lot of research has been done to the semantics and the use of different verb tenses in different languages. For some verb tenses an exact definition has been provided, however the Present Perfect has not been precisely defined yet. There are few theories adressing this verb tense, but no consensus has been reached. The Project Time in Translation studies the Present Perfect across languages and aims to find an exact definition of the verb tense. Large datasets containing different languages are used to find this definition. One corpus used to unveil the use of the Perfect in dialogue is *Harry Potter and the Philosopher's Stone*, because it contains both narrative and dialogue parts and exists in lots of different languages.

However, it is never tested whether Perfect use in dialogues in this book is a good reflection of the Perfect use in spontaneous dialogue. Of course this is important to know if we want to draw conclusions from this book about the way we use the Perfect in real life. This is where the subject of my research is coming from. I am going to investigate if the dialogue in *Harry Potter and the Philosopher's Stone* is comparable to spontaneous speech and thus is a right corpus to use for Time in Translation. To do so I formed the following research question: Is the Perfect tense use in dialogues written in *Harry Potter and the Philosopher's Stone* comparable with the Perfect tense use in spontaneous speech.

To answer my question I used a system called Dialogue Act Annotation which was created based on the Switchboard Corpus, a corpus consisting of spontaneous speech. I annotated the sentences in dialogues with a Perfect from the Harry Potter Corpus and compared the results to the results of the Switchboard Corpus. This comparison will be discussed in this paper.

The paper is structured in the following way. In the next chapter I will discuss why this research is relevant to the field of Artificial Intelligence. In chapter 3 I will discuss existing literature about the Perfect to see where linguistics stands with this topic at the moment. Then, in chapter 4 I will form my hypotheses based on the theory discussed in chapter 3. In the following chapter I will discuss what methodology I used to be able to answer my research question. In chapter 6 the results of this methodology will be stated. After that I will discuss my results and I will compare the data of both corpora. In the last chapter I will review my hypotheses and provide an answer for my research question. In this chapter I will also discuss issues I encountered during this research and I will do suggestions for further research.

2 Relevance for Artificial Intelligence

On the first glance the subject of my research might look like it has not got any relation to Artificial Intelligence. In this chapter I will briefly contradict this assumption.

The most important man for the rise of Artificial Intelligence has to be Alan Turing. In 1950 he designed the Turing Test to provide a definition of intelligence. During this test, a person will interact with a human as well as a computer trough written text. When the person is not able to tell the computer apart from the human, the computer passes the test because it means it was able to act humanly. To be able to pass this test it is, among other things, very important for a computer to be able to use and produce natural language (Russell and Norvig, 2010).

Production of natural language is where computers still struggle, especially within the field of automated translation. While programs are able to make translations that give speakers from different languages the ability to understand each other, the translations very rarely feel natural. This could be due to several factors, but one may be that the wrong verb tense is used. To be able to better the verb part of translation systems, it is important to have an exact definition for all the verb tenses in different languages. This definition is still missing for the Perfect. The project Time in Translation aims to find this definition and thus end the discussion about this tense. My research offers support in this research by making sure the corpus they use is the right one to utilise. In the following chapter I will provide some background on the current status of the Perfect tense and on the system I will use to do my research.

3 Theoretical Background

In this chapter I will explain the existing literature about the Present Perfect. In chapter 3.1 I will describe in what kind of sentences a Perfect can be used and in what sentences it is not possible to use the tense. In chapter 3.2 I will describe how the Perfect differs across languages. This is important to mention because it is one of the reasons why there is no clear meaning for the semantics of the Perfect yet. Chapter 3.3 describes and shows that it is very hard to provide a semantic characterization of the state that is introduced by a Perfect. This leads to chapter 3.4 which describes the aim of the current research which focuses on the use of the Perfect in dialogue. This chapter thus describes the small amount of existing literature about the Perfect in dialogue. In chapter 3.5 I will describe the Dialogue Act Annotation system that I will use in this research and how it has been created based on the Switchboard Corpus.

3.1 Restrictions on the English Perfect

Within linguistics a lot of research has been done into the semantics of the Perfect tense, however no consensus has been reached yet. First I think it might be useful to describe the difference between the Present Perfect and the Simple Past in English, because both tenses describe past events. The Present Perfect can be defined as describing a past event that introduces a present state, whereas a Simple Past only describes the past event (de Swart, 2007). The Present Perfect is formed by the auxiliary verb 'to have' combined with a past participle. There also is a Present Perfect Continuous in which the past event is not necessarily finished yet. This tense is formed by 'have/has been' combined with the present participle, which is the root of a verb +'ing'. From now on when talking about a Perfect, I refer to the Present Perfect.

De Swart (2007) describes a sytem introduced by Reichenbach (1947) which clearly distinguishes the English Perfect from the other English tenses. He uses three notions: the Event time, the Reference time and the Speech time. In the Simple Past the Event time and the Reference time are simultaneous and come prior to the Speech time. In the Present Perfect the Event time comes prior to the Reference Time and the Speech time, which in this tense are simultaneous. This difference between the tenses is made clear by example (1), in which (1a) contains a Simple Past and (1b) contains a Present Perfect. It is clear that in both these sentences the event of Sara leaving comes prior to the Speech time. The difference is the Reference time. Sentence (1a) only describes the event, whereas sentence (1b) describes the event of Sara leaving the party with the result that she is not there anymore. Citing de Swart: "it maintains the importance of the Speech time".

(1) a. Sara **left** the party.   (Simple Past)
     b. Sara **has left** the party.   (Present Perfect)

From this theory it follows that the Perfect can not be combined with time adverbials indicating the Event time because it is not simultaneous to the Reference Time. The Perfect can be combined with time adverbials indicating the Reference time. An example for this can be seen in example (2), in which (2a) is incorrect use of the Perfect (indicated with *) and (2b) is correct use of the Perfect (de Swart, 2007).

(2) a. *Sara **has left** at six o'clock.
     b. Sara **has left** this afternoon.

Moreover, de Swart (2007) gives some attention to the context in which a Perfect may be used. She describes how Boogaart (1999) states that a Perfect is an incorrect tense in narrative contexts. This is because while telling a story it is important to be able to describe a series of events. De Swart (2007) shows this by giving example (3a) in which a Present Perfect is used in the subordinate clause starting with *when*, which reflects narration. Example (3b) shows a Simple Past has to be used for contexts like this.

(3) a. *When John **has seen** me, he **has got** frightened.   (Present Perfect)
     b. When John **saw** me, he **got** frightened.   (Simple Past)

3.2 Cross-linguistic variation

While these rules for the use of the Perfect seem to be very clear, they are not found in every language. So there is some cross-linguistic variation in the use of the Perfect. For example, the Present Perfect in Dutch, French and German can be found with a time adverbial referring to Event time, while in English this is not possible as we saw above. De Swart (2007) shows this with the following examples in (4).

(4) a. *Sara **has left** at six o'clock.   (English)
     b. Sara **is** om zes uur **vertrokken**.   (Dutch)
     c. Sara **est partie** á six heures.   (French)
     d. Sara **ist** um sechs Uhr **abgefahren**.   (German)

Besides that, she shows that the rules for the *when-clause* stated before also apply to Dutch, but do not apply to French and German. This means the Perfect can not be used in narrative discourse in Dutch, which is also shown by Le Bruyn, van der Klis and de Swart (2019). This is a noteworthy finding because it creates a difference between narrative discourses in various languages. Furthermore, this is where a distinction between dialogue and narrative discourse arises, which is important for the course of my research. More about how the Perfect is used in dialogue can be found in chapter 3.4. The corpus-research by Le Bruyn, van der Klis, and de Swart (2019) also shows that Dutch has a broader distribution of the Perfect in dialogue than English. In their talk given in 2015 this is clearly demonstrated with the English example (5), which contains a Simple Past, but which is translated to Dutch with (6), which contains a Present Perfect.

(5)   Guys, come and have a look! I **found** a baby panda!   (English, Simple Past)

(6)   Jongens, kom kijken! Ik **heb** een baby panda **gevonden**!   (Dutch, Present Perfect)

On top of that Le Bruyn, van der Klis and de Swart (2019) show that in Dutch dialogue the Perfect is often used for event verbs, these are verbs that describe an action or happening in the past. The Simple Past on the other hand is often used for state verbs, which describe a condition or state in the past. They then suggest that the Perfect in Dutch is used to describe a state that is no longer holding at Speech time. This is different for English, in which the resultant state is still holding at Speech time.

3.3 The Perfect state

So as we see, the state a Perfect introduces varies across languages. However, for the English language alone there is no agreement about the meaning of the Perfect either.

Most research distinguishes between four semantic interpretations of the Perfect. Nishiyama & Koenig (2010) give the following four examples for these interpretations:

(7)   I can't come to your party tonight. I'**ve caught** the flu.   (Resultative Perfect)

(8)   I **have read** *Pirncipia Mathematica* five times.   (Existential Perfect)

(9)   I'**ve known** Max since 1960.   (Continuative Perfect)

(10) Malcolm X **has** just **been assassinated**.   ('Hot news' Perfect)

In a resultative Perfect like (7) the resultant state of the past event is still continuing in the Speech time. In an existential Perfect like sentence (8) an event has happened that is still relevant now, at Speech time, but it is not continuing anymore. In a continuative Perfect sentence like (9) the past event described is still going on in present time. The last use of the Perfect, the 'hot news' Perfect is used to report hot news. However, this last one can also be reduced to an existential Perfect. From now on the 'hot news' reading will be treated as such.

This seems like a clear distinction between the different interpretations of the Perfect. However, linguists do not agree on how to categorize the interpretations and whether the differences mean that there are different semantic definitions for each Perfect. In other words, there is no consensus about what state is introduced by a Perfect. Nishiyama and Koenig propose that this state can not be characterized semantically. They show several semantic specifications for the category of the Perfect state and their problems.

The first theory Nishiyama and Keonig describe is introduced by Kamp and Reyle (1993) and states that the Perfect state starts when the base eventuality (the past eventuality described by the verb) ends. Thus the relationship between the state and the base eventuality is temporal. The main downside to this theory is that there can be states that start when the base eventuality ends, but which are irrelevant to the utterance. The example given for this problem can be seen in (11). In this case (11a) is the resultant state of (11). Nevertheless, event (11b) could have also started after the base eventuality, but it is irrelevant to the utterance.

(11) Ken **has broken** his leg.   (Resultative Perfect)
       a. His leg is broken.
       b. Susan is married.

The second theory described defines the Perfect state as being a consequent state of the base eventuality. In this case, a causal relation between the base eventuality and the Perfect state is assumed. Example (7) makes this very clear. Me catching the flu has caused me to have the flu and not be able to come to the party. However, the problem to this theory is that it does not work for continuative Perfect readings. This is because the state introduced by a continuative Perfect is not the cause of the base eventuality. Example (9) explains this. The continuing state of me knowing Max is not caused because of me knowing Max since 1960.

The third semantic specification for the category of the Perfect state treats it as a permanent state. The permanent state in this case is the state of the base eventuality having occurred. There are a few problems to this theory. The most clear one in my opinion is that most resultant states of the Perfects do end at some point in time. For example, resultant state (11a) is a state that will end after some time and thus is not permanent.

In short, it is hard to provide a semantic characterization of the Perfect state. Nisiyama and Koenig thus propose a pragmatic characterization. In other words, we have to look at context to infer the meaning of a Perfect (Nishiyama & Koenig, 2010).

3.4   The English Perfect in dialogue

The project Time in Translation aims to specify a semantic as well as a pragmatic meaning for the Perfect. This is done by this using cross-linguistic variation. To do this, parallel-corpus techniques are applied to the different translations of *Harry Potter and the Philosopher's Stone*. This book is chosen as a corpus because it has a lot of translations and it contains both dialogue and narrative parts. In the current research, the goal is to focus on the pragmatic role of the Perfect in dialogue. Currently, very few research has been done to this subject. There are two main theories widely accepted.

Portner (2003) states the Perfect is used in dialogue to answer the question under discussion. An example for this can be seen in conversation (12).

(12) A: We need to get an explanation of George Eliot's style. Who can we ask?
       B: Well, George Eliot wrote *Middlemarch*, and if someone read an author's book, they understand her style. Unless they are stupid of course. Mary is smart and she **has read** *Middlemarch*.

Nishiyama and Koenig (2010) give another use for the Perfect in dialogue. They show the Perfect is often used to introduce a new topic in conversation. An example can be seen in conversation (13) (de Swart, 2018).

(13) A: **Have** you **seen** *Dancing with Wolves*?
       B: Yeah. I've seen that. That was a really good movie.

These are two very different theories. Following both theories it says that when a Perfect is used in sentences that are not questions, they most likely are used to answer a question. And when a Perfect is used in a question it most likely introduces a new topic. These theories are very small and do not really provide an exact and well-defined definition for the Perfect.

So to finally make an exact meaning of the Perfect, Time in Translation uses *Harry Potter and the Philosopher's Stone*. This is used because it contains both narrative and dialogue parts and it has a lot of translations, which makes cross-linguistic research possible. However, no analysis has been done to see whether the dialogue parts in this book are

a good representation of spontaneous speech. This is important to know, since the goal of the research is to make conclusions about the use of the Perfect real life conversations, like example (12) and (13). My research investigates the spontaneity of the dialogue parts containing a Perfect tense of the used corpus. The main question of my research thus is:

- Is the Perfect tense use in dialogues written in *Harry Potter and the Philosopher's Stone* comparable with the Perfect tense use in spontaneous speech?

3.5 Dialogue Act Annotation on the Switchboard Corpus

To answer the main question (Is the Perfect tense use in dialogues written in *Harry Potter and the Philosopher's Stone* comparable with the Perfect tense use in spontaneous speech?), I will use an annotation system called Dialogue Act Annotation. The system allows an analysis of discourse structure. A Dialogue Act indicates the intention of an utterance, this is done by looking at semantic, pragmatic and syntactic properties of the utterance. It thus is a very precise way to expose the form of conversation because it indicates with what intention utterances are expressed. The system was created based on the Switchboard Corpus, which is a corpus containing 260 hours of spontaneous phone conversations (Stolcke et al., 2000). These conversations were between people that did not know each other. An example of a Dialogue Act, which occurs most in the Switchboard Corpus, is the *Statement-non-opinion*-act (Jurafsky et al, 1997). This act is assigned to utterances that are statements where no opinions are expressed. Example (14) shows such an utterance.

(14)   Me, I'm in the legal department.   (Statement-non-opinion)

But the system also has labels for sentences that do not have semantic meaning. For example the *Acknowledge (Backchannel)*-act, which is also sometimes called a 'continuer'. This label is given to utterances that are used by speakers to let the conversation partner know they agree or understand what that person is saying. An example for this act is in (15).

(15)   Uh-huh.   (Acknowledge (Backchannel))

Furthermore, the system also distinguishes between different kinds of questions. To give an example they discriminate between the *Yes-No-Question* and the *Declarative Yes-No-Question*. Both of these Dialogue Acts refer to a question that can be answered with yes or no, however when the question is declarative it means the question is formed like statement with a question mark behind it. Examples for both Dialogue Acts can be found in (16) and (17) respectively.

(16)   Do you have any special training?   (Yes-No-Question)

(17)   So you can afford to get a house?   (Declarative Yes-No-Question)

So each utterance in the Switchboard Corpus was assigned a Dialogue Act label. The result of this task was a count for each act occurring in the corpus. In total 220 acts were created, which were clustered into 42 larger Dialogue Acts, also called SWBD-DAMSL labels (Stolcke et al., 2000). In Appendix A I added a list of the 42 acts with an example for each one from the Switchboard Corpus and how often they occurred in the corpus (Jurafsky et al, 1997). For my research the SWBD-DAMSL labels will be used because it is a compact set but still it clearly uncovers the structure of dialogues and this way the Harry Potter corpus can be compared to the Switchboard Corpus easily.

However, the Switchboard Corpus contains all kinds of sentences with all kinds of tenses. For the project Time in Translation and in my research, the main focus is on the use of the Perfect in dialogue. This is why Tellings, van der Klis, Le Bruyn and de Swart (2019) did an analysis to see how many times a Perfect occurred in each Dialogue Act in the Switchboard Corpus. This was done by running their Perfect Extractor on the Switchboard corpus to extract the Perfects. They found a significant association between the use of the Perfect and questions. The chance on use of a Perfect was 1.5 times higher in questions compared to statements. This supports the theory described in chapter 3.4 of Nishiyama and Koenig (2010) about the use of the Perfect for topic negotiation like in example (13), because this is often done in question form. Example (18) obtained from the Switchboard Corpus introduces a topic in question form. In this case the speaker wants to talk about the movie *Robin Hood*.

(18)   **Have** you **seen** *Robin Hood* yet?   (Yes-No-Question)

Also, many Perfects were found in the *Conventional-Closing*-acts. These utterances are used to close off a topic, an example obtained from the Switchboard Corpus can be seen in (19). This use of the Perfect can account for the theory of Portner (2003) that states a Perfect is often used to answer a question under discussion, and thus close off a topic, like in example (12).

(19)   Well, it has, it **has been** really fun.   (Conventional-Closing)

The main results of the analysis of Tellings, van der Klis, Le Bruyn and de Swart (2019) can be seen in table 1. I also added a column to give an example containing a Perfect of each Dialogue Act. Each Dialogue Act also has an abbreviation which were also used when annotating the Switchboard Corpus. The abbreviations are given between parantheses. There are also some subcategorical Dialogue Acts, these are indicated with the ˆ. An example of such a subcategory is the declarative one (`^d`). Important for my research are the different question acts and the *Conventional-Closing*-acts to be able to say something about the theories of Portner (2003) and Nishiyama and Koenig (2010) about the use of the Perfect in dialogue contexts.

Dialogue Act	# of Perfects		Examples
Negative non-no answers (ng)	37/203	18.2%	But I **haven't been** there.
Conventional-closing (fc)	144/1100	13.1%	Well, it has, it **has been** really fun.
Declarative yes-no-question (qyˆd)	93/1011	9.7%	You**'ve** never **seen** *Northern Exposure*?
Declarative wh-question (qw)	7/74	9.5%	What kind of degree **have** you **got**?
Yes-no question (qy)	400/4412	9.1%	**Have** you **seen** *Robin Hood* yet?
Statement-non-opinion (sd)	5931/69972	8.5%	I**'ve been** with them for seven years.
All other Dialogue Acts		<8.5%	

Table 1: Perfects in Dialogue Acts in the Switchboard Corpus

Furthermore, in this analysis Tellings, van der Klis, Le Bruyn and de Swart categorized the different Dialogue Acts in five bigger categories and looked at the ratio of tense use in these categories. They found that the Perfect is almost never used in agreement acts, backchannel acts and hedge acts. I have already given an example of a backchannel in (15). *Hedges* are used to reduce the certainty of a speakers statement or answer, an example for this act is given in (20).

(20)   I don't know if I'm making sense or not.   (Hedge)

The results of the analysis are shown in the diagram of figure 1. The bar charts of the agreement, backchannel and hedge acts show almost no blue, meaning these acts very rarely contain a Present Perfect.
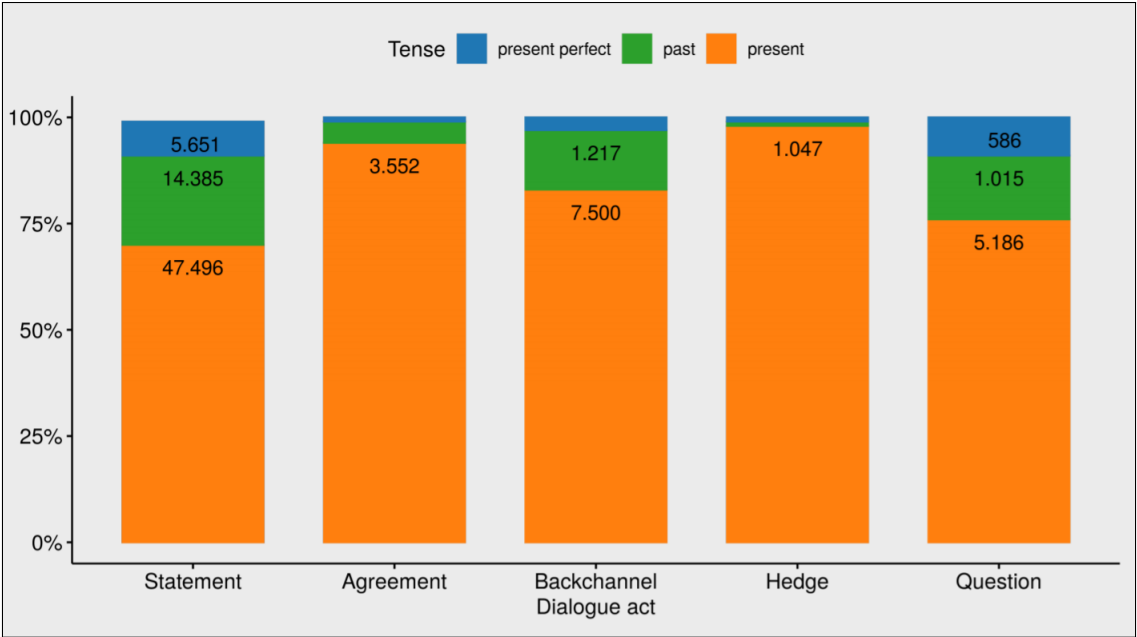
Figure 1: Tense use in larger Dialogue Act categories

My research, in which I will compare written dialogue to spontaneous dialogue, is a little similar to the corpus-research of Levshina (2017). She compared linguistic features in film subtitles to those in naturally occuring informal conversations. She found that subtitles contain fewer pause-fillers, reformulations and discourse markers. Discourse markers can be described as linking words or phrases in conversation. Hedges and backchannel acts like example (20) and (15) are examples of discourse markers. The Harry Potter Corpus can be compared to the corpus of film subtitles because it both consists of written and designed dialogue.

Based on the literature above and on the extensive analysis on the Switchboard Corpus, some hypotheses can be made about the use of the Perfect in certain Dialogue Acts in the Harry Potter Corpus. These hypotheses will be treated in the next chapter.

4   Hypotheses

As mentioned before, I aim to find an answer to the question:

- Is the Perfect tense use in dialogues written in *Harry Potter and the Philosopher's Stone* comparable with the Perfect tense use in spontaneous speech?

This will be done to hopefully justify the use of the Harry Potter Corpus to make conclusions about the use of the Perfect in dialogue. This way we can get closer to an exact meaning of the Perfect. So far we saw there is very few literature about the English Perfect in dialogue. One theory claims the Perfect is used to answer the question under discussion and the other one claims it is used to introduce a new topic in conversation. These different theories arise because there is no consensus yet about the state introduced by a Perfect. This is due to the different interpretations of Perfects and cross-linguistic variation in the use of the Perfect. Based on this literature, which is described in more detail in chapter 3, I formed some hypotheses.

The first and main hypothesis is the null hypothesis and is stated below.

H0. There is a similar Perfect distribution on the Dialogue Acts in the Harry Potter book as in the Switchboard Corpus.

To test this I will perform the Dialogue Act Annotation on the sentences in dialogue with a Perfect in the Harry Potter Corpus and compare this to the data of the Switchboard corpus. I expect the null hypothesis to be true because on the first impression the dialogues in the Harry Potter book come across spontaneous. This is the reason why it is already used as a corpus in the project Time in Translation. When taking a closer look I hope to see a similar Perfect distribution on the Dialogue Acts of the Switchboard corpus, since this corpus only consists of spontaneous speech. This main hypothesis can be divided in smaller subhypotheses, which I will now sum up.

The first hypothesis is called subhypothesis 1 and is stated below.

S1. When annotating sentences in dialogues containing a Perfect, questions introducing new topics and *Conventional-Closing*-acts will be found.

This second hypothesis is based on the literature of Portner (2003) and Nishiyama and Koenig (2010) about the use of the Perfect in dialogues. Furthermore, when annotating the sentences I expect to find a majority of the Dialogue Acts to be *Statement-non-opinion*-acts. When looking at the data of the Switchboard Corpus, we see that this act dominates the data. Namely, 36% of all utterances in dialogues are *Statement-non-opinion* (Stolcke et al., 2000). The corpus analysis of Tellings, van der Klis, Le Bruyn and de Swart shows this is not different when focusing on the utterances containing a Perfect.

The results in figure 1 of the analysis of Tellings, van der Klis, Le Bruyn and de Swart (2019) make me form my second subhypothesis.

S2. A very small amount of Perfects in agreement, backchannel and hedge acts will be found.

The corpus-research of Levshina (2017) in my opinion strengthens this hypothesis, because it shows that the written and designed dialogue of film subtitles does not contain a lot of discourse markers, for example hedges and backchannel-acts. Since the Harry Potter corpus also contains written and designed dialogue, I expect to see the same thing happening.

My third subhypothesis is stated below.

S3. There will be a higher chance on use of a Perfect in questions compared to statements.

This hypothesis is based on the finding of Tellings, van der Klis, Le Bruyn and de Swart (2019) that states the chance on the use of a Perfect was 1.5 times higher in questions compared to statements in the Switchboard Corpus.

In the next chapter I will describe how exactly I am going to collect my data to test my hypotheses.

5 Methodology

In this section I will describe what steps I will take to obtain the needed data from the Harry Potter corpus. This data consists of Dialogue Acts found in utterances containing a Perfect in the dialogue parts of the book.

5.1 PerfectExtractor

The first step in this research is to get a corpus consisting of all sentences in dialogues containing a Perfect tense in *Harry Potter and the Philosopher's Stone*. I will do this for the English as well as the Dutch book. This is done for both because when time allows it I will also annotate the Dutch translation. An algorithm called the PerfectExtractor is used to get this corpus. The algorithm allows extraction of Present Perfects and related forms in different languages (voltooid tegenwoordige tijd in Dutch). As we saw, the Present Perfect consists of the auxiliary verb 'to have' with a past participle, so it seems easy to find the forms. However, other words can come between auxiliary verb and the past participle in sentences. Furthermore, the Present Perfect Continuous is partly formed the same way, but it has to be categorized separately.

Despite this difficulties, the PerfectExtractor still extracts the right sentences. It does so by looking for auxiliary verbs. When it finds one, it looks for a past participle right from it, allowing words in between. If it finds punctuation or a different verb, it stops. It extracts occurrences of the Present Perfect Continuous as well and classifies them separately. I will run this tool on a corpus that consists of the dialogue parts of the book. Some modals were excluded, namely: "could/should/must have written" and "have got to write". This is done because these verbs have a different tense use than the non-modal verbs and there is a lot of cross-linguistic variation. After this step, a dataset is created containing all sentences from the dialogue parts of the English and the Dutch book that have a Perfect tense in them.

5.2 TimeAlign

Now that I have a dataset with all the necessary sentences, I need to align the verb phrases from both languages. This is done manually with a tool called TimeAlign. This shows a sentence from the original English book with a verb phrase marked green. On the right it shows the corresponding Dutch fragment of the book. In this fragment the verb phrase has to be selected that matches the green verb phrase in the English fragment. The tool also contains some checkboxes which can be checked if they apply. This is in case the PerfectExtractor made a mistake. An example of a screen in TimeAlign is in figure 2.

Not all translations were literal, some of them were translated a bit freely. They were accepted however, because otherwise the dataset would become small very quickly.

Figure 2: An example of TimeAlign

5.3 Dialogue Act Annotation

The output from the TimeAlign tool are Excel-files for each chapter, containing columns for the English fragments, the Dutch translations, the verb form from the source, the verb form from the translation and some other information which is less relevant for this research. The next step is to annotate each English sentence containing a Present Perfect with its Dialogue Act. As we saw in chapter 3.5 this is a good way to reveal the structure of dialogues and this way the Harry Potter data can be compared to the Switchboard Corpus. If time allows, I will also annotate the Dutch sentences. To annotate, I add two columns to the excel file. The first column is for giving the main category of the Dialogue Act. As we saw in chapter 3.5 some Dialogue Acts also have subcategories, an example of this is the *Yes-No-Question* (`qy`) which has a subcategory *Declarative Yes-No-Question* (`qy^d`). Those subcategories are given in the second added column. Sometimes it is necessary to look at the context of a sentence, to really make sure the right Dialogue Act is chosen. Figure 3 shows part of what the excel file for chapter 1 looked like after it was annotated. In column X it shows the utterance from the English source and column Q shows the Dutch translation. Column V shows in what tense was used in the source utterance and in column Y the Dialogue Acts are given. Column Z gives the subcategories if they were there.

	Q	R	S	T	U	V	W	X	Y	Z	AA
1	full frag	source	source	source	source	source	source	source	DAA	DAA subcategory	
131	Wees blij	50356	1.xml	s15.5	has gone	present perfect		Rejoice , f	sd		
163	'En ten sl	50088	1.xml	s22.1	have beer	present perfect cont		'And final	sd		
164	'En ten sl	50089	1.xml	s22.1	have repo	present perfect		'And final	sd		
165	Uilen jage	50097	1.xml	s22.2	have beer	present perfect		Although	sd		
169	'Nou , dat	50502	1.xml	s23.1	have beer	present perfect cont		'Well , Te	sd		
172	Kijkers uit	50090	1.xml	s23.2	've had	present perfect		Viewers a	sd		
174	Kijkers uit	50092	1.xml	s23.2	have beer	present perfect cont		Viewers a	sd		
180	'Eh - Petu	50368	1.xml	s25.5	haven 't h	present perfect		'Er - Petu	qy	^g	
278	'M'n best	50215	1.xml	s48.1	've seen	present perfect		'My dear	sd		
285	'Je zou de	50575	1.xml	s52.2	have notic	present perfect		'You 'd th	sd		
296	'Tenslotte	50063	1.xml	s53.2	've had	present perfect		'We *'ve*	sd		
308	Hij *is* to	50311	1.xml	s55.2	has gone	present perfect		I suppose	qy	^d	
315	'Zoals ik a	50382	1.xml	s59.2	has gone	present perfect		'As I say ,	sd		
317	Altijd maa	50012	1.xml	s60.2	have beer	present perfect cont		All this " Y	sd		
321	Ik *heb* r	50226	1.xml	s60.5	have seen	present perfect		I *have* r	sv		
336	Ik *heb* r	50254	1.xml	s64.2	haven 't b	present perfect		I *haven*	sd		
346	Blijkbaar h	50538	1.xml	s66.1	had been	present perfect cont		It seemed	sd		
372	Niemand	50294	1.xml	s71.6	's gone	present perfect		No one kn	sd		
379	'Na alles	50458	1.xml	s73.2	's killed	present perfect		'After all I	qy	^d	

Figure 3: The excel-file with Dialogue Act Annotation for chapter 1

Now I have created my dataset I will summarize it in the next chapter. Thereafter I will write a discussion in which I compare the data from the Harry Potter Corpus to the data of the Switchboard Corpus.

6 Results

In this chapter I will discuss the data acquired by applying Dialogue Act Annotation to the Harry Potter Corpus.

6.1 Relevant Dialogue Acts

First I think it is useful to shed some light on the relevant Dialogue Acts in the Harry Potter corpus. For this I created table 2, which gives every Dialogue Act found in the Harry Potter Corpus, its abbreviation and one or more random examples from the corpus.

Dialogue Act	Example
Statement-non-opinion (`sd`)	**I've come** to bring Harry to his aunt and uncle.
	Ron, you**'ve got** something on your nose.
	See, there's Potter, who**'s got** no parents,...
Declarative Yes-No-Question (`qy^d`)	I suppose he really **has gone**, Dumbledore?
	Oh, I see - so you**'ve** never **been** to Brazil?
	You**'ve met** Malfoy before?
Tag-Yes-No-Question (`qy^g`)	I mean, he **hasn**'t **gone**, has he?
	They**'ve** never **lost** a hundred and fifty points in one go, though, have they?
Yes-No-Question (`qy`)	But **have** yeh **seen** anythin', Ronan?
	Haven't you **heard** what it was like when he was trying to take over?
Wh-Question (`qw`)	What **have** we **got** today?
	What**'ve** you **done** to him?
Rhetorical-Question (`qh`)	After what McGonagall and Snape **have said**?
	My dear Friar, **haven**'t we **given** Peeves all the chances he deserves?
Statement-opinion (`sv`)	I **have** never **seen** any reason to be frightened of saying Voldemort's name.

Table 2: Examples from each Dialogue Act in Harry Potter

Some Dialogue Acts appear to be very similar, especially all the question acts. I have already made a distinction between the *Declarative Yes-No-Question* and the usual *Yes-No-Question* in chapter 3.5. Now we also see there is a *Tag-Yes-No-Question*. This is a question which is stated as a statement, but has a tag-sentence behind it which makes it a question. Furthermore, we now see the *Rhetorical-Question*. This Dialogue Act seems very similar to the *Declarative Yes-No-Question*. However, the *Declarative Yes-No-Question* is formed as a statement with a question mark behind it, whereas the *Rhetorical-Question* is formed like a real question. Furthermore the *Declarative Yes-No-Question* is meant to be answered, whereas the *Rhetorical-Question* is not.

6.2 The Perfects and their Dialogue Acts

Table 3 gives in column two for each chapter the amount of Present Perfects. The other columns show all the different Dialogue Acts found and how many times they were found. In the last row all these numbers are added up.

Chapter	# of Perfects	sd	qy^d	qy^g	qy	qw	qh	sv
Chapter 1	19	14	3	1	0	0	0	1
Chapter 2	1	0	1	0	0	0	0	0
Chapter 3	3	3	0	0	0	0	0	0
Chapter 4	8	7	1	0	0	0	0	0
Chapter 5	11	11	0	0	0	0	0	0
Chapter 6	30	27	1	0	1	0	1	0
Chapter 7	9	8	0	0	0	0	1	0
Chapter 8	3	2	0	0	0	1	0	0
Chapter 9	14	13	0	0	0	1	0	0
Chapter 10	13	11	0	0	1	1	0	0
Chapter 11	5	4	0	0	0	1	0	0
Chapter 12	20	19	1	0	0	0	0	0
Chapter 13	13	11	0	0	0	1	1	0
Chapter 14	20	19	1	0	0	0	0	0
Chapter 15	24	19	1	1	2	0	1	0
Chapter 16	14	10	1	0	1	1	1	0
Chapter 17	22	18	0	2	0	1	1	0
Total	229	196	10	4	5	7	6	1

Table 3: Amount of Perfects and their Dialogue Acts in Harry Potter

6.3   Questions containing a Perfect

Since the rest of the Harry Potter Corpus is not annotated with Dialogue Acts for other tenses yet, it is not possible to give a ratio of a certain Dialogue Act with a Perfect relative to the overall appearance of a certain Dialogue Act. However, it is possible to count all question marks in the dialogue parts of the Harry Potter book. This way I can make a ratio comparing the amount of questions containing a Present Perfect to the overall appearance of questions. This is done in table 4.

Chapter	# of questions	# with a Perfect	Ratio	
Chapter 1	34	4	4/34	11.8%
Chapter 2	12	1	1/12	8.3%
Chapter 3	17	0	0/17	0.0%
Chapter 4	40	1	1/40	2.5%
Chapter 5	63	0	0/63	0.0%
Chapter 6	85	3	3/85	3.5%
Chapter 7	26	1	1/26	3.8%
Chapter 8	20	1	1/20	5.0%
Chapter 9	36	1	1/36	2.8%
Chapter 10	23	2	2/23	8.7%
Chapter 11	22	1	1/22	4.5%
Chapter 12	49	1	1/49	2.0%
Chapter 13	24	2	2/24	8.3%
Chapter 14	30	1	1/30	3.3%
Chapter 15	65	5	5/65	7.7%
Chapter 16	77	4	4/77	5.2%
Chapter 17	53	4	4/53	7.5%
Total	676	32	32/676	4.73%

Table 4: Amount of questions in Harry Potter

In the next chapter, I will discuss the results of Dialogue Act Annotation on the Harry Potter Corpus and I will compare it to the data of the Switchboard Corpus.

7 Discussion

In this chapter I will compare the results of applying Dialogue Act Annotation on the Harry Potter Corpus to the data of the Switchboard Corpus. To do so, I used very extensive data from the Switchboard Corpus provided by Time in Translation which I added in Appendix B. This data was summarized in the poster by Tellings, van der Klis, Le Bruyn and de Swart (2019) which is shown in figure 4. For ease I presented this diagram again in figure 4 below. In this discussion I will use the act categories of this bar chart to draw conclusions about the Harry Potter Corpus.
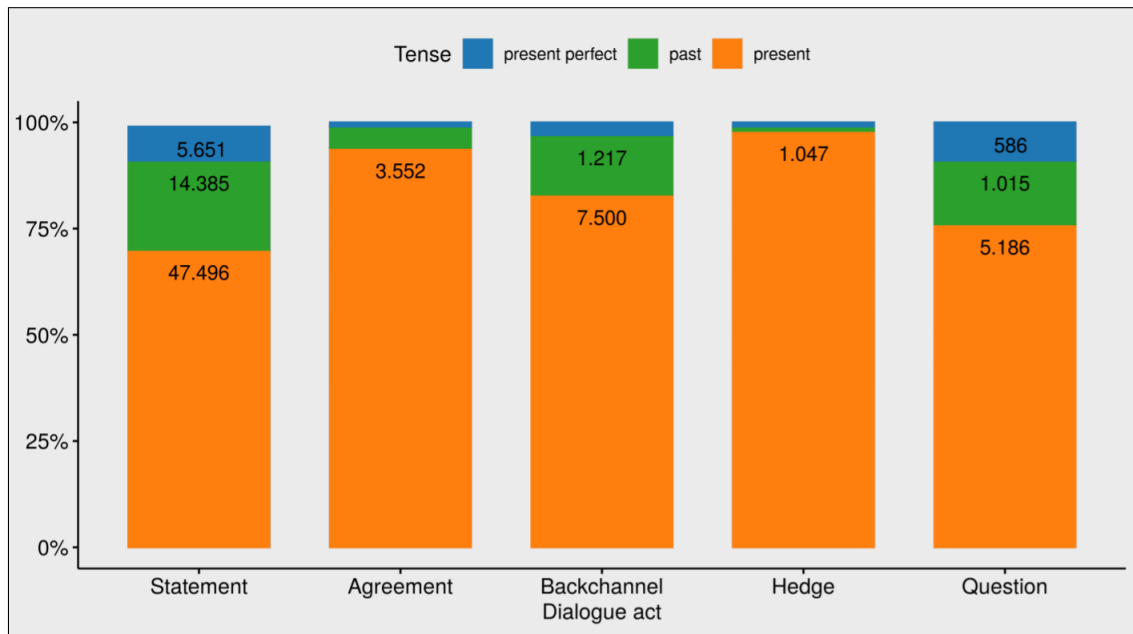

Figure 4: Tense use in larger Dialogue Act categories

As stated in the results above, it is not possible to compare the amount of each Dialogue Act in a Perfect tense to the overall appearance of that Dialogue Act in the Harry Potter Corpus, as done for the Switchboard Corpus in the analysis of Tellings, van der Klis, Le Bruyn and de Swart (2019). This is because only the sentences in dialogue containing a Perfect are annotated so far and there is not enough time to annotate more than that. However, there is one coordinating act I can do this for. This is the question-act, containing all Dialogue Acts that are questions. First I will discuss this act and after that I will look at the statement acts. Furthermore I will say somethings about questions introducing new topics and *Conventional-Closing*-acts. Lastly I will discuss the agreement, backchannel and hedge acts.

7.1 Questions

For the question act it is possible to compare the appearances with a Perfect to the overall appearance of the act. To do so I used table 4 in chapter 6.2. I also know the overall appearance of questions from the Switchboard Corpus (SB), which is shown in Appendix B. I put the data next to each other in table 5 and performed a Chi-Square test. This is a test to see whether two categorical variables are related or not. The test gives a p-value and if this value is lower than 0.05 it means that it is very likely there is a relationship between the variables. I also look at the effect size of the Chi-Square test, which tells how strong the association between both variables is. In this case, we want to see if the use of Perfects in questions is related to the corpus they are used in. If this is the case, it could be that differences or similarities in the use of the Perfect in questions is due to the corpora and not other phenomena. To get the number of questions without a Perfect in the Harry Potter Corpus (HP), all question marks in dialogue environments were counted. From this number the amount of questions with a Perfect was extracted.

	HP-questions	SB-questions	Total
#Perfects	32	585	617
#not Perfects	644	9079	9723
Total	676	9664	10340

Table 5: Amount of questions with and without Perfects in both corpora

I found that there is no significant association between the use of the Perfect in questions and the corpus they are in ($\chi^2(1) = 1.958, p \approx 0.161$), meaning that the use of Perfect of questions is not dependant on the corpus we are looking at. The effect is also very small ($\phi = 0.014$).

The amount of Present Perfects in questions compared to the amount of questions in both corpora is similar. In the Harry Potter Corpus 4.7% of the questions contain a Perfect and in the Switchboard Corpus this is 6.1%. These percentages are very close.

Now I am going to zoom in on each specific question act from the SWBD-DAMSL labels. I created table 6 with all question acts found in the Switchboard Corpus. The second column shows how often each act appeared with a Perfect compared to all found questions with a Perfect in Harry Potter. The third column shows the same for the Switchboard Corpus. Important to note is in the data of the Switchboard Corpus in Appendix B, the *Tag-Yes-No-Questions* were counted as *Yes-No-Questions*. For this table I counted the *Tag-Yes-No-Questions* separate from this category.

Dialogue Act	# with Perfect in HP		# with Perfect in SB	
qy^g	4/32	12.5%	18/585	3.1%
qy^d	10/32	31.3%	85/585	14.5%
qy	5/32	15.6%	348/585	59.5%
qw	7/32	21.9%	94/585	16.1%
qh	6/32	18.8%	18/585	3.1%
qo	0/32	0%	12/585	2.1%
qrr	0/32	0%	3/585	0.5%
qw^d	0/32	0%	7/585	1.2%

Table 6: Amount of question-acts with a Perfect in both corpora

This table shows that for some question acts there are big differences between the two corpora. I will now look at some of those differences.

7.1.1 Tag-Questions

The first thing standing out is the use of *Tag-Questions*. In the Switchboard Corpus they are almost never used while in the Harry Potter Corpus 12.5% of the questions with Perfects are *Tag-Questions*. To find out the reason for this difference, I looked at all sentences in Switchboard Corpus that were tagged as a *Tag-Question*. I found that there actually were more *Tag-Questions* containing a Perfect in the Switchboard Corpus than mentioned in my table. But why are they not counted as a *Tag-Question* containing a Perfect? This can be explained with help of example (21). It is part of a conversation extracted from the Switchboard Corpus. Behind each utterance it says with which tense and which Dialogue Act it was annotated.

(21)   A: Yeah, it's really different because that's kind of the, you know,
       that's the chalk hills down there,   (present, `sv`)
       B: Oh, and you've always,   (present, `qy^g`)
       A: lime-, limestone hills,   (none, `qy^g`)
       B: lived in Texas, have you?   (other, `qy^g`)

The question asked by person B in this conversation is tagged as a *Tag-Question*. However, because person A talks also during this sentence, the sentence is split. Because of this, the wrong tense is given to the sentence. The first part of the sentence is tagged as a Present and the second part is tagged as Other, meaning it is not one of the standard tenses.

However, if person A would not have talked at the same time as person B, this question would have been tagged as a Present Perfect. This of course is something that could occur more often than only in this example, which could change the number of *Tag-Questions* with a Perfect as well. It is one of the big differences between the Switchboard Corpus and the Harry Potter Corpus. In the Harry Potter Corpus multiple people talking at the same time is something that rarely happens, because this would make reading the book very hard. However, it is something that happens very often in real life conversations. This is something we have to keep in mind when drawing conclusions about dialogues with help of the Harry Potter Corpus.

7.1.2 Yes-No-Questions

When looking at the *Yes-No-Questions* and the *Declarative Yes-No-Questions* in table 6 we see that the first one is used much more often in the Switchboard Corpus and the second one is used relatively more often in the Harry Potter corpus. When taking a closer look at the sentences with these labels I see that the *Declarative Yes-No-Question* is often used when making assumptions about someone or something. The speaker seems to have an expectation of the answer the other person is going to give. The examples in table 2 from chapter 6.1 show this for the Harry Potter Corpus. This is also clear in example (22) and (23) from the Switchboard Corpus.

(22)   I'm, I'm sure you**'ve** probably **heard** of Marion Berry.   (qy^d)

(23)   I take it you **haven**'t **spent** any time in the military?   (qy^d)

When looking at the regular *Yes-No-Questions* I see that specific answers are not expected as often for these questions as for the declarative ones. Again this is shown for these questions from Harry Potter Corpus in table 2. Examples (24) and (25) show this for the Switchboard Corpus.

(24)   So. **Have** you **heard** about Saturns?   (qy)

(25)   **Have** you **tried** the, uh, the pool at the Spring Creek Fitness Center?   (qy)

The expectation for a specific answer in the *Declarative-Yes-No-Questions* can explain why they are seen more often in the Harry Potter Corpus than in the Switchboard Corpus. In this corpus the characters very often know each other when using this act, and thus can make better expectations of the other persons knowledge. Remember the Switchboard Corpus consists of conversations between people that have not talked before. For these people it is harder to make assumptions about the other persons knowledge and thus about the answer he or she is going to give.

On the other hand regular *Yes-No-Questions* are probably used relatively more often in the Switchboard Corpus because the speakers are just getting to know each other. In these situation you ask questions because you do not know the answer yet and you want to find out more about the other person. In the Harry Potter Corpus situations like this are less likely to occur.

7.1.3 Rhetorical-Questions

Another thing that stands out in table 6 is that *Rhetorical-Questions* are a much bigger portion of the questions in Harry Potter than in the Switchboard Corpus. It is not easy to find a reason for this fact. As I mentioned in chapter 6.1 the *Rhetorical Questions* are meant to be unanswered. When taking a closer look at the examples in table 2 these questions seem to be used to strengthen a message the speaker wants to communicate to convince the other person of something. When looking at *Rhetorical-Questions* containing a Perfect from the Switchboard Corpus, I see this act is also used to strengthen the speakers message. Example (26) and (27) show this. In (26) the speaker wants to emphasize he or she has gone many places. In (27) the speaker wants to emphasize how hard it is to find a cure for cancer and that scientists have been trying to find it for a very long time.

(26)   Oh, g-, where **have** I **gone** that's anywhere.   (qh)

(27)   How many years **have** we **tried** to find a cure for cancer.   (qh)

A similar reasoning as described above for the *(Declarative) Yes-No-Questions* can explain why this question act is used with a Perfect relatively more in the Harry Potter Corpus. In the Switchboard Corpus the persons do not know each other. They are forced to talk to each other about a subject. Their goal is not to convince the other person of something, the goal is just to talk so the Switchboard Corpus gets its data. Since the Harry Potter Corpus is a story and the characters know each other, they probably want to convince each other to do or not do something or to have the same opinion. This could explain why the *Rhetorical-Question* is more likely in the Harry Potter Corpus.

7.2 Statements

As mentioned before, it is harder to draw conclusions from the Harry Potter Corpus about statements. This is because only the dialogue sentences containing a Perfect are annotated with Dialogue Acts and thus I can not compare the amount of statements with a Perfect with the overall appearance of statements. What I can do is compare the odds of a sentence in the Perfect tense being a statement from both corpora. As shown in table 3 from chapter 6.2 the Harry Potter Corpus contained 229 Perfects. From these Perfects 196 were annotated *Statement-non-opinion* and 1 was *Statement-opinion*. This means 86.0% of all sentences containing a Perfect were statements. For the Switchboard Corpus this was 81.5%. These numbers are quite close to each other, meaning there is a similar distribution of statements in utterances with a Perfect tense.

What stands out however is that in the Switchboard Corpus a much bigger portion if these statements carry an opinion and thus are annotated *Statement-opinion*. Of all 5434 statements with a Perfect in the Switchboard Corpus, 852 carried an opinion, so 15.7%. In the Harry Potter Corpus only 1 out of 197 statements with a Perfect was a *Statement-opinion*, so 0.5%. There is only one reason I can think of which I will describe now. For this it is important to remember the Reichenbach system described in chapter 3.1. It stated that Present Perfect is used to describe an event that happened in the past and the result of that event in the present. Because Harry Potter is a story, the persons talk about things that happen in the moment. When expressing opinions about things, they are probably about things happening at that moment. Thus I think *Statement-opinion* acts are more often used with a Present tense in the Harry Potter Corpus. In the Switchboard Corpus people talk on the phone. They do not speak of things that are happening at that moment but more about things that happened and what they think of it at this moment, so the opinions are the result in the present time of the past event. This is the reason we see more *Statement-opinions* with a Perfect in the Switchboard Corpus than in the Harry Potter Corpus.

7.3 Topic negotiation and Conventional-Closing-acts

As we saw in chapter 3.5 the Switchboard Corpus contains a lot of questions introducing new topics and also a big portion of *Conventional-Closing*-acts. I will now discuss whether this is the same in the Harry Potter Corpus.

When looking at the examples from the Harry Potter Corpus of various question acts in table 2 of chapter 6.1, we see that most of these questions are used to introduce new topics. To repeat some of them:

(28)   You**'ve met** Malfoy before?   (qy^d)

(29)   **Haven**'t you **heard** what is was like when he was trying to take over?   (qy)

(30)   What **have** we **got** today?   (qw)

In (28) the speaker wants to talk about a meeting with Malfoy. In (29) the speaker wants to talk about what it was like when 'he' was trying to take over. And in (30) the speaker wants to talk about the schedule for the day. When counting all questions with a Perfect introducing a new topic, I found 19. This means 59.4% of all questions containing a Perfect introduced a new topic. So indeed in the Harry Potter Corpus we see a big portion of the questions with a Perfect are used for topic negotiation, like in the Switchboard Corpus.

However, there are no *Conventional-Closing*-acts found in the Harry Potter Corpus. Remember these were used to close off a topic. When looking at some examples of this act with a Perfect from the Switchboard Corpus presented in the poster of Tellings, van der Klis, Le Bruyn and de Swart (2019), we see that these acts are often used to close off a conversation. In (31), (32) and (33) I listed a few of the examples.

(31) It**'s been** pleasant talking to you. *(fc)*

(32) Well, it **has been** really fun. *(fc)*

(33) Well I**'ve** really **enjoyed** the conversation. *(fc)*

When seeing these examples it makes sense that this act is used more often in the Switchboard Corpus than in the Harry Potter Corpus. This is because the phone conversations in this Corpus have to be closed off, which is done with the *Conventional-Closing*-act. This is not often done in the Harry Potter Corpus since the characters have 'face-to-face' conversations. These sorts of conversations are not ended the same way as phone conversations. Furthermore conversation endings are maybe left out in the Harry Potter Corpus often, because they are not important for the story.

7.4 Agreement, backchannel and hedge acts

As shown in table 3 from chapter 6.2 there are no agreement, backchannel, or hedge acts in the Harry Potter Corpus with a Perfect. In table 7 below I listed how often those acts appeared with a Perfect in the Switchboard Corpus compared to all Perfects found.

Act	# with Perfect in SB	
Agreement	18/6666	0.3%
Backchannel	271/6666	4.1%
Hedge	8/6666	0.1%

Table 7: Discourse markers with Perfects in Switchboard

As this table shows, the agreement and hedge acts almost never appeared with a Perfect in the Switchboard Corpus. These acts are thus distributed similarly over the Perfects as in the Harry Potter Corpus. The appearance of the backchannel act was a little higher in the Switchboard Corpus. I think this again is due to the nature of both corpora. Since the Switchboard Corpus consists of phone conversations, people have to use speech to let the other person know they understand what is said. In Harry Potter this can be done by expressions or movements, since the dialogues in this book are most often when the characters can see each other. Furthermore, it would probably slow down the reading when these acts are constantly used, so I think they are consciously left out in books.

Now I have discussed several Dialogue Act categories that are or are not found in the Harry Potter Corpus, I will go back to my hypotheses and conclude whether they were right or not in the next chapter.

8 Conclusion

In this chapter I will go back to my hypotheses and draw conclusions about them. With these conclusions I will provide an answer to my research question:

- Is the Perfect tense use in dialogues written in *Harry Potter and the Philosopher's Stone* comparable with the Perfect tense use in spontaneous speech?

Furthermore I will discuss the problems in my research and I will do suggestions for further research.

8.1 Hypotheses

First I need to go back to my subhypotheses to be able to make a conclusion about my main and null hypothesis. The first subhypothesis I formed in chapter 4 is stated below.

S1. When annotating sentences in dialogues containing a Perfect, questions introducing new topics and conventional-closing acts will be found.

As we saw in chapter 7.3, this hypothesis turned out to be partly supported. I found 59.4% of the questions with a Perfect were used to introduce a new topic in Harry Potter. So indeed a lot of questions with a Perfect introducing new topics were found. This supports the theory of Nishiyama and Koenig (2010) described in chapter 3.4 about the use of a Perfect for topic negotiation.

However, there are no *Conventional-Closing*-acts found. This act is used to close off a topic and its use would support Portner's theory (2003) that states a Perfect is used to answer a question under discussion. Not finding this act does not contradict Portner's theory though. As described in chapter 7.3 the absence of this act in the Harry Potter Corpus can be due to the sorts of conversations that the characters have, which are very different than dialogues in the Switchboard Corpus. Furthermore, it could be that endings of conversations are left out in Harry Potter because they are not important for the story.

Now I go on to the second subhypothesis.

S2. A very small amount of Perfects in agreement, backchannel and hedge acts will be found.

This hypothesis is supported in chapter 7.4, we see that none of these acts were found with a Perfect in the Harry Potter Corpus. The agreement and hedge acts were also almost never found in the Switchboard Corpus with a Perfect. However, the backchannel act appeared a little more often in this corpus. But as explained in chapter 7.4 this can be due to the fact that the Switchboard Corpus consists of phone conversations. In Harry Potter backchanneling is probably done with expressions and movements more often. So since this hypothesis is not contradicted, it still stands.

I will now go on to the last subhypothesis.

S3. There will be a higher chance on use of a Perfect in questions compared to statements.

This is a hypothesis I am not able to support or contradict. Since there is no contradiction for this hypothesis, it still stands and can be used for further research. To be able to say something about it, we would need to know the odds of a statement containing a Perfect in the dialogues of Harry Potter. For that we need to know how many sentences in the dialogues of Harry Potter were statements, regardless of the tense these sentences are written in. Right now I only have this number for sentences written in the Perfect tense.

Now I go on to the main and null hypothesis.

H0. There is a similar Perfect distribution on the Dialogue Acts in the Harry Potter book as in the Switchboard Corpus.

As we saw, I am not able to contradict this hypothesis with the data available, therefore it is still presumed to be true. In order to say more about this hypothesis all sentences in dialogues of the Harry Potter Corpus have to be annotated. But, since my research provides no contradiction for the hypothesis, I expect it to be confirmed when this data is available. I can however show whether there is a similar Perfect distribution over the question act in

both corpora, because for this act I do not need a full annotation, it sufficed to count all question marks. I saw a similar distribution of Perfects over the question act in the Harry Potter book compared to the Switchboard corpus. Of all questions in Harry Potter 4.7% contained a Perfect and for the Switchboard Corpus this number was 6.1%. Since the Perfect distribution over the question act is similar in both corpora, my believe that the hypothesis will still stand when we know the Perfect distribution over all Dialogue Acts in Harry Potter, is strengthened.

Furthermore, I can say something about the distribution of Dialogue Acts on the Perfects in both corpora. This tells us whether the Perfect is used for similar acts in the Harry Potter book and the Switchboard Corpus. I created table 8 below as an overview of chapter 7. It shows the distribution of the acts presented in the poster of Tellings, van der Klis, Le Bruyn and de Swart (2019) on the Perfects in the Harry Potter Corpus and the Switchboard Corpus.

Act	HP	SB
Statement	86.0%	81.5%
Agreement	0%	0.3%
Backchannel	0%	4.1%
Hedge	0%	0.1%
Questions	14.0%	8.8%

Table 8: The odds of an utterance with a Perfect being a certain act in both corpora

This shows a very similar distribution in both corpora. We saw however that for the smaller SWBD-DAMSL-labels, like the *Tag-Questions*, *Yes-No-Questions*, *Rhetorical-Questions* and *Statement-opinions*, there are bigger differences in distributions over the Perfect in both corpora. But as we also saw in chapter 7, this can be due to the fact that both corpora exist of very different sorts of dialogues.

8.2   Research Question

I now go back to my research question:

- Is the Perfect tense use in dialogues written in *Harry Potter and the Philosopher's Stone* comparable with the Perfect tense use in spontaneous speech?

With the data obtained for this research I am not able to say whether there is a similar Perfect tense distribution between *Harry Potter and the Philosopher's Stone* and spontaneous speech. I can say that in both corpora the Perfect is used in similar acts, mainly questions and statements. My research has shown that the Harry Potter Corpus and the Switchboard Corpus are two very different corpora. However, since my null hypothesis is not contradicted, it is not denied that the dialogues in the Harry Potter book have similar Perfect tense use as spontaneous speech. I thus think the Harry Potter Corpus can be used as a corpus to look into the Perfect use in spontaneous speech. Nonetheless, it is important to remember that it is written in a way that makes it easy and exciting to read. So discourse markers are not used often and sentences are a lot less chaotic than in real life.

8.3   Further Research

To be sure my null hypothesis is not contradicted for all Dialogue Acts, it is necessary to annotate all sentences in dialogue from the Harry Potter Corpus. This would also strengthen the answer to my research question. Full Dialogue Act Annotation is something that could be done in future research. Since annotating takes up a lot of time, I was not able to do this. Another interesting area for future research is looking at the distribution of Perfects on Dialogue Acts in different languages to see whether there is cross-linguistic variation here. To save a lot of time in the annotating process it would also be interesting to see if the system can be automated. This way much larger datasets can be annotated and thus more data can be compared.

Lastly it would be interesting to see whether Dialogue Act Annotation can contribute to bettering translation programs. When sentences are assigned a Dialogue Act, it might be easier to find correct translations. An English sentence like example (5) in chapter 3.2

with a Simple Past can maybe be translated to Dutch sentence (6) with a Perfect when the sentence has been given a Dialogue Act and to another translation with a different tense if that feels more natural for that act.

9 References

1. Boogaart, R. J. U. (1999). *Aspect and temporal ordering: A constrastive analysis of Dutch and English.* Doctoral dissertation, Free University, Amsterdam.

2. Jurafsku, D., Shriberg, L., Biasca, D. (1997, August 1). *Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, Draft 13.* Retrieved from https://web.stanford.edu/ jurafsky/ws97/manual.august1.html.

3. Kamp, H. & Reyle, U. (1993). *From discourse to logic*, part 1, 2. Kluwer. Dordrecht.

4. Klis, M. van der, Le Bruyn B. & Swart, H. de (2015). Extracting Present Perfects from a multilingual corpus. Paper presented at CLIN26, Amsterdam, The Netherlands, December 18, 2015.

5. Le Bruyn, B., Klis, M. van der & Swart, H. de (2019). The Perfect in dialogue: evidence from Dutch. *Linguistics in the Netherlands, 36*(1), 162-175.

6. Levshina, N. (2017). Online film subtitles as a corpus: an n-gram approach. *Corpora, 12*(3), 311-338.

7. Nishiyama, A., & Koenig, J. P. (2010). What is a perfect state? *Language*, 611-646.

8. Portner, P. (2003). The (temporal) semantics and (modal) pragmatics of the perfect. *Linguistics and Philosophy, 26*(4), 459-510.

9. Reichenbach, H. (1947). *Elements of Symbolic Logic.* The Free Press/Collier–Macmillan. New York/London.

10. Russell, S.J. & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach.* Pearson Education. Upper Saddle River, New Jersey.

11. Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Ess-Dykema, C. van & Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics, 26*(3), 339-373.

12. Swart, H. de (2007). A cross-linguistic discourse analysis of the Perfect. *Journal of Pragmatics, 39*(12), 2273-2307.

13. Swart, H. de (2018) Tense use in discourse and dialogue: a parallel corpus approach. Paper presented at Workshop Ways of Reference in Romance Languages, Köln, Germany, 2018.

14. Tellings, J., van der Klis, M., Le Bruyn, B., & de Swart, H. (2019). Tense use in dialogue. Poster presented at SemDial 2019, London, England, 2019.

Appendix A   List of Dialogue Acts in the Switchboard Corpus

SWBD-DAMSL	SWBD	Example	Cnt	%
Statement-non-opinion	sd	*Me, I'm in the legal department.*	72,824	36%
Acknowledge (Backchannel)	b	*Uh-huh.*	37,096	19%
Statement-opinion	sv	*I think it's great*	25,197	13%
Agree/Accept	aa	*That's exactly it.*	10,820	5%
Abandoned or Turn-Exit	% -	*So, -*	10,569	5\%
Appreciation	ba	*I can imagine.*	4,633	2%
Yes-No-Question	qy	*Do you have to have any special training?*	4,624	2%
Non-verbal	x	*[Laughter], [Throat_clearing]*	3,548	2%
Yes answers	ny	*Yes.*	2,934	1%
Conventional-closing	fc	*Well, it's been nice talking to you.*	2,486	1%
Uninterpretable	%	*But, uh, yeah*	2,158	1\%
Wh-Question	qw	*Well, how old are you?*	1,911	1%
No answers	nn	*No.*	1,340	1%
Response Acknowledgement	bk	*Oh, okay.*	1,277	1%
Hedge	h	*I don't know if I'm making any sense or not.*	1,182	1%
Declarative Yes-No-Question	qy^d	*So you can afford to get a house?*	1,174	1%
Other	o,fo,bc,by,fw	*Well give me a break, you know.*	1,074	1%
Backchannel in question form	bh	*Is that right?*	1,019	1%
Quotation	^q	*You can't be pregnant and have cats*	934	.5%
Summarize/reformulate	bf	*Oh, you mean you switched schools for the kids.*	919	.5%
Affirmative non-yes answers	na,ny^e	*It is.*	836	.4%
Action-directive	ad	*Why don't you go first*	719	.4%
Collaborative Completion	^2	*Who aren't contributing.*	699	.4%
Repeat-phrase	b^m	*Oh, fajitas*	660	.3%
Open-Question	qo	*How about you?*	632	.3%
Rhetorical-Questions	qh	*Who would steal a newspaper?*	557	.2%
Hold before answer/agreement	^h	*I'm drawing a blank.*	540	.3%
Reject	ar	*Well, no*	338	.2%
Negative non-no answers	ng,nn^e	*Uh, not a whole lot.*	292	.1%
Signal-non-understanding	br	*Excuse me?*	288	.1%
Other answers	no	*I don't know*	279	.1%
Conventional-opening	fp	*How are you?*	220	.1%
Or-Clause	qrr	*or is it more of a company?*	207	.1%
Dispreferred answers	arp,nd	*Well, not so much that.*	205	.1%
3rd-party-talk	t3	*My goodness, Diane, get down from there.*	115	.1%
Offers, Options Commits	oo,cc,co	*I'll have to check that out*	109	.1%
Self-talk	t1	*What's the word I'm looking for*	102	.1%
Downplayer	bd	*That's all right.*	100	.1%
Maybe/Accept-part	aap/am	*Something like that*	98	<.1%
Tag-Question	^g	*Right?*	93	<.1%
Declarative Wh-Question	qw^d	*You are what kind of buff?*	80	<.1%
Apology	fa	*I'm sorry.*	76	<.1%
Thanking	ft	*Hey thanks a lot*	67	<.1%

Appendix B Extensive Dialogue Act data from the Switchboard Corpus

Rijlabels	gerund	infinitive	modal	none	other	participle	past	present	present perfect	present perfect continuous	Eindtotaal
%	61	136	298	10219	336	17	553	4417	80	3	16120
^2	35	110	47	499	60	11	50	236	5		1053
^g			1	58	1		4	32	1		97
^h	1	202	26	110	33		35	133	13		553
^q	2	56	121	48	142	1	52	491	18		931
aa	28	91	147	8133	201	5	112	3127	20	2	11866
aap_am	2	3	14	36	8	1	4	47	1		116
ad		255	80	45	103		17	227	7		734
ar		1	10	258	7		12	62			350
arp_nd		2	5	49	24		29	89	10	1	209
b	250	523	375	40022	1283	83	849	3825	227	15	47452
b^m	9	22	5	597	10		14	55	1	1	714
ba	30	55	189	2376	119	5	126	2476	20	4	5400
bd	1	1	1	34	1		3	64			105
bf	24	58	54	245	97	3	105	421	16	2	1025
bh	3	6	11	616	15	1	87	411	1		1151
bk	2	6	7	1206	19	4	20	171	5		1440
br	1	32	3	166	10	1	13	77	1		304
fa		9		17	4		3	44	1		78
fc	91	191	67	1489	92	2	209	304	94	43	2582
fo_o_fw_"_by_bc	4	20	13	687	9	2	14	133	4		886
fp	6	1		155	1			60			223
ft	5	30		28	2			8			73
h		9	25	62	46		8	1047	8		1205
na	4	5	27	354	43	2	74	321	29		859
ng	2	1	10	103	14		24	115	34	1	304
nn		5	1	1365	3		2	9			1385
no		3	14	32	23		16	197	8		293
ny	3	6	3	2985	5	1	5	43	5		3056
oo_co_cc		33	46	4	13		1	17			114
qh	1	12	33	38	127		28	316	18		573
qo	2	3	7	233	41		27	326	12		651
qrr	1		8	13	26	1	23	136	3	1	212
qw	7	7	54	235	160	6	238	1207	94	8	2016
qw^d		1	1	7	12		10	43	7		81
qy	12	35	102	416	515	15	573	2760	365	18	4811
qy^d	11	17	57	224	152	3	139	534	85	1	1223
sd	140	287	2852	4561	14636	100	13344	32682	4582	344	73528
sv	59	201	1300	1521	5997	18	1041	14814	852	51	25854
t1	1	2	2	14	8		24	47	5		103
t3	1	15	8	42	5		6	44			121
x	28	90	68	3845	165	10	113	486	34	5	4844
Eindtotaal	827	2542	6092	83147	24568	292	18007	72054	6666	500	214695