

**UNIVERSITY OF TARTU
DEPARTMENT OF ENGLISH STUDIES**

**THE ANALYSIS OF CANONICAL AND NON-CANONICAL QUESTIONS
IN AN ENGLISH LANGUAGE PODCAST
BA thesis**

**KÄRT ROOMÄE
SUPERVISOR: *Lect.* JANE KLAVAN, PhD**

**TARTU
2019**

ABSTRACT

This bachelor's thesis studies direct questions in "Grammar Day," an episode of a linguistics podcast *Talk the Talk*. The aim is to analyze the formulation and function of canonical and non-canonical direct questions in natural oral discourse. The approach used is similar to that of conversation analysis and, hence, the analysis does not proceed from any specific hypotheses. Instead, the thesis takes a data-driven approach.

The thesis consists of five sections: the introduction, the section comprising the literature review, an analysis of direct questions in the annotated transcript of the podcast episode, the conclusion, and the list of references. The introduction highlights the importance of examining canonical and non-canonical questions as well as the reasons for choosing podcast as a source for compiling the corpus of this study.

The theoretical part contains three sections. Section 1.1 introduces the categorization of questions into two large categories: canonical and non-canonical questions. Section 1.2 provides the reader with some of the subdivisions of non-canonical questions used by different authors. Section 1.3 will present the categorization of non-canonical questions used in this study and give reasons for deciding on this kind of division.

The empirical part consists of an analysis of direct questions in "Grammar Day." Section 2.1 explains why podcast was chosen as the source material, Section 2.2 introduces the process of narrowing down the podcasts and their topics. Section 2.3 gives an overview of transcribing "Grammar Day" by using a semi-automatic transcription tool Temi and annotating the episode with a modified version of the Jeffersonian Transcription System. Section 2.4 studies the frequency of questions and their division to different categories. The last, Section 2.5, gives examples for all categories and discusses the results.

The thesis ends with a conclusion where the most crucial aspects of the previous sections are highlighted. The list of references gives a comprehensive overview of all the primary and secondary sources mentioned in the thesis. The thesis also has 4 Appendixes. Appendix 1 introduces the similarities and differences present in 4 different semi-automatic transcription tools via examples. Appendix 2 further compares features present in each of the tools. Appendix 3 contains 5 sample pages from the annotated transcript of "Grammar Day," wherein direct questions have been highlighted, and Section 4 lists all direct questions present in the podcast episode, category by category.

TABLE OF CONTENTS

ABSTRACT	2
INTRODUCTION	4
1 CANONICAL AND NON-CANONICAL QUESTIONS	6
1.1 Typology of Questions.....	6
1.2 Subtypes of Non-canonical Questions.....	9
1.3 Categorization of Non-canonical Questions Used in This Study	12
2 ANALYSIS OF DIRECT QUESTIONS IN “GRAMMAR DAY”	15
2.1 Podcasts as a Source of Linguistic Data	15
2.2 Choosing the Podcast Episode	17
2.3 Transcribing the Text.....	19
2.4 Results.....	24
2.5 Discussion.....	28
CONCLUSION	34
LIST OF REFERENCES	37
Primary Sources	37
Secondary Sources	37
Appendix 1: Approximately one-minute long extracts of the transcripts.	39
Appendix 2: Comparison of websites offering (semi)automatic transcription services.....	41
Appendix 3: Extracts from the annotated transcript of “Grammar Day.”	42
Appendix 4: Direct questions in “Grammar Day.”.....	48
RESÜMEE	53
Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks	54
Autorsuse kinnitus	55

INTRODUCTION

Questions form a considerable part of our everyday conversations. The aim of this study is to see how questions are formulated in natural oral discourse and what the function of the questions is. For that reason, “Grammar Day,” an episode of a linguistics-themed podcast *Talk the Talk* was chosen as the source material for compiling the corpus. Physical paralinguistic features (such as gestures and posture), which are no doubt important for human communication purposes, are excluded from the present study.

Podcast is a 21st-century phenomenon, making it very topical for modern society. The term *podcast* itself is a blend of two words, *iPod* and *broadcast*, and it refers to a digital audio recording meant for offline listening, which is either downloaded or streamed (Chandler and Munday 2016).

Linguists from different subfields can study questions, and the present thesis will focus on three aspects: the syntax, semantics, and pragmatics of questions. The properties of each question present in the corpus of this study will be evaluated on the basis of one or more of these three aspects, depending on the particular case.

Linguists have so far paid more attention to other media that often combine audio and visual features, even though there have been several in-depth studies on questions in (semi)oral discourse, such as news broadcasts and talk shows. Podcasts, given their relatively short existence, have not been dealt with to that extent. The topic is important both in relation to filling a gap in the field of theoretical linguistics and providing ample material for a detailed, but not overly large-scale study, giving an opportunity to get acquainted with the methodology used by conversation analysts.

This study intends to investigate how native speakers of English formulate questions in natural oral discourse and what kind of function the questions serve. No hypotheses were formulated beforehand, because the conversation analytic approach adopted here is data-

driven. The primary source was the podcast episode “Grammar Day,” the transcription of which required acquiring skills of manually annotating transcripts, with the help of an online semi-automatic transcription tool Temi, and the Jeffersonian Transcription System. Several handbooks and monographs were used in order to get acquainted with the relevant terminology and the work done before.

This thesis is divided into two chapters: the first, theoretical one, is composed of the following sections: 1) Canonical and/or Non-canonical Questions, and 2) Categorization of Non-Canonical Questions. The second, empirical one, has five sections: 1) Podcasts as a Source of Linguistic Data; 2) Choosing the Podcast Episode; 3) Transcribing the Text; 4) Results, and 5) Discussion. The goal of the second chapter is the analysis of direct questions in “Grammar Day,” while the first chapter intends to provide the reader with the contextual basis required for comprehending the analysis of the podcast episode as well as the categories formulated.

1 CANONICAL AND NON-CANONICAL QUESTIONS

Questions play a key role in interpersonal communication and as a subject for linguistic research, they provide numerous opportunities. Asking the so-called right questions may save a lot of time and make conversations considerably more efficient. It is possible to analyze questions from different perspectives of various subfields of theoretical linguistics, but in the present thesis, syntactic, semantic, and pragmatic aspects of questions are in the center of attention. In this chapter, the division into two large categories, *canonical* and *non-canonical questions* (henceforth, CAQ and NONCAQ), following the example of Dayal (2016) and Kania (2016), will be described. After that, different subtypes of non-canonical questions, given that they are the focus of this thesis, will be explained alongside with the categorization developed for this study, using Kania's (2016) and Dayal's (2016) categorizations as an example.

1.1 Typology of Questions

In order to understand the difference between the two types of questions, one must consider their function as well as syntax. CAQ are used to obtain new information, while NONCAQ function differently. NONCAQ can have many purposes, some of which will be introduced later in the chapter, but the main characteristic is that their primary goal is not to seek information. Their syntactic form may also deviate from the traditional interrogative syntax, which in the case of analyzing natural discourse may additionally be explained by the relatively short turns taken by the speakers, i.e., for how long one speaker talks at a time.

Dayal, in her book *Questions* (2016), has approachably explained how the CAQ are distinct from the NONCAQ. In the following, her comprehensive treatment of the subject is used as the main source, given that it is likely the first concise one-volume exploration of questions. Further, her coverage is accessible even without extensive background

knowledge. Dayal (2016: 1-2) first reminds the reader of the traditional properties of questions, which most of us have learned in English grammar classes. These include inverted subject and auxiliary as well as positioning the interrogative word at the beginning of the sentence, in addition to rising intonation. The CAQ also require that the one asking the question is unaware of the answer. However, the formal conditions are often not met when it comes to natural discourse, which is analyzed in this study. Thus, it is essentially a question of “the relation between the interrogative form and the speech act of questioning” (Dayal 2016: 2).

In order to further clarify what exactly is meant by a speech act of questioning, three concise criteria proposed by Dayal (2016: 4) are quoted here:

- 1) S does not know the truth about p.¹
- 2) S wants to know the truth about p.
- 3) S believes H knows the truth about p.

Dayal (2016: 5) claims that the criteria of the speech act of questioning are universal, and that deviation from these conditions gives the interrogative forms a specialized character. While the goal of the interrogative forms with a specialized character is not getting the truth about the proposition, they can still be semantically treated as questions. Dayal (2016: 268) has also listed the criteria for the NONCAQ, one or more applicable at a time:

- 1) They elicit information but also convey a bias about what that information is likely to be;
- 2) Their syntactic form does not conform to the standard interrogative form;
- 3) Their purpose is to engage in some other kind of speech act instead of, or in addition to, eliciting information.

¹ S – Speaker, p – proposition, H – Hearer.

While the first two properties of the NONCAQ are quite digestible – compare, for example, questions ‘The exam isn’t tomorrow?’ and ‘The exam tomorrow?’ – the third one likely requires further explanation.

The speech act theory has been discussed thoroughly by the British philosopher of language Austin in his book, *How to Do Things with Words* (1962) and by the American philosopher Searle in *Expression and Meaning: Studies in the Theory of Speech Acts* (1979). Kania (2016: 74) and Dayal (2016: 4) also touch upon the same topic, connecting speech acts with the analysis of questions. Both claim that questions not functioning as information-seekers, or those fulfilling some additional goal, can be defined as *indirect speech acts*. According to Searle (1979: 30-31), one sentence can be semantically multi-layered: one meaning is heard and the other understood. Searle’s (1979: 31) example, an ordinary question, ‘Can you reach the salt?’ indicates the presupposed ability of the addressee as well as a request, and therefore it deviates from the speech act of questioning. One must thus pay attention to discourse, i.e., turns taken before and after the question in order to understand its exact function in the specific context.

These abovementioned criteria and restrictions were all taken into account when deciding whether the questions analyzed in the present study should be allocated to the category of CAQ or NONCAQ. However, in several cases, an issue arose concerning the overlap between syntax and pragmatics. This shows again that even though different subfields of theoretical linguistics study distinct phenomena, they are still intrinsically linked. Within the scope of this thesis, for example, it appeared that while a question could have an interrogative syntax, this did not necessarily mean that it functioned as a CAQ. Polar and rhetorical as well as ability questions in particular posed problems. Specific examples will be given in the empirical part of this thesis, alongside with the discourse context required.

1.2 Subtypes of Non-canonical Questions

NONCAQ serves as an umbrella term for various types of questions that fall under the same broad category either syntactically and/or pragmatically. As became apparent over the course of writing the present thesis, it was at times quite difficult to assign them to specific subtypes. Nevertheless, before a closer study of the subdivisions compiled by different authors can be conducted, Dayal (2016: 268) has highlighted two characteristics that the NONCAQ have in common:

- 1) It is often possible to opt for a potentially simpler form distinct from the non-canonical one. The choice made by the speaker in favor of the NONCAQ form indicates that the purpose of the question is different.
- 2) Embedded clauses, i.e., subordinate clauses placed in the middle of the matrix clause, cannot generally be used with the NONCAQ.

Regarding the criteria mentioned in the discussion above, the categorization of non-canonical yes-no-questions by Kania (2016) will be the first to be introduced. While her treatment focuses on the acquisition and use of yes-no questions, Kania's (2016) division was used when sorting the questions analyzed in this study into classes. She differentiates between non-inverted yes-no questions, wherein subject comes before auxiliary (*You can't see it?*) (Kania 2016: 63); questions, where predicate is present, but other elements, i.e., auxiliary and/or subject, have been omitted (*<Do> You want an apple?*) (Kania 2016: 63); and fragmented questions, referring to one-word questions (*Now?*), phrases without a predicate (*The guy from the shop?*), and subordinate clauses (*Because he hit on you?*). The last three categories are illustrated in Kania (2016: 64). In her study of the British National Corpus (henceforth, BNC), Kania's (2016) classification goes further. Considering the question types of interest for this thesis, wh-questions have been divided into 1) full; 2)

auxiliaryless; 3) echo-, clarification, and quiz-questions, and 4) one-word and ‘oblique’ wh-questions (Kania 2016: 93-94). Subcategories of tag questions comprise full, short, and invariant tags (Kania 2016: 95-96).

To be more precise, Kania (2016) conducted a study which dealt with yes-no questions in adult-to-adult speech. Kania (2016: 86) points out that there is a lack of quantitative research based on corpora when it comes to non-canonical questions used in conversations between adults. Kania’s (2016) methodology serves as the basis for the methodology used in this thesis. As referred to in Kania (2016: 87), the corpus study consisted of the following steps:

- 1) Summary of the most relevant information on the (sub)corpus used and discussion of the criteria for extracting data;
- 2) Results about the general frequency of questions;
- 3) Coding and analysis of different question types found in the broader categories, concentrating on NONCAQ yes-no questions in particular.

The data for Kania’s (2016) study were taken from the BNC. Kania (2016: 87) specifies that the reason for selecting this corpus of both spoken and written English (British variety) dating back to the early 1990s was its size and accessibility as well as the detailed categories of the texts of that corpus. It is also easy to export data to MS Excel, among other applications. Kania (2016: 88) created a subcorpus, into which face-to-face unscripted spontaneous conversations were collected. Then all interrogative sentences were counted, the criteria being that the utterance must end with a question mark, which included wh-, alternative, tag-, and yes-no questions.

Additionally, the frequency of questions as compared to all utterances was mentioned (Kania 2016: 90). After that, “a random selection of 5000 hits was exported to MS Excel and then coded by hand” (Kania 2016: 91), helping to avoid systematic biases. Then the reason

for excluding certain utterances was explained, which left 4665 observations that were subsequently divided into four categories: 1) wh-question; 2) tag-question; 3) alternative question, and 4) yes-no question (Kania 2016: 92).

Subcategories		Example	Results from BNC (n = 1729)	Estigarribia (2007, pp. 33-35)
Canonical			948 (55%)	
	DO (untensed)	Do you want some?	167 (10%)	AUXF
	DO (tensed)	Did you buy any?	139 (8%)	
	MODAL/HAVE/BE	Have you seen him?	368 (21%)	
	copula BE	Is he a writer?	274 (16%)	
Noncanonical			777 (45%)	
uninverted	Strong assertive		156 (10%)	
	w/medial DO	You do realize that?	15 (1%)	SERTAUX
	w/MODAL./HAVE/BE	You can see him?	50 (3%)	
	w/BE (copula)	You are a writer?	45 (3%)	
	short uninverted	You did?	15 (1%)	
	without medial aux.	You saw him?	31 (2%)	SERT
	reduced		247 (14%)	
	WSERT	You want some?	95 (5%)	WSERT
	SPRED	They gone already?	56 (3%)	SPRED
	APRED	Don't like this?	13 (1%)	APRED
	PRED	Want some?	83 (5%)	PRED
	fragments		312 (18%)	FRAG
	NP	A man?	288 (17%)	
	PP	Under the table?	11 (1%)	
	other one-word	Now?	9 (0%)	
	other multiword	Seriously, though?	4 (0%)	
other	subordinate	Because of him?	66 (4%)	-
	errors	No like it?	n.a.	ERRAUX

Figure 1. The distribution of question types in Kania's (2016: 86-102) corpus study (original title: Table 4.2: Coding scheme and results for yes-no questions in the BNC sample; from Kania 2016: 99)

A pie chart featuring all four categories was included in the study, and then all but one category, alternative questions, which made up only around 1%, were further subcategorized (Kania 2016: 93). For specific subcategories of NONCAQ, see Chapter 1, Section 1.2 in Kania (2016). All subcategories were depicted as pie charts to better illustrate the results. Examples were given from the BNC sample and from other studies. Since yes-

no questions had the largest number of subcategories and the book focuses on that type, the results have been shown in greater detail (Kania 2016: 98). Namely, in Figure 1, a table can be seen where (sub)categories with examples and numbers in the BNC sample have been summarized. These results are important in the context of the present thesis, because they will be discussed in comparison of the results of the present study in the second part of the thesis. Kania (2016) also explains the more complex cases and finally, discusses the results.

Dayal (2016: 268) has mentioned specific NONCAQ categories in the chapter on non-canonical questions, focusing on more than just yes-no questions. Her list includes 1) negative polar and declarative questions (*Isn't it raining?*) (Dayal 2016: 271); 2) echo questions (*It is raining?*); 3) rhetorical, ability, and inclination questions (*Would you mind telling me if it rains?*), and 4) tag questions (*It's raining, isn't it?*). She admits that the list is not all-inclusive. Dayal (2016: 270) then explains specific properties of NONCAQ through examples of relevant subtypes.

The first above-mentioned category is a request biased either positively or negatively due to the question formulation, while the second is used when what has been said previously needs clarification, being thus dependent on discourse context. Rhetorical, ability, and inclination as well as tag questions all have an obvious answer on the part of the speaker. (Dayal 2016: 282) The constituents of the third above-mentioned category have already been discussed in this section as indirect speech acts. Rhetorical questions can also be interpreted as CAQ in certain contexts, as can be ability and inclination questions, when the answer is not established (Dayal 2016: 283-285). The last category, tag questions, includes both assertive form and conveys a need for acknowledgement/confirmation (Dayal 2016: 288).

1.3 Categorization of Non-canonical Questions Used in This Study

The categorization used in this thesis was compiled after analyzing the results, since the approach taken is data-driven, without any hypotheses formulated beforehand. The

categorization is fully based on the questions in the corpus, which were collected, sorted into two and then categorized further. Nevertheless, Kania's (2016) study serves as an example when it comes to methodology. Because of that, a detailed overview of the study on BNC was given above. The same steps are followed and, as stated previously, this thesis also uses two levels of categorization. First, the questions are grouped into two: CAQ and NONCAQ; after that, subcategories are formed according to the data available for analysis. Kania's (2016) corpus included face-to-face spontaneous conversations, though, and was considerably larger. Additionally, her categories were much more detailed. The present thesis takes a more general approach. For example, tag questions are not further divided into full, short, and invariant tags (Kania 2016: 95-96).

The classification suggested by Dayal (2016) has also been used, so that the final categorization used in this study is a combination of the ones proposed by both Kania (2016) and Dayal (2016). The reason for synthesizing the two is that while some categories overlapped, namely, biased questions (negative polar and declarative) and tag questions, Kania (2016) has gone further and divided elliptical constructions into reduced questions and fragments. Such minuteness is not required for the purposes of the present study. Additionally, Kania's (2016) classification includes four types of wh-questions, but Dayal (2016) has only named echo questions in her list. A more general subcategory was preferred. Moreover, rhetorical, ability, and inclination questions were not the center of attention for Kania (2016), since her focus was on yes-no-questions, but they appeared several times in the corpus of this study, so Dayal's (2016) third category was included here. The final categorization thus comprises five rather broad subtypes, as can be seen in alphabetical order in Table 1. One should take into account that this is just another possible version of formulating subcategories of NONCAQ. There might be other classifications proposed by other researchers.

The five subtypes of NONCAQ used in this study thus include: 1) fragments; 2) rhetorical, ability, and inclination questions; 3) (strong) assertive questions; 4) tag questions, and 5) wh-questions. It has also been indicated that negative polar and declarative questions belong to the third subtype, while echo and one word questions form a part of the fifth subtype. An example sentence, formulated by the author, has been altered according to the constraints imposed by each subcategory.

Table 1. Classification of NONCAQ

Subtype of NONCAQ	Example
Fragmented questions	The exam tomorrow?
Rhetorical, ability, and inclination questions	Can you take the exam tomorrow?
(Strong) assertive questions, including negative polar and declarative questions	The exam isn't tomorrow?
Tag questions	The exam is tomorrow, right?
Wh-questions, including echo and one word questions	What exam you taking tomorrow?

2 ANALYSIS OF DIRECT QUESTIONS IN “GRAMMAR DAY”

The empirical part of the present thesis will focus on the analysis of canonical and non-canonical direct questions in one particular category of broadcast media, *podcast*. This type of digital audio recording was chosen after considering that it usually has audio only. The rising popularity of podcasts serves as another reason. The author also familiarized herself with the work that has been done before on television shows. In the following paragraphs, before a closer examination of a particular podcast episode can be conducted, key terms will be defined and the process of narrowing down the type of the broadcast media to be analyzed will be introduced.

Then, the exact order of procedures taken, in addition to transcription and annotation conventions will be explained, in order to interpret the results of the corpus. According to the methodology used by conversation analysts, all other categories are compiled only after counting the questions, except for CAQ and NONCAQ, since the latter are the focus of this study. Later, NONCAQ are further divided into subcategories and variation among speakers will also be introduced. Results will be illustrated with graphs.

2.1 Podcasts as a Source of Linguistic Data

It was decided that the focus should be on oral media, rather than written records. When speaking, people have less time to think about what they are saying, consider their wording, as well as concentrate on producing grammatically correct, eloquent questions. Thus, the following analysis will shed light on the features of natural discourse, and the patterns forming as a result. Another decision was made quite early on: linguistic, rather than paralinguistic features (volume and pauses, among others) will be at the center of attention. Even though semantic and pragmatic function of questions have also been dealt with, in addition to purely grammatical constructions, rhythm and intonation – two main foci of

prosody, a category of paralinguistics according to Matthews (2014) - are largely not covered here. However, attention will be devoted to more notable cases of the tone of voice and some sounds: transcription includes instances when the podcasters imitated certain animals or changed their tone strikingly, for example sounding very ominous all of a sudden. Additionally, laughter and audible inhaled breaths, as well as other context-specific noises, such as songs featured in the podcast were indicated with comments in double parenthesis. These sounds were included so that the transcription would retain most original features, and could be read as authentically as possible.

The scope of this thesis deals with a subcategory of broadcast media, *podcasts*. According to Chandler and Munday (2016), broadcast media refers to “radio, television, direct-broadcast satellite broadcasting, and webcasting.” At first, the goal was to analyze some type of television show, those which provide ample dialogue and interpersonal communication, such as late-night talk shows, being of greatest interest. After delving into the articles and theses that have been written over the years, it turned out, however, that several in-depth studies have been conducted on the same topic, although not all of those to be mentioned below focused on the English-language programs. As far as talk shows are concerned, Danileiko’s “Formal and functional questions in an American talk show *Late Night with Conan O’Brien*” (2005) as well as a book chapter, “The Role of Questions in Talk Shows” by Schirm (2009), show that the subject has been discussed from a similar angle that will be used in this study. Furthermore, an article on debates, “Interviewers’ challenging questions in British debate interviews” (Emmertsen 2006), and several studies on broadcast news interviews (see, for example, Cohen 1989, Heritage & Roth 1995, Heritage 2002) confirm that other types of television shows have also received attention.

Nevertheless, podcasts have not been received such attention yet in the context of linguistic research. No study focusing specifically on questions in podcasts has been

conducted, according to the sources that were available for the author. The present thesis will therefore provide a small-scale contribution regarding a very new type of entertainment. The history of the term *podcast* can be traced to the beginning of the 21st century. *Podcast* was first coined by Hammersley (2004) in his article “Audible revolution,” which was published in *The Guardian*. The word is a blend that combines elements of two words, *iPod* and *broadcast*. Chandler and Munday (2016) have defined it as “a digital audio recording stored online but (unlike broadcasts) designed for offline use on the user’s computer or mobile device, being either downloaded as a series through an RSS feed or streamed to a media player when needed.” Podcast may or may not include video material and it usually has a specific topic (Stanbrough n.d.: para. 1).

Distributed widely both online and through apps on our smartphones, free of charge or, sometimes, through paid services, podcasts are easily obtained almost everywhere, even when no Internet connection is available. The latter option, however, can usually be employed when the particular episode has been downloaded beforehand. Listening to someone’s voice establishes a closer connection, too. Taking into consideration the aforementioned factors, it is understandable why podcasts, once “a low-concept cottage industry” (Woods 2018: para. 5), are so popular these days, with 6.0 million adults in the UK listening to them on a weekly basis (RAJAR Midas 2018: slide 3). The widespread presence of podcasts in people’s lives is also one of the reasons why they have been used in the present thesis.

2.2 Choosing the Podcast Episode

Opting for podcasts proved to be a virtually inexhaustible source of inspiration. At first, it was surprising just how many different categories there are. To give some examples, according to Castos (2019), a website providing hosting for podcasters, the 13 genres currently available on iTunes include anything from sports to comedy, from business to

politics, from technology to culture, *et cetera*. It is beyond doubt that all of those who are interested in listening to one, will find a podcast with a theme to their liking. As for this thesis, most of the aforementioned genres were excluded soon after delving into the topics they were concerned with. However, art, culture, and language prevailed, taking into account the personal preferences of the author. Since the criterion of overriding importance was language, meaning that the podcast to be analyzed had to be hosted by native speakers of English, only websites featuring English-language podcasts were browsed.

As the next step, eight shows in total were chosen, having either art, culture, or language as a theme. From the category of art, *the Allusionist* and *Art Detective* were reviewed more closely; from culture-themed podcasts, *Mostly Lit* (all three featured in Smith Galer 2017) and *Pop Culture Happy Hour (PCHH)* (Herman 2015) were chosen; lastly, *The World in Words*, *Talk the Talk*, *A Way with Words*, and *The Word Nerds* (the last four found from Hayward and Stimola n.d.) were researched more closely from podcasts dealing with the topic of language. For the reason that linguistics is closer to the heart of the author of the present thesis, it was finally decided that the choice will be made from the last four shows. Then, as they were previously unknown to the author, they all had to be studied more carefully. This included skimming the titles of the episodes, as well as listening to some extracts. At last, it was decided that the Perth-based linguistics podcast *Talk the Talk* would be the best choice, providing the reader with enjoyable, yet intellectually stimulating content.

Talk the Talk is a show where host(s) discuss current topics in linguistics, some episodes have also evolved from the comments and questions sent to the hosts, such as episode 353: “Mailbag of Highly Intelligent Listeners”, which aired Feb 27, 2019, and episode 334: “Mailbag of Darkness,” aired Aug 21, 2018. To this date, they have created over 350 audio recordings, and show no signs of stopping anytime soon. The main hosts include linguist Daniel Midgley, who speaks the American variety of English; a media

studies teacher Ben Ainslie, and a former philosophy teacher Kylie Sturgess, who are both speakers of Australian English, as well as an occasional Swedish host Hedvig Skirgård, who is currently completing her doctoral studies at the Australian National University. Sometimes they also include guest speakers, or do solo episodes, depending on the topic. The podcast airs 11 a.m. every Tuesday on RTRFM 92.1. (Talk the Talk podcast 2019) Having selected their show, one had to opt for one episode from the 354 available, which took some time and effort. Nevertheless, deriving again from personal preferences, the choice was finally narrowed down to episodes 300: “Is Grammar Elitist?” which aired Sept 12, 2017; 315: “Grammar Day” from Feb 28, 2018, and “327: How We Talk,” aired May 29, 2018.

All three episodes were co-hosted, a requirement for the analysis of questions in oral speech, since at least two people speaking serve as a prerequisite for a natural-sounding dialogue. The first two were hosted by Midgley, Ainslie, and Sturgess, while the third had a fourth contributor, Skirgård, and a linguistic anthropologist, Prof. Nick Enfield, as a guest. Taking into consideration the topics, episode 315: “Grammar Day” was decided on as the primary source for analysis. This episode dispels several grammar myths, clarifies terminological issues, and brings fun, yet educative examples of the rules people (un)consciously follow. Furthermore, “Grammar Day” features discussion of listeners’ comments, received online prior the airing of the episode, taking listeners as participants of the episode in a sense, a common trend in today’s media.

2.3 Transcribing the Text

Thanks to the rapid progress of science and technology, it has become much easier to analyze human speech. The advancements of speech recognition software and websites providing (semi)automatic transcription services have greatly facilitated carrying out preparatory work for compiling the corpus for this study. Without the necessary programs,

the early stages of the present thesis would have been considerably more challenging and time-consuming. However, as will be shown in the following analysis, it is not yet possible to fully rely on automated processes when it comes to speech recognition and automatic transcription. Human speech, especially with multiple speakers, is far too complicated for computers to recognize all its nuances, especially in such detail as is required in linguistic data analysis.

Taking into account the length of the podcast episode “Grammar Day” and lack of previous experience of transcription and annotating linguistic data, it was decided that a semi-automatic transcription tool will be used. It then became apparent, however, that there is a multitude of transcription tools available on the Internet. For that reason, several criteria were established in order to narrow down the websites providing such services. One of the first criteria was that the tool has to provide a free transcript. Another requirement was that it must be possible to later annotate all the features relevant to this study in the same program, thus avoiding switching between tabs. Then, some more detailed requirements were taken into account, including the program’s ability to identify speakers and how the user interface design looked like.

Considering the criteria established beforehand, a pilot study was conducted. Several websites offering (semi)automatic transcription were browsed to have an idea of their services and what conditions were applied by the owners. This process left the author with four websites which were then examined more thoroughly: sonix, Temi, Transcribe, and Trint. A sample transcript of two episodes was requested from all four, which can be seen in Appendix 1. One of the reasons for choosing two episodes instead of one was that both sonix and Trint only provided 30 minutes of free transcription. Another reason was that at the time of the request a definite decision had not been made regarding the specific episode to be used as the source material. For a detailed comparison and links for accessing these websites, see

Appendix 2. To give a brief overview of the results, it can be said that all four software provide the user with a satisfactory result. Depending on how detailed the task at hand needs to be, they are all suitable to some extent. However, considering the requirements of this thesis, a choice was made to use Temi. The fact that one audio file of any length could be transcribed at no cost during the free trial, its user-friendly design, and surprisingly accurate punctuation, including question marks, were the main reasons for deciding in favor of this software. The appearance of the user interface of Temi is shown in Figure 2.

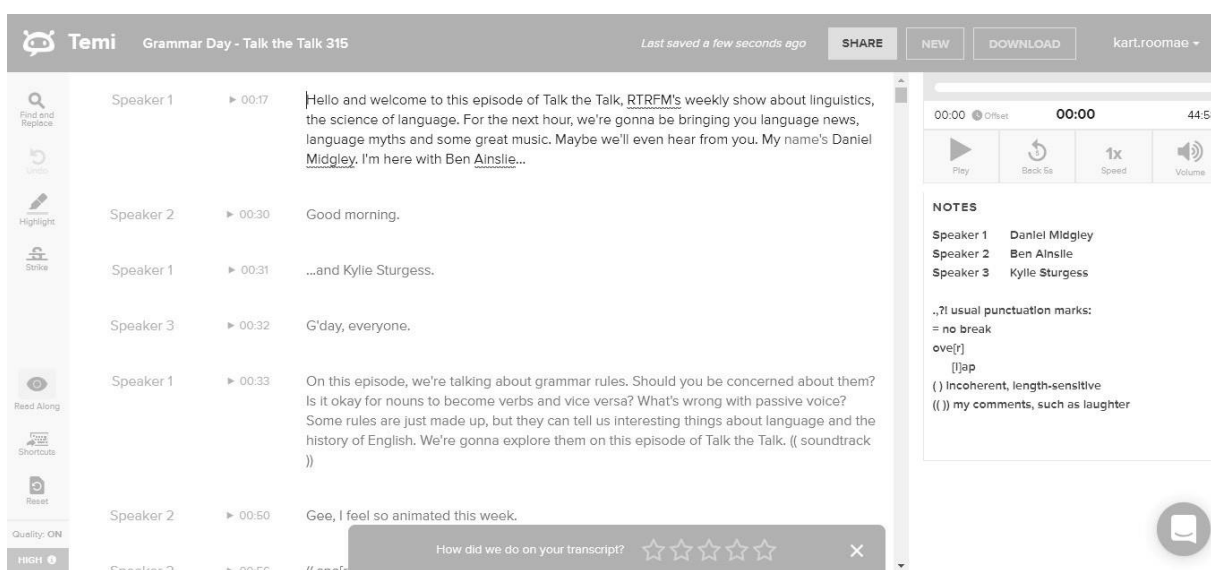


Figure 2. User interface of Temi

Furthermore, it is also possible to annotate the data in Temi. Linguistic annotation of a corpus refers to “the association of descriptive or analytic notations with language data.” This could mean transcribing, tagging, or labeling, among many other procedures. (Ide 2017: 2) Using the free transcript editor offered by Temi, one can easily annotate and adjust who is speaking and what is being told, timestamps included. Despite the user-friendly design, however, and due to the minuteness required for this study, the process of editing and annotating 44 minutes and 58 seconds of raw data took about 1300 minutes, which amounts to over 21 hours of manual work. Additionally, there were several shortcomings of Temi. The software misnamed speakers when people started talking at the same time, it left out

laughter as well as other indecipherable sounds and words, and timestamps were no longer accurate when there was overlapping speech. In order to preserve original features, such as laughter, audible inhales, and striking changes of tone, they were included during the process of annotation. However, without this tool, the preparatory work would have taken significantly more time, even with the time spent verifying names, terminology, and the written works referred to during the episode.

In order to provide the reader with an accurate reading of “Grammar Day,” Jefferson’s (2004) transcription system was used. Gail Jefferson is considered to be one of the founders of conversation analysis (henceforth, CA) together with Harvey Sacks and Emanuel Schegloff. At first, the idea was to edit the transcript as a whole so as to give the reader the context, providing him/her with a dialogue which is comprehensible to “common reader” and does not require previous exposure to transcription conventions used in CA, a text similar to that of a play. According to the initial plan, only questions would be annotated in greater detail, according to the conventions of the Jefferson Transcription System, paying attention to turns and overlaps, among other features. During the process of improving the transcript offered by Temi, however, it became clear that such differentiation would only complicate things. Because of that, the author opted for annotating the entire episode, after it had been downloaded in .docx file format, with a modified version of the Jefferson Transcription System, with some additional symbols used. In the end, the annotated transcript was 33 pages long, 10,031 words in total.

The transcription system developed by Jefferson which had its start in her transcribing Harvey Sack’s lectures in the 1960s (Lerner 2004: 2) is very detailed and complex. The annotation used in this study has been done according to a considerably simplified version of it, but it should be kept in mind that the main focus is also different. Traditional CA transcripts require comprehensiveness, since pertinent aspects cannot be

predicted (Koshik 2005: 4). While Jefferson's attempt was to write down practically every "pronunciational particular", pause, and laugh featured in the conversation (Jefferson 2004: 15), the goal here is to concentrate on the syntax and pragmatics of the categories of questions. Therefore, a selection was made from the transcript symbols used in the Jefferson Transcription System, which can be seen in Table 2 (for sample pages of the annotated transcript of "Grammar Day," see Appendix 3).

Table 2. Transcription conventions (adapted from Jefferson 2004: 24-31)²

Symbol	Signification
[overlap starts
]	overlap ends
=	no break/gap within one line
[text]= =[text]	no break/gap between two lines
.,?!)	punctuation markers
(())	transcriber's descriptions
gap	pauses
...	short pauses and instances where sentence is not finished
()	indiscernible

It has been taken into account that professional conversation analysts would not approve of including any other symbols, but the annotated transcription of "Grammar Day" serves another function. It should also be kept in mind that the resulting transcript is very subjective. The author remains entirely responsible for any remaining mistakes and realizes that some sections may be differently interpreted by other transcribers. In fact, it was later

² It should be noted that gaps as markers of pauses, symbol '...' as well as the exclamation mark have been added for a better reading experience. They do not appear in the conventions proposed by the Jefferson Transcription System.

found out that somebody's attempt at transcribing "Grammar Day" was later added to the episode's website (Ainslie et al 2018), where it is available for listening and download. Nevertheless, the version created by the author was retained, except for indiscernible sections, namely the titles of the tracks and the sounds which were imitated by the hosts, in order to provide the reader with a clearer text. Indistinguishable words and sounds were corrected according to the anonymous transcription available on the episode's website.

2.4 Results

After the quite lengthy process of transcribing and annotating raw data in the corpus of this study, the focus was shifted toward questions. It was decided that only direct questions would be analyzed and categorized for the reason that the material would have otherwise been too substantial for the scope of this study. Therefore, each question selected for this study ends with a question mark, contrary to reported ones, which also tend to be syntactically more complex. Adhering to the principles of CA methodology, no further categories nor research questions were established in advance. According to Koshik (2005: 3), "[a] CA analysis is data driven." The questions were therefore assigned to subcategories by interpreting the data collected during transcription and annotation. A parallel can be drawn with Sidnell's (2013: 77) introductory notes, wherein he compares CA with "a kind of exploration." His description of this process involves a series of steps:

The classic conversation analytic procedure begins with the noticing of some distinctive bit of behavior in social interaction. The analyst then works to locate other instances, and, in the process, begins to identify the boundaries of the phenomenon of interest. As instances are gathered into a collection, the analyst can begin to describe the practice or phenomenon in terms of its generic, context-independent properties, moving away from the particularities of any single case. However, despite this use of collections, the analyst always remains accountable to each individual case and its particularities. (Sidnell 2013: 78)

Koshik (2005: 4-5) also uses the word "micro-analysis" for examining examples of a phenomenon, a procedure which aids in constructing a comprehensive picture that plays a part in understanding fundamental topics in language.

In order to conduct a closer study by interpreting the data collected and annotated, and to see what kind of patterns emerge, the data were grouped together. Kania's (2016: 86-102) corpus study on adult-to-adult speech based on the British National Corpus was used as an example for the present analysis. Since the focus of this study is on the phenomenon of direct questions present in the episode of *Talk the Talk*, "Grammar Day," the questions were first imported into Excel. From 33 pages of annotated transcript (10,031 words), 132 direct questions were detected, the full list of which can be seen in Appendix 4 alongside with categories applicable to each question. No changes were made to the syntactic form of the questions. Considering the definitions and features proposed in the theoretical part of this thesis, the questions were first divided into two main categories: canonical and non-canonical questions (henceforth, CAQ and NONCAQ).

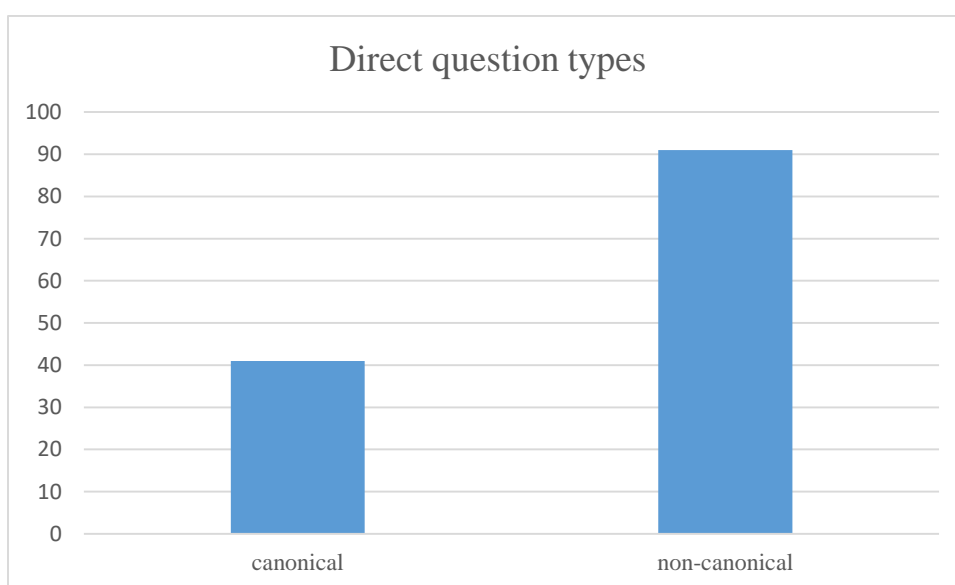


Figure 3. The distribution of direct question types in "Grammar Day"

As can be seen in Figure 3, non-canonical questions predominate. There are 91 cases (69%) belonging to the category of NONCAQ against 41 (31%) cases of the CAQ. This result can be explained by the fact that the raw data come from a podcast episode, wherein language use is less restrained than, for instance, in a news broadcast, albeit both exemplifying oral discourse. The setting of "Grammar Day" is informal and the hosts may

speak quite freely. Their main focus is not on producing absolutely grammatical sentences, and because of that, the data were much more interesting to analyze, reflecting the actual use of language. Additionally, there are three different people talking, which leads to overlaps and frequent reactions to each other's statements.

After the first division had been made, the NONCAQ were assigned to five subcategories, bearing in mind that they are the phenomenon of greatest interest to the author. The subcategories of the NONCAQ were: 1) fragments (*The exam tomorrow?*); 2) rhetorical, ability, and inclination questions (*Can you take the exam tomorrow?*); 3) (strong) assertive questions (*The exam isn't tomorrow?*); 4) tag questions (*The exam tomorrow, right?*), and 5) wh-questions (*What exam you taking tomorrow?*). All five will also be further treated in the discussion of the results (see Section 2.4). The bar plot, which can be seen in Figure 4, demonstrates the exact division.

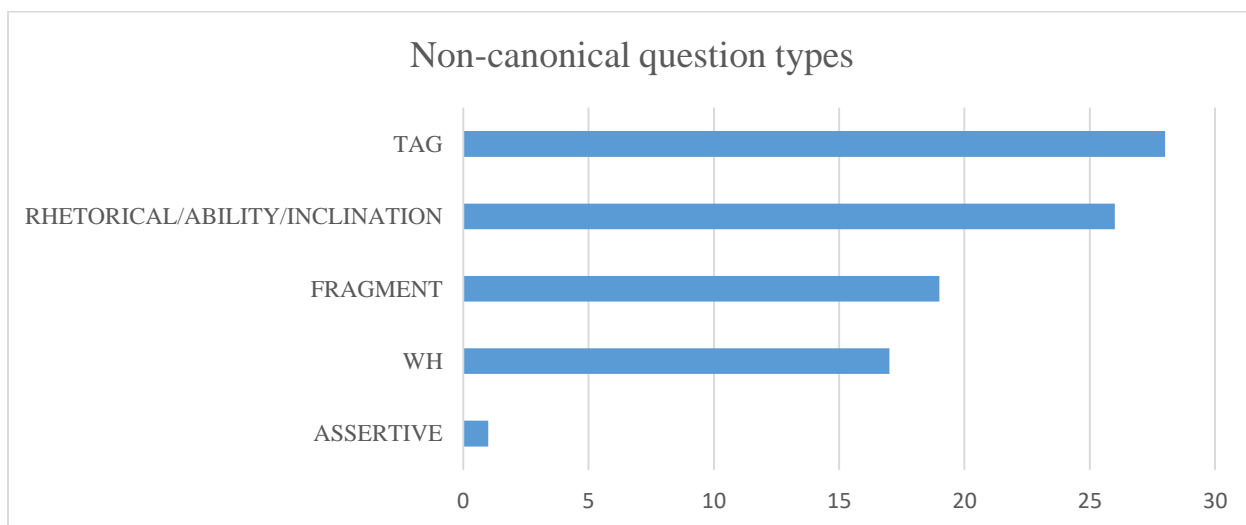


Figure 4. The distribution of non-canonical question types in “Grammar Day”

It is possible to deduce from Figure 4 that the two largest categories are the categories of tag (31% or 28 observations) and rhetorical/ability/inclination questions (29% or 26 in absolute numbers), followed by fragments (21% or 19 instances) and wh-questions (19% or 17 questions). The least frequent category is the category of assertive questions, being equal to just 1%, or 1 observation. A similar two-fold categorization: first CAQ vs. NONCAQ

marking and then allocation to subcategories was employed in Kania's (2016: 86-102) corpus study, but her work was significantly more detailed. The comparison of the results is given in the discussion.

There was also considerable variation present among speakers when it comes to direct questions. As was introduced in Chapter 1, Section 1.1, episode 315: "Grammar Day" was hosted by three people: Daniel Midgley, Ben Ainslie, and Kylie Sturgess. For the purposes of this study, they were identified as Speaker 1, Speaker 2, and Speaker 3, respectively, given that this was the order of appearance in this episode. Figure 5 shows the exact number of questions they all asked.

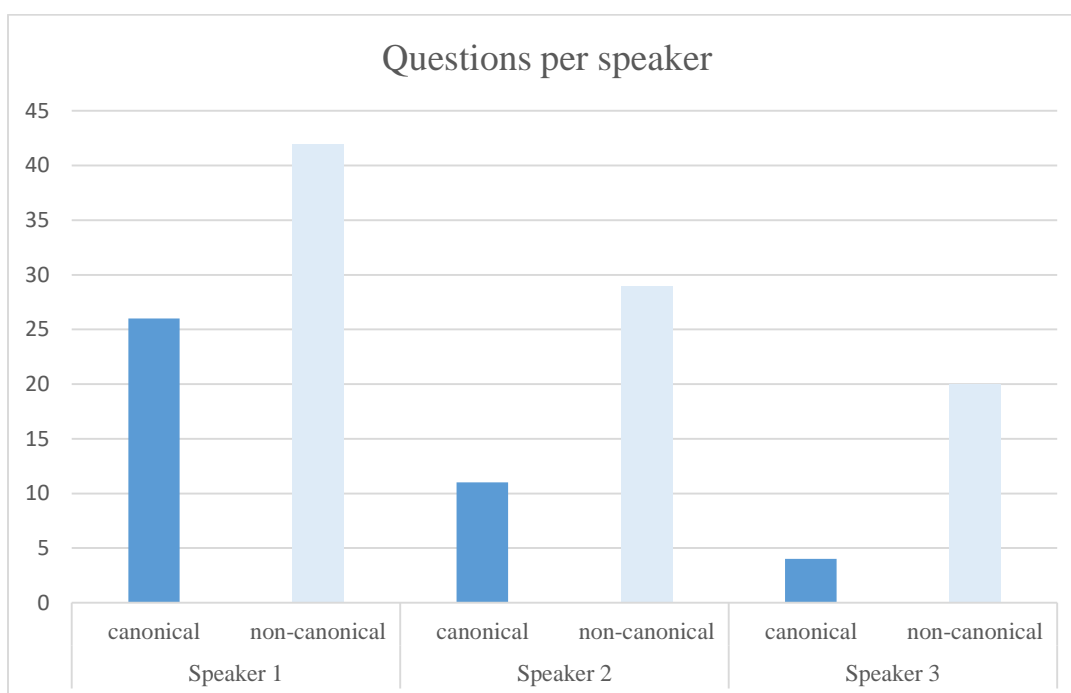


Figure 5. The distribution of direct question types among speakers in "Grammar Day"

It is evident that Speaker 1 completely dominated the discussion throughout the episode. The number of questions asked by Midgley was 68 (52%) in total, out of which 26 were CAQ and 42 NONCAQ. The number of questions asked by Midgley is considerably larger than that of Ainslie and Sturgess. Speaker 2 asked 40 questions – 11 CAQ and 29 NONCAQ, making up 30%, while Speaker 3 asked 24 questions, 4 CAQ and 20 NONCAQ

respectively, which constitutes 18%. The fact that Speaker 1 asked significantly more questions than the other two hosts may be explained by the fact that he was the only professional linguist participating in that episode. He is obviously better acquainted with the terminology used and he is also the one who compiled quizzes present in “Grammar Day.” Therefore there are several reasons for his leading role. Kylie (Speaker 3) is often referred to as a skeptic, which may also indicate the markedly smaller number of questions asked. Her role resembles more of an observer, who provides outsider insight upon request. Daniel (Speaker 2) appears as curious and keen on learning new things about languages and linguistics.

Personality traits may also play a role. Some are less likely to ask questions than others. Individual differences have been studied by Dąbrowska (2015). While Dąbrowska’s (2015) approach was usage-based, the focus was on grammatical knowledge, and the data were obtained from linguistic experiments, rather than by analyzing conversation wherein people are not aware of being studied, which is the approach used in the present thesis. Still, the general observations can be made: “underlying cognitive differences” affect people’s performance (Dąbrowska 2015: 659). Additionally, the “linguistic experience” is not the same for everyone because of education and profession (Dąbrowska 2015: 660). Those who work more with language normally read more as well, which adds variation to their language use. The lexis and grammar used by people also differs, as opposed to the principles of shared Universal Grammar (Dąbrowska 2015: 663). It would also be interesting to study whether the variety of English the participants speak and possibly even gender have an influence on the number of questions.

2.5 Discussion

As it turned out during the process of analyzing the direct questions in the transcribed and annotated version of the podcast episode “Grammar Day,” the frequency of non-

canonical questions is larger than, for example, in the sample of the British National Corpus (cf. Kania 2016). In the present analysis, 91 cases (69%) belonged to the category of NONCAQ against 41 (31%) cases of the CAQ. Kania's (2016: 99) results, however, show that canonical questions predominated with 948 cases (55%), while there were 777 cases (45%) assigned to the category of NONCAQ. One should nevertheless be careful with drawing conclusion based on these findings, since the total number of questions analyzed in the two afore-mentioned studies is very different. Therefore, it would be interesting to carry out a more in-depth research on the topic in the future by increasing the number of podcast episodes in the analysis. Difference between mediums, oral language studied in Kania's (2016) study and podcast in the present thesis could be of importance, in addition to the specific variety of English – British vs. American vs. Australian in this case. Moreover, it would be interesting to do a statistical analysis on how many questions were asked during one turn.

The first large question type of the present thesis, CAQ, which consisted of 41 cases, forms a significant minority in the compiled corpus. Returning to Chapter 1, Section 1.2, which was meant to aid in differentiating between CAQ and NONCAQ, this category was established according to the functional and/or syntactic properties of the particular question. People use CAQ when their aim is to obtain new information, contrary to NONCAQ. Additionally, the constraints of interrogative syntax may be ignored in case of NONCAQ. The criteria of a speech act of questioning also matter (cf. Dayal 2016: 4-5). As for the specific examples of CAQ, in addition to the so-called ordinary questions, such as (29)³ *Are we going to wear to like T-shirts and something or [come up wi]th slogans? =* and (74) *Do you think that you 'made a demand' or do you think that you could 'demand something' first?*,

³ The number preceding the examples refers to the order of appearance of the question in the annotated transcript (cf. Appendix 4). In case of all examples, the original form has been retained, i.e., no changes have been made to their syntactic structure.

a pattern formed, which well exemplifies the nature of oral discourse. Namely, contracted forms were used. For instance, questions (10) *'Why don't you tell me when you don't understand me?'* and (42) *Doesn't your university do this as well with their slogan?=. Moreover, several cases began with a conjunction, which is normally frowned upon in writing (see, for example, question (24) *And do you know what it does?*). It is clear that while these questions all belong to the same category, variation is very much present.*

Looking again at the results of the NONCAQ subdivision in the present thesis (see Figure 4), some general remarks will be made first in order to give an overview of the overall trends. It is evident that two of the categories, tag questions and rhetorical/ability/inclination questions, are of approximately the same frequency, the former comprising 28 cases, while the latter has 26 instances assigned to it. Following a similar pattern, there are 19 fragmented questions and 17 wh-questions. Last but not least, one question belonged to the subcategory of assertive questions. Some of the questions in those five subcategories were easily divided, whereas others presented difficulties, mostly because of the contrast between direct and indirect speech acts. For example, questions (25) *'Would you like some coffee?'*, (50) *Is it Roger Moore?* as well as (58) *Why do you care?* Looking at the problematic cases, it was first decided to mark them as 'ambiguous,' but with the help of discursive context, the category was omitted later. Namely, questions (25) and (50) are syntactically CAQ, but they are used as NONCAQ when considering their pragmatic properties. Thus they were categorized as rhetorical/ability/inclination questions. Question (58), however, may resemble a NONCAQ at first, but is actually pragmatically used as a CAQ, since the speaker truly wants to know the answer.

Moving on to the more specific analysis of NONCAQ and following the alphabetical order decided upon earlier, the first subcategory to be examined is *fragmented questions*. The first subcategory, fragmented questions, with 19 examples in the corpus exemplifies one

of the main NONCAQ criteria (cf. Dayal 2016: 268), non-standard syntactic form. The name of the category owes to the fact that some elements that normally occur in questions have been omitted, such as auxiliary, subject, or predicate (cf. Kania 2016: 63-64). The length of questions belonging to this subcategory in the corpus of this study is strongly inclined towards 1-3 words on average. To illustrate the aforementioned characteristics and constraints, questions (21 and 22) *'huhh?'* and (108) *Dddo you...?* show the unrestrained spontaneous dialogue of "Grammar Day," while questions (69) *I hate to inconvenience you, 'the verb or, 'it's such an inconvenience, 'the noun?* and (93) *You sure you don't want to rephrase it?* show relatively high contrast with the former. One should also take into account the discourse context and whether the question is meant to signal one's reaction or to ask for information, although in a less eloquent way than ordinary questions.

It was notably harder to classify questions as *rhetorical, ability, and inclination question*. The speech act theory which was briefly introduced in the theoretical part of this thesis (see Chapter 1, Section 1.1) defined questions belonging to this category as examples of indirect speech acts. Searle's (1979: 30-31) several semantic layers, or meanings that are heard versus those that are understood, are also featured in this corpus. For that reason, it was also difficult in the beginning to categorize these questions as CAQ or NONCAQ. For example, question (30) *Can a wooden house be dilapidated?* can technically serve as a CAQ in some other context. In "Grammar Day," however, it was clearly a rhetorical question, not seeking an answer, since the person asking it already knew the answer. Question (96) *Do you know what it's like for me?* is also rhetorical, since it is obvious that the addressee cannot see into another person's mind. Sometimes, though, this type can go unnoticed. Namely, question (50) *Is it Roger Moore?* is syntactically absolutely standard question, but in reality, the focus was on the word 'more' and no one really wanted to know whether it is Roger Moore.

The third subcategory, *(strong) assertive questions*, consisted of just one question. It was the case (31) *You think so?* The classification used in this study actually included both negative polar and declarative questions under this subcategory. Dayal (2016: 282) has given them some attention and characterizes these questions as biased. The person asking already has some kind of an opinion and the way the question has been formulated confirms it. The one and only case represented in the corpus of this study has an affirmative syntactic form, which expresses the supposition resulting from what the addressee has just verbalized during the previous turn and will repeat right after the question.

The fourth category included *tag questions*. According to Dayal (2016: 288), this type functions as a way of asking for acknowledgement or confirmation. This claim is supported by the fact that words ‘right’ and ‘really’ were present in most of the sentences of this category, 14 and 6 cases, respectively, the total being 28. Some of these consisted of just this one word, such as (29) *Right?* and (78) *Really? =*, while others had the aforementioned tag in the end of the question: (113) *All the lights go down, right?* The questions containing either ‘right’ or ‘really’ have an affirmative matrix clause, i.e., the part of the sentence that comes before the tag, which is usually separated by a comma. Some of the questions, however, combine affirmative and negative form. This happens when the question tag has an auxiliary in it. Questions (43) *Well, it’s vaguely poetic, isn’t it? =* as well as (79) *They come a lot later, don’t they?* show that both auxiliaries ‘be’ and ‘do’ are present in the sentences. One question, (122) *=And some of them are bloody crisis actors, you know what I mean?* highlights that there are different ways of asking for confirmation.

The last subcategory, *wh-questions*, refers to questions containing a wh-word. This corpus has 9 out of 17 cases with ‘what,’ 6 having ‘why’ and 2 with ‘how.’ Some of them consist of just one word, such as (20) *What?*, (57) *Whyyy?*, some are quite fragmented, conveying the openness of these questions, namely, questions (95) *Why should I...?* and (53)

So what part of speech does that...? This kind of lack of restriction again characterizes unscripted dialogue, where people speak spontaneously, although the main points have been written down prior to recording the episode. People speak in turns, often interrupting each other and as a result, some of the question cannot be completed. Nevertheless, question (81) *I mean, why invent two separate words for closely related things?* displays that not all wh-questions are so short and fragmentary. Still, questions (41) *[What']s up?* and (60), (102) *You know what?* make it rather evident that shorter types are preferred.

The discussion of the results shows that the NONCAQ form the majority with 69% over 31% of the CAQ, which contrasts with Kania's (2016) study, even though these two studies are very different, both in terms of categorization and the sample size. The types discussed in this section were constructed according to the functional and/or syntactic properties of the particular question, keeping in mind the distinction established by different authors mentioned in Sections 1.1 and 1.2. The two larger categories were CAQ and NONCAQ, and 5 subtypes of the NONCAQ were later analyzed separately. Most frequent were the use of tag (28 cases) and rhetorical/ability/inclination questions (26 cases), followed by 19 fragmented and 17 wh-questions. There was also one assertive question in the corpus. Variation was present in each of the categories, which exemplifies once more the unrestrained speech in "Grammar Day." People say what they first think of, and there is no pre-written dialogue, only some general topics and words to be discussed.

CONCLUSION

The present thesis aimed at studying direct questions in the annotated transcript of a linguistics podcast “Grammar Day.” The corpus was compiled from a podcast episode for two main reasons. There is audio without non-verbal communication and therefore it was not necessary to study physical paralinguistic features, such as gestures and posture of the hosts of the podcast. The other reason lies in the fact that podcasts have existed for just 15 years, since the publication of Hammersley’s (2004) article. The term denotes a blend of *iPod* and *broadcast*, a digital audio recording to be listened offline, having been either downloaded or streamed beforehand (Chandler and Munday 2016). Other shows featuring (semi)oral discourse, like talk shows and news broadcasts were therefore not analyzed, considering also the fact that there have been several comprehensive studies on them, including from the perspective of questions.

The direct questions were first divided into two categories, canonical and non-canonical questions, which differ from each other in their syntactic form and/or function. Canonical questions are used to seek information, while non-canonical questions are either biased, have a non-standard syntactic form, or function as indirect speech acts (Dayal 2016: 268). Two questions from the corpus of this study aid in highlighting the difference: (5) *Daniel, what’s been going on in the world of linguistics in the week gone past?*= and an example of a non-canonical question: (66) *Noun or verb?* Non-canonical questions were further allocated to five different subcategories, according to their syntactic form as well as the semantic and pragmatic properties, using a combination of the methodology used in two studies, Dayal (2016) and Kania (2016).

It was revealed during the analysis that in a natural discourse context, speakers tend to prefer non-canonical forms. Out of 132 questions, 91 were non-canonical questions (69%), the rest of them, 41 (31%), were canonical. Out of 91 non-canonical questions, tag

questions were the most popular choice (31% or 28 cases), followed by rhetorical/ability/inclination (29% or 26 cases), fragmented (21% or 19 cases), and wh-questions (19% or 17 cases). The fifth category, assertive questions, has just 1 example in the corpus, which amounts to 1%. There were three speakers participating in the episode and the most active of them, Daniel Midgley, asked 68 questions (52%), while Ben Ainslie came second with 40 questions (30%). The third speaker, Kylie Sturgess, asked 24 questions (18%) in total. In Kania's (2016) study, the proportion of canonical and non-canonical questions was different: Kania's (2016: 88) sample from the British National Corpus, which consists of face-to-face spontaneous conversations of the British variety, had a larger number of questions and there were 948 canonical (55%) over 777 (45%) non-canonical questions.

The scope of this thesis was relatively limited and did not allow for a detailed analysis of all the question types mentioned. Given that the classification of questions into canonical and non-canonical questions is a rather new approach, not many papers have been written about such categorization, especially with respect to questions featured in podcasts, although it is possible that we do not know all the research papers written on the topic of canonical and non-canonical questions. This kind of approach, therefore, might not be that new.

This thesis serves as a pilot study for a possible MA thesis, establishing the ground and giving a chance to get acquainted with the typology of questions as well as the phenomenon of podcasts. At this stage, it seems most interesting to continue studying non-canonical questions, since these have not been treated sufficiently thus far. The focus would likely be on one subcategory of them, narrowing down the study and allowing for a more fine-grained analysis. Additionally, a larger sample of podcast episodes to be studied would be advisable. A larger corpus would provide more data and this would enable to see what kind of patterns form on a larger scale, across different podcasts and different speakers. Therefore, using one of the paid transcription tools could facilitate the process. At the same

time, this corpus can also form the basis of future research for others. The annotated transcript of “Grammar Day” can be used for analyzing some completely different linguistic phenomenon. The full corpus can be accessed via the QR code provided in Appendix 3.

LIST OF REFERENCES

Primary Sources

- Ainslie, Ben, Daniel Midgley, and Kylie Sturgess (hosts). 2018. Grammar Day. *Talk the Talk*. [Audio podcast]. Available at <https://talkthetalkpodcast.com/315-grammar-day/>, accessed Feb 17, 2019.
- Cox, Patrick (host). 2018. When an American says 'sure' to a Brit, does it mean yes or no? *The World in Words* [Audio podcast]. Available at <https://www.pri.org/stories/2018-10-10/when-american-says-sure-brit-does-it-mean-yes-or-no>, accessed March 25, 2019.

Secondary Sources

- Austin, J. L. 1962. *How to Do Things with Words*. Oxford: Oxford University Press.
- Castos. 2019. iTunes Podcast Category List. Available at <https://castos.com/itunes-podcast-category-list/>, accessed March 5, 2019.
- Cohen, Akiba A. 1989. Answers without questions: a comparative analysis of television news interviews. *European Journal of Communication*, 4: 4, 435-451.
- Chandler, Daniel and Rod Munday (eds). 2016. *A Dictionary of Media and Communication*. 2nd ed. Oxford University Press. Available at <http://www.oxfordreference.com.ezproxy.utlib.ut.ee/view/10.1093/acref/9780191800986.001.0001/acref-9780191800986-e-3088?rskey=0Gbbvp&result=2>, accessed Feb 20, 2019.
- Dąbrowska, Ewa. 2015. Individual differences in grammatical knowledge. In Dąbrowska, Ewa and Dagmar Divjak (eds). *Handbook of Cognitive Linguistics*, 650-668. Berlin: De Gruyter Mouton.
- Danileiko, Nadia. 2005. Formal and Functional Questions in an American Talk Show *Late Night with Conan O'Brien*. University of Jyväskylä. Available at https://jyx.jyu.fi/bitstream/handle/123456789/7405/URN_NBN_fi_jyu-2005470.pdf?sequence=1, accessed Feb 20, 2019.
- Dayal, Veneeta. 2016. *Questions*. Oxford: Oxford University Press.
- Emmertsen, Sophie. 2007. Interviewers' challenging questions in British debate interviews. *Journal of Pragmatics*, 39: 3, 570-591.
- Hayward, Brittany and Maureen Stimola. n.d. The top 15 language learning podcasts for curious multilingual minds. Available at <https://www.fluentu.com/blog/language-podcasts/#>, accessed March 5, 2019.
- Heritage, John. 2002. The limits of questioning: negative interrogatives and hostile question content. *Journal of Pragmatics*, 34: 10-11, 1427-1446.
- Heritage, John C. and Andrew L. Roth. 1995. Grammar and institution: questions and questioning in the broadcast news interview. *Research on Language and Social Interaction*, 28: 1, 1-60.
- Herman, Alison. 2015. The 20 best culture podcasts. Available at <http://flavorwire.com/539890/the-20-best-culture-podcasts>, accessed March 5, 2019.
- Ide, Nancy. 2017. Introduction. In Ide, Nancy and James Pustejovsky (eds). *Handbook of Linguistic Annotation*, 1-18. Dordrecht: Springer.

- Jefferson, Gail. 2004. Glossary of transcript symbols with an introduction. In Lerner, Gene H. (ed). *Conversation Analysis: Studies from the first generation*, 13-31. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kania, Ursula. 2016. *The Acquisition and Use of Yes-no Questions in English: A Corpus-Study from a usage-based perspective*. Tübingen: Narr Francke Attempto Verlag GmbH + Co. KG.
- Koshik, Irene. 2005. *Beyond Rhetorical Questions: Assertive questions in everyday interaction*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Lerner, Gene H. 2004. Introductory remarks. In Lerner, Gene H. (ed). *Conversation Analysis: Studies from the first generation*, 1-11. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Matthews, P.H. 2014. *The Concise Oxford Dictionary of Linguistics*. 3rd ed. Oxford University Press. Available at <http://www.oxfordreference.com.ezproxy.utlib.ut.ee/view/10.1093/acref/9780199675128.001.0001/acref-9780199675128-e-2731?rskey=07afIE&result=1>, accessed Feb 20, 2019.
- RAJAR Midas Audio Survey. 2018. Available at https://www.rajar.co.uk/docs/news/MIDAS_Spring_2018.pdf, accessed Feb 28, 2019.
- Schirm, Anita. 2009. The Role of Questions in Talk Shows. In Dynel, Marta (ed). *Advances in Discourse Approaches*, 147-173. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Searle, John R. 1979. *Expression and Meaning: Studies in the Theory of Speech Acts*. New York: Cambridge University Press.
- Sidnell, Jack. 2013. Basic Conversation Analytic Methods. In Sidnell, Jack and Tanya Stivers (eds). *The Handbook of Conversation Analysis*, 77-100. Chichester: Wiley-Blackwell.
- Smith Galer, Sophia. 2017. The 25 culture podcasts that will blow your mind. Available at <http://www.bbc.com/culture/story/20170804-the-25-culture-podcasts-that-will-blow-your-mind>, accessed March 5, 2019.
- Talk the Talk. 2019. About us. Available at <http://talkthetalkpodcast.com/about-us/>, accessed April 24, 2019.
- Woods, Bob. 2018. The podcasting revenue boom has started. Available at <https://www.strategy-business.com/article/The-Podcasting-Revenue-Boom-Has-Started?gko=40bef>, accessed Feb 28, 2019.

Appendix 1: Approximately one-minute long extracts of the transcripts.⁴

Trint https://trint.com/	sonix https://sonix.ai/
"A Family Divided by English" – <i>The World in Words</i> episode 464	
[00:00:00] Viner is British. His wife Lynne is American.	[00:00:00] Viner is British. His wife Lynne is American.
[00:00:04] So we met in England and she'd only been here about one year two years two years.	[00:00:04] So we met in England and she'd only been here about one year two years two years.
[00:00:11] That was 16 years ago. Like all couples the communication hasn't always been smooth. They can get stuck even though what we might think of as the simplest of words. I'll ask a question and Lynne will say.	[00:00:11] That was 16 years ago. Like all couples the communication hasn't always been smooth. They can get stuck even though what we might think of as the simplest of words.
[00:00:26] Which to me means I couldn't care less. I just hear it as a now whereas I see it as a yes yes.	[00:00:21] I'll ask a question and Lynne will say which to me means I couldn't care less.
[00:00:38] No. Day night. Real Fake British American.	[00:00:30] I just hear it as a now whereas I see it as a yes yes no. Day night real fake British American it's the World in Words.
[00:00:49] It's the World in Words. I'm Patrick Cox. Today you say aluminum changing variations of English still muddying those transatlantic waters back to wife and husband Lynne and Phil who still aren't sure about sure what Americans say what Brits here mean and as a yes.	[00:00:50] I'm Patrick Cox. Today you say aluminum changing variations of English still muddying those transatlantic waters back to wife and husband Lynne and Phil who still aren't sure about sure what Americans say what Brits here mean and as a yes.
Transcribe https://transcribe.wreally.com/	Temi www.temi.com
"Grammar Day" – <i>Talk the Talk</i> episode 315	
Speaker 1: [00:00:10] Hello and welcome to this episode of talk the talk RTR FM's weekly show about Linguistics the science of language for the next hour. We're going to be bringing you language [00:00:25] news language myths and some great music. Maybe we'll even hear from you. My name is Daniel Midgley. I'm here with Ben Ainslie. Good morning and Kylie stretches. Should I everyone on this episode we're talking about grammar rules. Should you be	Speaker 1 00:17 Hello and welcome to this episode of talk, the Talk Rtr fms weekly show about linguistics, the science of language. For the next hour, we're going to be bringing you language news, language myths, and some great music made me believe in here from you. My name's Daniel. Mentally, I'm here with Ben Ainslie. Good morning, and Kylie's judges everyone on this episode, we're talking about grammar rules. Should you be concerned about them? Is it okay for nouns to become verbs and vice versa?

⁴ These extracts present the beginning of the podcast episodes. More striking differences have been highlighted in gray.

<p>concerned about them? Is it? Okay for nouns to become verbs and vice versa [00:00:40] what's wrong with passive voice? Some rules are just made up, but they can tell us interesting things about language and the history of English. We're going to explore them on this</p>	<p>What's wrong with passive voice? Some rules are just made up, but they can tell us interesting things about language and the history of English. We're going to explore them on this episode of</p>
<p>Speaker 3: episode of Talk The Talk.</p>	<p>Speaker 2 00:48 the talk.</p>
<p>Speaker 2: Jay I feel so animated this week</p>	<p>Speaker 3 00:50 Gee, I feel so animated this week. I'm so glad so many people have been checking them out on our social media. Those</p>
<p>Speaker 3: wasn't it beautiful? [00:00:55] I love those but I'm so glad so</p>	
<p>Speaker 2: many people have been checking them out on our social media those.</p>	

Appendix 2: Comparison of websites offering (semi)automatic transcription services.

Features	sonix	Trint	Temi	Transcribe
integrated editor	yes	yes	yes	yes
speaker identification	yes	yes	yes	yes
turn-taking modification	yes	yes	yes	no
timestamps	yes	yes	yes	yes
adjustable timestamps	yes	yes	no	yes
punctuation	yes	yes ⁵	yes	yes
contracted forms identification	yes	yes	yes	yes
acronym expansion	no	no	no	yes
strikethrough	yes	yes	yes	no
highlight	yes	yes	yes	no
find & replace	yes	yes	yes	no
accompanying app for iOS	no	yes	yes	no
accompanying app for Android	no	no	yes	no
available offline	no	no	no	yes
several export formats	yes	yes	yes	no
export of specific sections	yes	yes	yes	no
separate account needed	no	no	yes	yes
individual license	yes	yes	yes	yes
business/group license	yes	yes	no	yes
30 minutes free	yes	yes	yes	yes ⁶
more than 30 minutes free	no	no	yes	no
price per minute	no	no	yes	yes
price per hour	yes	yes ⁷	no	no
price per month	yes	yes	no	no
price per year	no	no	no	yes

⁵ Only period included.

⁶ Only available in 1-minute long packages, a restriction which is not advertised on the website and will only be seen after signing up.

⁷ Applicable only in case of exceeding monthly upload limit.

Appendix 3: Extracts from the annotated transcript of “Grammar Day.”⁸

Page 1

- Speaker 1: Hello and welcome to this episode of Talk the Talk, RTRFM's weekly show about linguistics, the science of language. For the next hour, we're gonna be bringing you language news, language myths and some great music. Maybe we'll even hear from you. My name's Daniel Midgley. I'm here with Ben Ainslie...
- Speaker 2: Good morning.
- Speaker 1: ...and Kylie Sturgess.
- Speaker 3: G'day, everyone.
- Speaker 1: On this episode, we're talking about grammar rules. Should you be concerned about them? Is it okay for nouns to become verbs and vice versa? What's wrong with passive voice? Some rules are just made up, but they can tell us interesting things about language and the history of English. We're gonna explore them on this episode of Talk the Talk.

⁸ The 5 sections presented here refer to pages 1, 11, 13, 32, and 33 in the transcript. In addition to the first and last page of the transcript, it was decided to choose pages that may be of interest, especially when paying attention to overlaps, pauses, and sounds featured in the episode. The sample pages also draw attention to the shortness of turns, common in natural discourse, although pages 32 and 33 show that this is not always the case. All direct questions have been highlighted. The full corpus can be accessed at <https://bit.ly/2Q8rD9X> or by scanning this QR code:



((soundtrack))

Speaker 2: Gee, I feel so animated this week.

Speaker 2: ((sno[rt])=) [mm]

Speaker 3: [=An]d wasn't that beautiful? I love those vid.=I'm so glad so many people have been checking them out on our social media. Those [videos are a-mazing.]=

Speaker 2: =[Foor listeners who mi]ght not be aware, a particularly enthusiastic listener has animated small segments of our show.

Speaker 3: Mm.=

Speaker 1: =And we're pretty cute in these segments too.=

Speaker 2: =Mm[hmh]=

Speaker 3: =[Well], I immediately made my profile pic [on Facebook ((la]ughter)), we're just so beautiful.=

Speaker 1: [Yes, I've noticed.]

Speaker 2: =If I were to use the parlance, I believe we have been chibified.

Speaker 3: Aaah.=

Speaker 1: =Woow, [I like that.]

Speaker 2: [Chibi. Like] 'cute' in like Anime language.

Page 11

Speaker 2: [Pause, pause, pause, pause.] First of all, you were definitely that kid in school. Second of all, ((chuckle))

Speaker 3: Argh. It was just irrit... I think it was because she kept on going on and on [and I just...]

Speaker 2: [Sounds of] awesome. [I actually kinda like]that.

Speaker 1: [Sounds of awesome.]

Speaker 2: Like don't get me wrong. If I...=

- Speaker 3: =((mutter))
- Speaker 2: If I was to sit in a round table at a marketing pitch meeting and that was floated, it would definitely not be going on the top of my list, but at the same time, 'sounds of awesome.' Like 'the tinkling sounds of awesome emanating from the speakers.'
- Speaker 3: No[oo] [Nooo.]
- Speaker 2: [I'm] on board. I'm on board. [Kylie], what is your actual problem though?=
 Speaker 1: =Yeah. Can you describe what is wrong with this grammatically?
 Speaker 3: She was using 'awesome,' which is an adjective, and she was trying to use it as a noun.
 Speaker 1: Right... Okay.
 Speaker 2: 'Sounds of awesome,' awesome being a thing.
 Speaker 1: Yeah.=
 Speaker 3: =Ye[ah.]
 Speaker 1: [She] was trying to sort of convert that.=I heard a lot of examples of this watching the Olympics and one of the announcers said, 'It's a solid run, but is it enough to podium?' Right.= Okay.
 Speaker 3: =Yes!
 Speaker 2: So good!
 Speaker 1: I know.
 Speaker 2: So good! I'm so on board with this.
- Page 13:**
- Speaker 3: Yep. That's another one.
- Speaker 1: Uum... Some people call it nominalization or verbing, if it's going the other way.=
- Speaker 3: =Zero derivation is another way of calling [it.]
- Speaker 2: [U]uuh, that's fun.=

- Speaker 1: =That's a good one, 'cause... 'cause you're not adding anything like -
ment, as in 'development'.
- Speaker 3: Yeahh.
- Speaker 1: Or... Or... Umm...
- Speaker 2: Zero derivation.=
- Speaker 1: =Yeah. You're just taking that noun, not adding anything, making a
verb.
- Speaker 3: So you're gonna see a lot of it in advertising. People trying to match
words.
- Speaker 2: Well, it's vaguely poetic, isn't it?=
[Mm]h. Ummh.
- Speaker 3: =Ye[ah.]
- Speaker 1: [Mm]h. Ummh.
- Speaker 2: Like you take this noun and you create a verby verby verb out of it.
Because a 'scarf,' right, is such an idiosyncratic noun, right? A 'scarf'
is a scarf and so, 'to scarf around you' is just soo evocative.
- Speaker 1: Mm[mm.]
- Speaker 3: [Mm]... Have a look at these guys and tell me if any of them
jump out to you as 'Hey, that's useful.' / 'Hey, that's crap.'
- Speaker 2: Should we do it one at a time?
- Speaker 1: Sure. You go ahead.
- Speaker 2: I've... I've got a couple of fabulous ones.=I'll run through them. So,
from Thai tourism, 'Find your fabulous.' From California lottery, 'Go
directly to fabulous.'
- Speaker 1: Go directly.=
- Speaker 2: =ULTA. 'Welcome to fabulous.' And Mindtree, 'Welcome to
possible.'
- Page 32:**
- Speaker 1: [He... H]e would say that.

Speaker 2: Yeah, he would. Crisis actor.

Speaker 3: ((laughter))

Speaker 1: Hey, let's listen to a track. And this one iTAL tEK with 'Challenger Deep' on RTRFM 92.5.

((music))

Speaker 1: Lots and lots of great responses. Let's hear 'em. Aaron sent an email. 'It's the new promotional age. Exacerbate your tedious!' Seemingly, the popularization of grammatical dead ends. Good point. Garth sent me an email studio@rtrfm.com.au. 'This seems to be a common issue for me in my favorite discipline of motorsport, rallying. Slogans like 'I love rally!' used by the promoters of the World Rally Championship, no less. Shit me to tears! To me, you enter a 'rally' to go 'rallying' because you are a 'rallyist.'" Ooh, Garth is insisting on the morphology. I think that's really interesting. 'That's rallying' is a timeless quote and used to justify about any misfortune experienced in or around an event, but, some people are starting to say, 'That's rally.' **What is wrong with the world?** All right, well, remember, this is English we're talking about, sooo... It's not such a problem. This has been going on for thousands of years. Still says, 'One year, the Christmas time marketing slogan for Starbucks, where I worked was 'Let's merry!' **Like, it's Christmas time, merry, right?** If I didn't want to get fired, I would've called corporate and be like, **'Can you fire the entire marketing department?'** 'Kay. Thanks. Bye.' The marketing department is useless because they didn't make a drink called the covfefe.' Uh, he continues. 'It's not the word class changing that I object to.' That's good because if you had, you would've been in trouble when you said that you wanted to fire people. 'I just object to it sounding stupid.' N' I think that is actually a legitimate complaint that people think that if it sounds vacuous or if it sounds put-on, they mind that a lot more than they actually mind the shifting of lexical categories. So that's interesting. Mike gives me a poser. He says, 'Try expressing 'He was born in 1987 in active voice.'" ((mumble)) Okay, well it's got the 'was' and it's got the past participle 'born.' **But what is the main form? Would you believe? 'To bear.'** So, I guess... If we add that participant back in, 'He was born in 1987,' becomes 'His mother bore him in 1987,' which is... **You could see why that would be passive voice, right? Because the person who... I mean I know mothers are important, right? Come on. But...** The person who was born really is the subject of the sentence, and so I can see why passive voice sounds a lot better. So Mike is making a fantastic point. ((Khmhm.)) John on Facebook says, 'The only rule that matters in English is English will do what it wants when it wants with, but no prior precedent making a difference.' Case in point.

Arkansas English giving no f's about Kansas. I remember as a kid saying 'Are-Kansas' and being corrected. Why? Why is it a '-sa'? Silent 's,' weird thing. Matt chimes in: 'Rules I would like to happily ((laughter)) be rid of.' Nice infinitive split, Matt, I appreciate what you're doing. Love your work. Simon wanted to know if I could comment on this sentence: 'Colorless green ideas sleep furiously.' Ahh, yes. A very famous sentence by

Page 33:

ummm... Trotsky! Leon Trotsky, who was a secret linguist, but also a secret environmentalist who had lots of green ideas, but they weren't very exciting, so they were kind of colorless. Nobody liked them. Those ideas would have to wait for many more years during which time they would sleep. But they would arrive with a vengeance, so I guess these colorless green ideas must have been sleeping furiously. Good old Leon Trotsky. Let's take a moment and remember his linguistic achievements. That's all for today's episode of Talk the Talk. I'd like to thank you for listening. Thanks to Matt for taking us Out to Lunch very shortly. And... I would also like to exhort you to check out our Facebook page and we're doing great stuff on Patreon. Thanks for listening and until next time, keep talking.

Speaker 2: This has been an RTRFM podcast. RTRFM is an independent community radio station that relies on listeners for financial support. You can subscribe online at rtrfm.com.au/subscribe.

Speaker 3: Our theme song is by Ah Trees and you can check out their music on ahtrees.com and everywhere good music is sold.

Speaker 2: We're on Twitter @talkrtr, send us an email talkthetalk@rtrfm.com.au, and if you'd like to get lots of extra linguistic goodies, then like us on Facebook or check out our Patreon page. You can always find out whatever we're up to by heading to talkthetalkpodcast.com.

Appendix 4: Direct questions in “Grammar Day.”

I CANONICAL QUESTIONS

	Speaker	Question
1	Speaker 1	Should you be concerned about them?
2	Speaker 1	Is it okay for nouns to become verbs and vice versa?
3	Speaker 1	What's wrong with passive voice?
5	Speaker 2	Daniel, what's been going on in the world of linguistics in the week gone past?= =But when it comes to light immediately, as it often does, how did you know that that person didn't understand you?
10	Speaker 1	'Why don't you tell me when you don't understand me?'
12	Speaker 1	=But when it comes to light immediately, as it often does, how did you know that that person didn't understand you?
13	Speaker 2	=[What is the 'zed'?]
17	Speaker 1	Now, if you're a neuroscientist and you wanted to look for some kind of brain signal that someone had misunderstood, what would you look for?
18	Speaker 1	What do you think, Kylie?
19	Speaker 2	How'd you do it, how?
24	Speaker 1	And do you know what it does?
28	Speaker 3	Are we going to wear to like T-shirts and something or [come up wi]th slogans?= =What's, [what's your gripe?]=
35	Speaker 2	[Kylie], what is your actual problem though?= Doesn't your university do this as well with their slogan?= What is more anyway?
36	Speaker 2	[Kylie], what is your actual problem though?= Doesn't your university do this as well with their slogan?= What is more anyway?
42	Speaker 1	Doesn't your university do this as well with their slogan?= What is more anyway?
48	Speaker 1	What is more anyway?
51	Speaker 1	What is... what is 'more' as a part of speech?
54	Speaker 2	=Is it a verb...?
56	Speaker 1	So what is that doing?
58	Speaker 2	Why do you care?
62	Speaker 1	Are you ready?
63	Speaker 1	Think to yourselves, was this a verb first or a noun first?
67	Speaker 1	Which came earlier, 'You could access something' or 'you had access to something'?
70	Speaker 1	Which one came first?
72	Speaker 1	Do you like verb or noun, 'to quiz somebody' or 'to have a quiz'?
73	Speaker 1	Do you think that you 'made a demand' or do you think that you could 'demand something' first?
74	Speaker 1	Which came first?
76	Speaker 1	Which came first?
84	Speaker 2	Who did the 'You can't start a sentence with 'because''?
86	Speaker 2	=What is a split infinitive?
88	Speaker 1	'Why do teachers hate passive voice so much?'
94	Speaker 1	Woo, what is passive voice?
98	Speaker 3	'is it... Is it active?'
99	Speaker 3	Is it action filled?
101	Speaker 2	[But...] But does it only sound floppy to you because you've had a series of people you respect tell you that that's the thi[ng?]
104	Speaker 1	What... What is this?
106	Speaker 1	What... What's the deal with that?

116	Speaker 2	Have we got some more?
123	Speaker 3	'Well, what would you say to the conspiracy theorists who think you're just a crisis actor?'
127	Speaker 1	But what is the main form?
132	Speaker 1	Why is it a '-sa'?

II NON-CANONICAL QUESTIONS

1) ASSERTIVE

	Speaker	Question
31	Speaker 1	You think so?

2) FRAGMENTS

	Speaker	Question
9	Speaker 1	Like [what?]
21	Speaker 2	'huhh?'
22	Speaker 1	'huhh?'
33	Speaker 2	Kylie?
55	Speaker 3	=Nominal?
61	Speaker 3	Do I...?
64	Speaker 1	Kylie?
66	Speaker 1	Noun or verb?
69	Speaker 1	I hate to inconvenience you,' the verb or, 'it's such an inconvenience,' the noun?
71	Speaker 1	'The inconvenience?'
75	Speaker 1	'Orange,' the color or 'orange,' the fruit?
77	Speaker 1	'Orange,' the color, or 'orange,' the fruit?
92	Speaker 1	Ready?
93	Speaker 3	You sure you don't want to rephrase it?'
100	Speaker 3	Are we seeing something happen that goes?'
108	Speaker 2	Ddo you...?'
110	Speaker 1	That being a crisis actor?'
119	Speaker 3	Oohh?
128	Speaker 1	Would you believe?'

65	Speaker 1	How about this one?
68	Speaker 1	How about 'inconvenience'?
80	Speaker 1	You know, why wouldn't they?
81	Speaker 1	I mean, why invent two separate words for closely related things?
95	Speaker 1	Why should I...?
102	Speaker 1	You know what?
103	Speaker 1	And you know what?
121	Speaker 1	Why they love them so much?
131	Speaker 1	Why?

3) RHETORICAL/ABILITY/INCLINATION

	Speaker	Question
4	Speaker 3	[=An]d wasn't that beautiful?
6	Speaker 1	May I ask you a question?
7	Speaker 1	What does it mean to understand language?
8	Speaker 1	Can you tell when someone has misunderstood you?
14	Speaker 3	Is it something they forgot to tell us?
23	Speaker 3	=Is that the frontier at the front...?= =
25	Speaker 1	'Would you like some coffee?'
27	Speaker 2	Should we take a track?
30	Speaker 1	Can a wooden house be dilapidated?
32	Speaker 1	Would you like to-to...?
34	Speaker 2	Did you get your grammar grouch on?
37	Speaker 1	Can you describe what is wrong with this grammatically?
38	Speaker 1	'It's a solid run, but is it enough to podium?'
45	Speaker 2	Should we do it one at a time?
47	Speaker 2	'Would you let me see beneath your beautiful?'
50	Speaker 1	Is it Roger Moore?
59	Speaker 3	Must I really 'free my think' or...?
82	Speaker 1	Remember, if you have any questions or comments about anything you hear, why don't you get those to us?

85	Speaker 2	Would you've said that he Lowthed it?
91	Speaker 2	Can you explain it to me?
96	Speaker 2	Do you know what it's like for me?
105	Speaker 2	Ooo, must [we?]
107	Speaker 1	'Can you tell me right now that you will not accept a single donation from the National Rifle Association?'
120	Speaker 3	Aahh, do you think we'll end up having lizard people being to blame?
124	Speaker 1	What is wrong with the world?
126	Speaker 1	Can you fire the entire marketing department?

4) TAG

	Speaker	Question
11	Speaker 2	That's the thing, right?
16	Speaker 2	To be fair to Americans, it does sound like an off-brand bad guy from a superhero film, right?
26	Speaker 2	Right?
29	Speaker 2	Right?
40	Speaker 1	So ((chuckling)) 'Pod-', 'podijum' is a noun and it's a thing, right?
43	Speaker 2	Well, it's vaguely poetic, isn't it?=-
44	Speaker 2	Because 'a scarf,' right, is such a idiosyncratic noun, right?
49	Speaker 2	'More' is sooo vague though, isn't it?
52	Speaker 3	It's 'amounts,' isn't it?
78	Speaker 3	Really?=-
79	Speaker 3	They come a lot later, don't they?
83	Speaker 1	It's weird, isn't?
87	Speaker 2	Right?
89	Speaker 1	Really?
90	Speaker 3	Real[ly?]
97	Speaker 2	Right?
109	Speaker 2	We know that this is an actual thing, right?
111	Speaker 2	So what happens, is in a big basketball stadium, which seats like 15,000 people, the central court gets taken away, right?
112	Speaker 3	Really?

113	Speaker 2	All the lights go down, right?
114	Speaker 2	Oh really?
115	Speaker 3	Really?
117	Speaker 2	Aaah, now this goes back quite a way, doesn't it?
118	Speaker 2	Well, 'false flagging' is like Age of Sale stuff, is[n't it?]
122	Speaker 2	=And some of them are bloody crisis actors, you know what I mean?
125	Speaker 1	Like, it's Christmas time, merry, right?
129	Speaker 1	You could see why that would be passive voice, right?
130	Speaker 1	I mean I know mothers are important, right?

5) WH

	Speaker	Question
15	Speaker 3	You know, in quantum physics, what?
20	Speaker 3	What?
39	Speaker 3	What we see?]
41	Speaker 3	[What]s up?
46	Speaker 2	Ooh, you know what?
53	Speaker 1	So what part of speech does that...?
57	Speaker 2	Whyyy?
60	Speaker 1	You know what?
65	Speaker 1	How about this one?
68	Speaker 1	How about 'inconvenience'?
80	Speaker 1	You know, why wouldn't they?
81	Speaker 1	I mean, why invent two separate words for closely related things?
95	Speaker 1	Why should I...?
102	Speaker 1	You know what?
103	Speaker 1	And you know what?
121	Speaker 1	Why they love them so much?
131	Speaker 1	Why?

RESÜMEE

TARTU ÜLIKOOL
ANGLISTIKA OSAKOND

Kärt Roomäe

The Analysis of Canonical and Non-canonical Questions in an English Language Podcast

Kanooniliste ja mittekanooniliste küsimuste analüüs ingliskeelses *podcast*'is

bakalaureusetöö

2019

Lehekülgede arv: 55

Annotatsioon:

Käesolev bakalaureusetöö uurib otseseid küsimusi keeleteadusliku *podcast*'i „Talk the Talk“ episoodis „Grammar Day“. Töö eesmärgiks on kindlaks teha, kuid erineb töö korpuses olevate küsimuste lauseehitus ja missugune on nende funktsioon, võttes arvesse, et tegu on loomuliku, suhteliselt mitteametliku suhtlussituatsiooniga. *Podcast* on küllalt uus meedium ja *podcast*'i kui keeleteaduslike andmete allikat käsitlevaid uurimusi pole seni arvukalt esinenud. Lisaks esindab töös kasutatud küsimuste tüpoloogia uut lähenemist teoreetilises lingvistikas.

Töö jaguneb kahte ossa: esimene, teoreetiline osa selgitab küsimuste liigitamist kanoonilisteks ja mitte-kanoonilisteks, misjärel antakse ülevaade eri autorite poolt välja pakutud mitte-kanooniliste küsimuste alajaotustest ning antud töös kasutatud jaotust. Töö empiirilises osas analüüsitakse otseseid küsimusi ning nende kategoriseerimist.

Töös analüüsitud 132 küsimust on pärit kolmveerandtunnise (44:58) *podcast*'i episoodi anoteeritud transkriptsioonist, mis on 33 lk ja 10 031 sõna pikk. Küsimused jagati esmalt kahte suurde kategooriasse: kanoonilised ja mitte-kanoonilised küsimused. Seejärel keskenduti teisele kategooriale ning jaotati mitte-kanoonilised küsimused omakorda viide alakategooriasse vastavalt nende süntaktilistele, semantilistele ja pragmaatilistele omadustele. Selgus, et osalejad eelistasid *podcast*'ile omases suhteliselt vabas õhkkonnas mittekanoonilisi vorme. Arvuliselt jagunesid 132 otsest küsimust 91 kanooniliseks ning 41 mittekanooniliseks, vastavalt 69% ja 31%. Mitte-kanoonilised küsimused jaotusid omakorda: 28 (31%) küsijätkega küsimust, 26 (29%) retoorilist/oskus-/kallutatud küsimust, 19 (21%) fragmenteeritud küsimust, 17 (19%) wh-küsisõnaga küsimust ning 1 (1%) assertiivset küsimust.

Märksõnad:

Inglise keel, süntaks, pragmaatika, semantika, vestlusanalüüs, *podcast*

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Kärt Roomäe,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose

The Analysis of Canonical and Non-canonical Questions in an English Language Podcast,

mille juhendaja on Jane Klavan,

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

[allkiri]
Kärt Roomäe

Tartus, 24.05.2019

Autorsuse kinnitus

Kinnitan, et olen koostanud käesoleva bakalaureusetöö ise ning toonud korrektselt välja teiste autorite panuse. Töö on koostatud lähtudes Tartu Ülikooli maailma keelte ja kultuuride kolledži anglistika osakonna bakalaureusetöö nõuetest ning on kooskõlas heade akadeemiliste tavadega.

[Autori allkiri]
Kärt Roomäe

Tartus, 24.05.2019

Lõputöö on lubatud kaitsmisele.

[Juhendaja allkiri]
Jane Klavan

Tartus, 24.05.2019