1 INTRODUCTION

It is safe to say that most students of Arabic become aware of a particular contradiction regarding the use of case markers in the formal register Standard Arabic. The contradiction lies between what *should be* and what *is*. That is, between how the system with case markers *should be* used according to the codified grammar, and how it actually *is* used in real life. Many students in Western universities have text books, according to which the system of case markers should be used. But they usually notice that in reality, it very rarely is used. What we have here is a tension between a prescriptive and a descriptive grammar of Standard Arabic, one based on a codified norm, and the other on an empirical norm. The highly prescriptive Arabic grammar has not changed much since its codification in the Middle Ages, and is still the grammar condoned by education systems in the Arabic speaking world. The system with case markers is particularly cherished and has been described as synonymous to grammar as a whole (Haeri, 2003, s. 40). In contrast to this, however, a correct use according to prescriptive grammar is rarely to be found in real life use of Standard Arabic, and it might even be considered inappropriate in many situations (Hallberg, 2019, p. 15).

Previous research investigating the discrepancy between prescriptive and descriptive Arabic grammar has typically focused on the broader issue of Standard Arabic versus the spoken dialects. The use of case markers, however, constitutes a register *within* Standard Arabic, and as such it has not been as widely researched. Studies have indicated that a vast majority of native speakers lack the proficiency needed to use case markers according to the prescriptive grammar (Parkinson, 1993), and that speakers actually risk coming across as ridiculous if they do use them to any large extent while speaking Standard Arabic (Ferguson, 1959). The most comprehensive study to date of the use of case markers in spoken Standard Arabic is Hallberg (2016), which analyses 17 televised interviews with highly educated native speakers, totalling 5 hours and 22 minutes of extemporaneous speech. The results show that case markers were only pronounced in 7.5% of the potential cases. Hallberg (2019) identifies common patterns of how case markers are used in extemporaneous speech, revealing a descriptive grammar of the use and non-use of case markers, with "a set of conventions, non-codified grammatical rules, governing where case is and is not to be marked" (Hallberg, 2019, p. 24). However, Hallberg's study only includes extemporaneous speech, and he points out another important category of spoken Standard Arabic, of which there are no linguistic studies to date: reading aloud. The vast majority of texts in Standard Arabic, e.g. news articles and novels, are undiacritized, that is omitting almost all of the case markers. An Arabic speaker who reads such a text aloud is therefore left to decide the extent to which case markers are to be pronounced. Hallberg (2019) argues that it is reasonable to assume that a person reading aloud would display a similar use of case markers as in extemporaneous speech, but he points out that there is little direct empirical evidence at hand (Hallberg, 2019, p. 17).

Therefore, the purpose of this study is to provide such direct empirical evidence, as a way to fill the gap identified by Hallberg. This is done by conducting an empirical study of native speakers reading undiacritized texts in Standard Arabic aloud. The main aim is to contrast the results to those in Hallberg's studies of extemporaneous speech. Are case markers pronounced to the same extent, and following the same patterns of use and non-use? Since the use of case markers are such a central part of the prescriptive grammar, it is probable that most native speakers would try to maximize their use, would they be aware that this is what is being observed. It would indeed be relevant to conduct such a study, as it might show the participants' actual abilities to use case markers, as was done in Parkinson's (1993) study. This study, however, aims at observing the use of case markers when native speakers are reading aloud in a situation resembling everyday use of Standard Arabic as much as possible.

The study is thus designed in a way that the participants' focus stays as much as possible on the content of the texts that are read aloud, rather than on linguistic form. The overall question asked in this study is therefore: To what extent, and according to what patterns, do native Arabic speakers use case markers when reading a text in Standard Arabic aloud, while focusing on content rather than on linguistic form?

This study might contribute to approaching a more refined descriptive grammar of spoken Standard Arabic; How the system of case markers actually *is* used when native speakers read texts aloud with focus on content, and not on language form, rather than how it *should be* used according to a prescriptive grammar. As Hallberg (2019) points out, this can in turn be valuable for the development of curricula for teaching Arabic, and setting "relevant and realistic proficiency aims" (Hallberg, 2019, p. 29) for students of Standard Arabic.

2 BACKGROUND

This chapter gives a background to different aspects of the case system in Standard Arabic. The first section gives a general linguistic overview of the system and how it functions according to the prescriptive grammar. The second section puts the case system in perspective in terms of different varieties of Arabic, arguing that it is part of an exceptionally elevated register *within* the already formal variety Standard Arabic. The third section gives a brief account of the role of the case system in the instruction of Standard Arabic, making the point that there is an ambivalent view of its importance at Western universities. These three sections serve to explain what the case system is, and to motivate the relevance of this study.

Transliterations of Arabic letters in this study follow the literary system "Litt" (Isaksson). This system is usually recommended for studies in literature, rather than for linguistic studies, since it is more accessible for readers without knowledge in Arabic. Even if this is indeed a linguistic study, it is sufficient to use this less detailed representation in order to understand and follow the examples given.

2.1 The case system in Standard Arabic

The case system in Standard Arabic mainly consists of short vowels that are added as endings to nouns, adjectives and adverbs through diacritic signs. In the Arabic grammar, it is part of the bigger system *iʿrāb*, which also includes endings on verbs according to different moods, using the same diacritic signs. This study, however, only deals with the case system, since it would be too cumbersome to include both case and mood, and since it can be assumed that use of case and mood play out in similar ways. Some categories of words are inflected for case by way of orthographical variations in spelling of the word's ending, rather than by adding diacritics. These endings are then visible as letters in words in written texts, and there is no reason to think that they would not be pronounced when reading aloud. By contrast, the diacritic signs are only rarely included in written texts. It is these diacritics that are in focus in this study, since their absence leaves it up to the reader to pronounce them or not. When the term "case marker" is used throughout this study, it is always these latter markers that are referred to. The diacritic signs correspond to the vowels *-u*, *-a*, and *-i*, with an option to add the sound of a final -n for indefinite words, creating the endings *-un*, *-an* and *-in*, or so called nunation. The Arabic names for the diacritic signs are *ḍamma*, *fatḥa* and *kasra*, and the term for nunation is *tanwīn* (Table 1). There is also a diacritic sign called sukūn which indicates that no vowel is to be pronounced.

Name of diacritic sign	Diacritic sign	Phoneme
ḍamma	◌ُ	*-u*
fatḥa	◌َ	*-a*
Kasra	◌ِ	*-i*
ḍamma tanwīn	◌ٌ	*-un*
fatḥa tanwīn	◌ً	*-an*
kasra tanwīn	◌ٍ	*-in*
sukūn	◌ْ	-

Table 1.1 Diacritic signs in Standard Arabic

There are three cases in the Arabic case system: Nominative, accusative and genitive, indicating syntactic functions and relationships to other words. Most Arabic words have the possibility to take different endings for any of these three cases, and are therefore labelled triptotes. If the word stands in the indefinite, it takes nunation. Table 1.2 shows a paradigm for case markers for the triptote word *bayt*, which means house.

Case	Definite		Indefinite	
Nominative	*al-bayt**u***	البيتُ	*bayt**un***	بيتٌ
Accusative	*al-bayt**a***	البيتَ	*bayt**an***	بيتاً
Genitive	*al-bayt**i***	البيتِ	*bayt**in***	بيتٍ

Table 1.2 Cases in Standard Arabic

The case markers are commonly called case endings, and in most words, they do indeed appear as word-final endings, such as in the paradigm above. However, there is a common exception to this: words with possessive suffix pronouns. In Arabic, possession is not always expressed with a separate word as in English, but with a suffix attached to the noun. For example, the suffix for "her" is *-ha*, and so the phrase "her house" is expressed with the word *bayt* + *ha*. In spoken Arabic this could just be pronounced *baytha*. According to the codified grammar of Standard Arabic, however, the correct case marker for the word *bayt* must be inserted before the suffix. Hence, if the case is nominative the correct pronunciation is *bayt-u-ha*, with the case marker for nominative, the *ḍamma* pronounced as *-u*, inserted before the suffix. Note that this is still only one word with an attached suffix pronoun: *baytuha*. Words with possessive suffix pronouns thus constitute a special category of words, since they have internal case markers rather than word-final endings. This might have effects on the extent to which they are pronounced, as we will see later.

Case	Noun with suffix pronoun	
Nominative	*bayt**u**ha*	بيتُها
Accusative	*bayt**a**ha*	بيتَها
Genitive	*bayt**i**ha*	بيتِها

Table 1.3 Possessive suffix pronouns with case markers

At a first glance, the Arabic case system might appear as uncomplicated. However, there are a number of exceptional declension categories that deviate morphologically from words following the triptote declension described above. Two of these categories are diptotes and words in the sound feminine plural. These two categories get special attention in this study, since they both depend on diacritic signs to mark case, but do so in a way that deviates from the normal triptote words. They can thus be said to have irregular inflection for case, making it potentially more difficult for a reader to pronounce the correct case marker when reading a text that has no diacritic signs. The diptote category takes its name from the fact that it only has two different case markers when indefinite, so that it cannot take *kasra* (Ryding, 2005, p. 167). Moreover, it does not take nunation. The diptote category consists of words with a number of different morphological patterns that need to be learnt by the speaker in order to identify a word as belonging to the category.

Case	Definite		Indefinite	
Nominative	*al-manāṭiq**u***	المناطقُ	*manāṭiq**u***	مناطقُ
Accusative	*al-manāṭiq**a***	المناطقَ	*manāṭiq**a***	مناطقَ
Genitive	*al-manāṭiq**i***	المناطقِ	*manāṭiq**a***	مناطقَ

The declension category sound feminine plural consists of words ending with *-āt* in the plural, and are thus more easily recognized than diptotes. Words in the sound feminine plural cannot take *fatḥa*, neither in definite nor indefinite words.

Case	**Definite**		**Indefinite**	
Nominative	*al-qaṭarāt**u***	القطراتُ	*qaṭarāt**un***	قطراتٌ
Accusative	*al-qaṭarāt**i***	القطراتِ	*qaṭarāt**in***	قطراتٍ
Genitive	*al-qaṭarāt**i***	القطراتِ	*qaṭarāt**in***	قطراتٍ

There is in the prescriptive grammar a common exception from the obligation to pronounce case markers: words before a pause, *waqf*. According to traditional descriptions of Arabic, final short vowel and nunation is to be omitted in recitation, a rule that is also applicable to spoken Standard Arabic (Hallberg, 2016, p. 245). This implies that case markers should not be pronounced for words that occur before a pause. What is otherwise seen as insufficient use of case markers (to read words aloud without case markers) is in the *waqf* position not only permissible, but necessary. To pronounce a case marker in the *waqf* position is to commit a mistake, according to prescriptive grammar. How, then, is a pause position defined? Traditional grammar provides detailed rules for *waqf* in Quranic recitation, *tajwīd*, but there is little information about when a pause is otherwise condoned (Hallberg, 2016, p. 74). In theory, it is possible for a speaker to omit all case markers but produce an audible pause after each word, and then claim that the absence of case markers was a correct rendition of the text, according to grammatical rules. In reality, it might raarely be taken to that extreme, but the rules about the *waqf* position "leaves room for interpretation" (Hallberg, 2016, p. 244) of when it is to be applied.

2.2 Use of case markers in different varieties of Arabic

An extensive use of the Arabic case system is a skill mastered mainly by a small elite of scholars and specialists in Arabic grammar (Ryding, 2005, p 167). An important factor to keep in mind, in order to understand the challenges that the case system poses for the average native speaker, is the fact that it is almost completely absent in the spoken dialects. The linguistic situation in the Arabic speaking world is often described with the word diglossia, coined by the socio-linguist Charles Ferguson (1959, p. 336). This means that the mother tongue for people in the Arabic-speaking world is a spoken variety, such as the Levantine or Egyptian dialects. The formal Standard Arabic is then learned through formal instruction throughout the educational system, where only Standard Arabic is deemed appropriate to teach. The case system is thus present only in Standard Arabic, but even there it is often left unused. Case markers are written mainly in religious texts such as the Quran and the Bible, and in children's books (Hallberg, 2016, p. 76). In most texts encountered in everyday life, such as newspapers, magazines and novels, there are no, or very few, case markers. A substantial use of case markers is therefore not only reserved for the formal variety Standard Arabic – it also constitutes a higher register *within* Standard Arabic. It is possible, and indeed very common, to read, write and speak Standard Arabic without using case markers, or using very few of them. Figure XX shows the phrase "I want to go to the house" in three different registers:

1. The Levantine dialect
2. Standard Arabic without case and mood markers
3. Standard Arabic with case and mood markers

As can be seen, the main shift in register is the one from the spoken dialect to Standard Arabic which has distinctly different words for the verbs "want" and "go". Then, there is the next shift which is to add the case and mood markers. Note that if this phrase would be followed by a pause, the *kasra* case marker *-i* on the last word, *al-bayti*, must not be pronounced, according to the rules about *waqf*. The shift between Standard Arabic without and with case markers might seem like a small one, since it only consists of adding the mood and case markers. It should be remembered, though, that the example phrase given here is extremely simple. Even normal language use would include much more complex constructions, including use of declension categories with irregular inflection, such as diptotes and the sound feminine plural mentioned above. A fully correct use of case markers thus demands that the speaker apply grammatical rules while using a register which normally doesn't include them, and which is in and by itself highly different from the mother language, which is the spoken dialect. This is not an automatic skill, as might be the case in other languages such as German, which also has a system of case markers. Rather, it takes formal education and is a "rigorous task, even for educated native speakers" (Ryding, 2005, p. 167). One of the reasons for this is the linguistic distance between the native languages and the highest register of Standard Arabic with case and mood markers, as exemplified in figure X.

Levantine dialect	ʾana I	biddi want		ʾarūḥ go	ʿa to	l-bēt the house	أنا بدي أروح ع البيت
Standard Arabic without case/mood markers	ʾana I	ʾurīd want	ʾan to	ʾadhhab go	ʾilā to	l-bayt the house	أنا أريد أن أذهب إلى البيت
Standard Arabic with case/mood markers	ʾana I	ʾurīdu want	ʾan to	ʾadhhaba go	ʾilā to	l-bayti the house	أنا أريدُ أن أذهبَ إلى البيتِ

2.3 The role of case markers in the instruction of Standard Arabic

The system with case markers has enjoyed a central position in the Arabic grammar tradition from its inception in the Middle Ages until today's curricula in the Arab world. Its role has been cherished to the point where it is sometimes described as synonymous to the word grammar itself: "Indeed, grammar as a whole for most people comes to mean the case endings" (Haeri, 2003, p. 40).

The Arabic grammar tradition harks back to the first centuries after the life of the prophet Muhammad, and was systematized and codified in the period until approximately the year 1100 CE. After this followed a period of elaboration of the way in which grammar treatises were presented in order to transmit the codified grammar. From this time on, there

would be no more significant changes to the grammatical theories or ways of systematization. Treatises like *sharḥ al-mufaṣṣal* by Ibn Yaʿīsh (d. 1245 CE) are still used as standard textbooks for university level grammar instruction in the Arab world (Bohas, 2006, p 16). Any grammatical treatise would start with a chapter establishing the different word classes, and then going straight into an elaborate and detailed presentation of the rules of *iʿrāb* with case markers being in the focus (Bohas, 2006, p. 54). At the core of Arabic syntax lies the theory of government, referring to the idea that there are governing words that often cause other words to take a certain case or mood ending (Bohas, 2006, p. 50). The theory thus revolves around ways of explaining and codifying how case and mood markers are to be distributed, elevating them to what is perhaps the most prevalent feature of Arabic grammar. If the use of case markers is taken for granted as a central part of grammar instruction in the Arab world, Western universities have a more ambivalent view. A study (Wennerholm, 2019) showed that two of the most widely used textbooks at Western universities have opposite views on the use of case markers. The book *Standard Arabic - An elementary-intermediate course* (Schulz et al. 2000) demands that students use them in most parts of the book, and include case markers in grammatical presentations and exercises. On the other hand, the book *Al-Kitaab fī taʿallum al-ʿarabiyya* (Brustad et al. 2011) deems it inappropriate for the elementary- and intermediate-level students to learn how to use case markers, and reserves it for later stages. The same study also showed that three out of four Swedish universities place great importance on the use of case markers already during the first semester of instruction in Standard Arabic, whereas one university didn't place any importance at all on this during the first semester (Wennerholm, 2019).

In light of this background, and particularly the different views on the use of case markers for students of Arabic as a foreign language, there might be a need for empirical studies to serve as basis for future decisions on curricula at Western universities. In countries such as Sweden, where significant numbers of Arabic native speakers live, it can be informative to contrast the Arabic instructed at the universities with the language used in real life situations by the students' potential conversation partners. As mentioned, it is already a common observation that case markers are often left out even in formal Standard Arabic, and that an extensive use is reserved for specialists in Arabic. It will therefore come as no surprise that the results of this study show a limited use when participants read texts aloud while focusing on the content rather than the language form. Nevertheless, there are no empirical studies to date that establishes this as an empirical fact, nor what patterns there are of use and non-use of case markers when reading aloud. This study might serve as a small but useful part of a larger effort to build a more solid empirical foundation on which to formulate a descriptive grammar of the use of case markers, which can then inform curricula for learners of Standard Arabic as a foreign language.

3 THEORY

This chapter gives a theoretical framework to the use of case markers in Standard Arabic. The first section introduces the standard language ideology, and argues that its resulting prescriptive grammar fails to identify variations in the use of case markers, and that an identification of empirical norms is required in order to understand the use of case markers in reading aloud. The second section describes the most comprehensive study to date on this empirical norm, a dissertation by Hallberg (2016). The third section goes on to identify a gap in empirical studies, namely the use of case markers when reading aloud.

3.1 Standard language ideology and linguistic variation

The term standard language ideology is not specific for Arabic, but also for views on a wide range of languages such as English, French and Spanish (Milroy, 2006, p. 133). The main idea behind the ideology is that there exists one correct variety of a language, which is imposed on its speakers. This standardization means that only one form of the language is seen as correct, while other variations are deemed wrong. An English-speaking person who writes or says *you were* is considered to be right, while somebody who says *you was* is considered to have committed a mistake (Milroy, 2006, p. 133). However, as much as this ideal language might be perceived as the only correct one, it is never achieved in practice. In written language it might come closer to the ideal, but in spoken language it's usually further away. Therefore, the standard language ideology, according to Milroy, is an idea in people's mind, and not something that can actually be achieved in the everyday reality of language use (Milroy, 2006, p. 134). This creates the tension mentioned in the introduction, between what *should be* and what *is*. On the one hand, there are a set of rules that should be followed, but on the other hand there are variations in how a language is used in the real world of human communication. A key word here is variation. There can only be one standard language, with one set of common rules, and there is no room for language variation in the standard language ideology: "the ideology of standardisation is inimical to change and variation" (Milroy & Milroy, 1991, p. 26). When the standardisation of a language has reached a certain point, a codification takes place, for example in eighteenth century England, when dictionaries and grammar-books were published (Milroy & Milroy, 1991, p. 27). As a natural consequence of this comes prescription, an imposition to follow the codified grammar (Milroy & Milroy, 1991, p. 27). There is, thus, a prescriptive grammar which is codified and represents an ideal form of language use, while at the same time broad layers of speakers normalise a use that deviates from codified rules. This normalisation creates certain norms as to how the language should be used in these real-life situations, but these norms are not codified in prescriptive grammar and can only be traced empirically. The result is two competing norms: A codified norm and an empirical norm (Hallberg, 2016, p. 46).

In the context of this study, this model fits well to describe the use of case markers in Standard Arabic: It is clear that a full use is a central part of the codified norm, and it is equally clear that there is only a limited use according to the empirical norm. The codified norm gives a prescriptive grammar with rules on how to use case markers, but it is not equipped to deal with variation in real life use. As mentioned in the background section, it is possible to speak Standard Arabic without, or with few, case markers. If a speaker would use case markers to a full extent when speaking Standard Arabic, then it would break the empirical norm and the speaker would risk being seen as pretentious or overly formal (Hallberg, 2016, p. 54). But what is this empirical norm? Even though it is common knowledge that speakers only use case markers to a very limited extent in Standard Arabic, this is not to say that there is no use at all. And obviously, terms like "common knowledge"

and “limited extent” don’t provide any evidence about the empirical norm regarding use of case markers. Hence, the empirical norm must be identified by means of empirical studies, which is the purpose of the present study. Variation in use of case markers implies that there might be different patterns of use and non-use that can only be discerned by observing real life use. By doing this, steps can be taken to describe the empirical norm for the use of case markers. In light of this theoretical framework, an underlying research question can be formulated for the present study: What grammar is needed to read Standard Arabic aloud according to the empirical norm?

3.2 Hallberg’s study

The present study follows in the footsteps of a dissertation by Hallberg (2016) that deals precisely with the empirical norm for the use of case markers in spoken Standard Arabic. The dissertation is the most comprehensive study to date on the use of case markers, and together with a following article by Hallberg (2019), it is the main source of inspiration for this present study. Hallberg’s dissertation analyses extemporaneous speech in Standard Arabic in 17 televised interviews with highly educated native speakers, totalling 5 hours and 22 minutes of speech. The results show that case markers were pronounced in 7.5% of the potential cases (Hallberg, 2016, p. 168). This overall finding thus confirms the general observations in older studies, namely that case markers are only used to a small extent in extemporaneous speech in Standard Arabic. Apart from the main finding that there is only a sparse use of case markers in extemporaneous speech in general, Hallberg identifies more in detail when case markers were used in the analysed corpus of interviews. He sees five patterns of how case markers are used, and not used, in extemporaneous speech:

1. **Preference for nouns with enclitic pronouns.** (In the present study, the term used for enclitic pronouns is suffix pronouns, as explained in the background section.) In Hallberg’s study, these words have a 43,9% rate of case marking (Hallberg, 2016, p. 211), which is by far the highest rate for any of the investigated categories. As a possible reason for this Hallberg (2016, p. 209) mentions the lack of pause form for words with enclitic pronouns, preventing speakers from justifying a non-use of case markers by treating them as standing in the pause form.
2. **Almost completely missing for words with definite article.** With a rate of 0,3%, words with the definite article are characterized by an almost total absence of case markers in the study.
3. **Preference when case would be orthographically represented in text.** Some case markers are usually always written out in texts, for example for masculine singular in the accusative. In extemporaneous speech, these markers tend to be pronounced, just as they are usually visible in texts. This category of markers is not present in the present study, which only includes case markers that are not visible in the tested texts
4. **Omission in pause position.** The corpus text was segmented into utterances, and pause defined as the end of an utterance. The result shows a predicted sparse use of case markers in pause form of 1,1% (Hallberg, 2016, p. 245). Of the marked words in pause form, all except two have orthographic case endings, and are thus in and by themselves likely to be marked for case, regardless of position (Hallberg, 2016, p. 245).
5. **Very sparse use in all other positions not mentioned above.**

3.3 The lack of empirical studies of reading aloud

Hallberg's study gives a thorough analysis of the use and non-use of case markers according to the empirical norm, rather than the codified norm, that is, in real life language production as opposed to an imagined ideal according to prescriptive grammar. It is however limited to a certain context: extemporaneous speech by highly educated speakers giving television interviews. In a subsequent article, Hallberg (2019, p. 5) identifies three contexts of oral production of Standard Arabic:

- Extemporaneous speech
- Recitation
- Reading aloud

Recitation is defined as "the verbatim, oral reproduction of a text from memory" (Hallberg, 2019, p. 18). Typically, these are religious texts such as the Quran or poetry, that have full case marking in their written form. Recitation is of course closely connected to the written source, and therefore the case markers are memorised without the reader necessarily being aware of their grammatical functions. The case markers are such a natural part of recitation that omitting them would "strike many Arabic speakers as absurd" (Hallberg, 2019, p. 19). Reading aloud of undiacritized texts, on the other hand, can occur either with or without pronouncing case markers. In formal situations such as news broadcasts and speeches, they would typically be pronounced, and reading aloud with case markers is also the natural linguistic form for instruction of Standard Arabic in the Arab world and in many Western universities (Hallberg, 2019, p. 15). This is however a register that most speakers are only rarely required to use. Rather, the default form for reading an undiacritized text in Standard Arabic aloud is uninflected reading, that is without pronouncing case markers (Hallberg, 2019, p. 14). This includes for example reading something from a news article aloud to a friend when the focus of attention is the content of the text rather than linguistic correctness (Hallberg, 2019, p. 14). Moreover, according to Hallberg, uninflected reading aloud is probably an oral manifestation of silent reading (Hallberg, 2019, p. 14). In light of the centrality that is being ascribed to this register, Hallberg (2019, p. 17) notes that it is "problematic" that there are to the best of his knowledge no linguistic studies providing direct empirical evidence of how case markers are being used while reading aloud. In lack of this evidence, he hypothesises that it is done in similar ways as in extemporaneous speech:

> In both extemporaneous speech and inflected reading aloud the speaker/reader has to rely on their own proficiency in the grammatical system to compute and enunciate diacritic case markers, and it is therefore reasonable to assume that case marking in uninflected reading aloud plays out in similar ways as in extemporaneous speech, on which we have better knowledge. (Hallberg, 2019, p. 17)

This study sets out to fill the gap and provide empirical evidence of use of case markers in the context of reading undiacritized texts in Standard Arabic aloud, with focus on language content rather than linguistic form, similar to Hallberg's example with a person reading parts of a news article aloud to a friend. One of the aims will then be to contrast the results to those in Hallberg's study as described above, and to test his hypothesis that the use of case markers plays out in in similar ways as in extemporaneous speech.

4 METHOD

The method used in this study consisted of recording the sound of twenty native Arabic speakers reading texts in Standard Arabic aloud, and then noting every case marker they pronounced. The texts were taken from websites of state and non-state actors giving practical information about the coronavirus pandemic 2020. The recorded sound was used to analyse participants' use of case markers with a quantitative approach. A total number of 184 words were tested. Participants for the study were recruited among native Arabic speaking students in a school teaching Swedish For Immigrants (SFI) in Göteborg, Sweden, during the spring of 2020. This method serves to answer the overarching question to the study: To what extent, and according to what patterns, do native Arabic speakers use case markers when reading a text in Standard Arabic aloud, while focusing on content rather than linguistic form? The latter part of this question, "while focusing on content rather than linguistic form", is central to this study, and has had far-reaching consequences for the design of the method. It would of course have been possible to simply give participants a text to read aloud and tell them that the objective is to study their use of case markers. This would however have risked to encourage a strong focus on linguistic form rather than content, with participants trying to pronounce case markers according to prescriptive grammar. Given that the pronunciation of case markers has such a central and cherished role in formal education, there is an obvious risk that the collected data would have been the result of speakers trying to live up to an ideal register, rather than an expression of natural reading aloud with focus on the text's content. Hallberg (2019, p. 17) predicts that a study of speakers reading texts aloud will "constitute a difficult case of Observer's paradox". Consequently, an effort has been made to design a method that diverts the attention away from linguistic form and directs it to the text's content as much as possible.

4.1 Collection of data

The way the data was collected will here be described in seven points. First, I will give a presentation of the participants, and then of the texts. The third point elaborates on the preparations of the texts. This point includes a detailed account of how tested words were coded with different tags, using Hallberg's methodology. The fourth point presents two categories of words (words in *waqf* position and words preceding *hamzat al-waṣl*) that are being excluded from the study, even though they can take case markers. The fifth point describes how the texts were manipulated before the study, and the sixth points gives a detailed account of the procedure during the interviews. Finally, a seventh point deals with ethical considerations.

4.1.1 The participants

The participants in the study were recruited among native Arabic speakers in classes in Swedish For Immigrants (SFI), study path 3, at a school in Göteborg. The fact that I sometimes work there as substitute teacher helped in contacts with the school. The option to study SFI is open for all immigrants from 20 years of age who have a registered social security number in Sweden and don't have Swedish as a native language (Göteborgs Stad, 2020). When registering for SFI, all applicants attend an interview in the municipality's central administrational office for adult education. During the interview they are asked about their educational background, and directed to a suitable "studieväg", study path, in order for people with similar educational backgrounds to end up in the same classes. There are three

study paths: Study path 1 for people with none or short school background, study path 2 for people with six to nine years school background, and study path 3 for people with 10 years or more of school background (Göteborgs Stad, 2020). The recruitment for participants for this study was done solely among persons enrolled in study path 3, since people in this category can be expected to have the best possibilities to use case markers.

The recruitment was done by means of an invitation in an e-mail written by me, which was sent to all registered persons with Arabic as native language registered in study path 3, in the school. The total number of recipients was 54 persons. In the e-mail I introduced myself as being substitute SFI-teacher, as well as student of Arabic at the university of Gothenburg, conducting a research study about the Arabic language. I invited the recipient to participate in a 30-minute online interview about short texts in Arabic dealing with current topics, and expressed my hope to get response from people who would like to participate. After additional mouth-to-mouth advertising from an Arabic speaking assistant teacher, a total of 20 persons agreed to participate. During the interviews for this study, additional information was collected about each participant: Gender, age, education level in the home country, and spoken Arabic dialect. The following information was obtained in that way:

- **Gender:** 11 men and 9 women
- **Age**: Between 20 and 52 years old. Average age 36 years.
- **Spoken dialect**: 12 Levantine, 3 Iraqi, 2 Palestinian, 2 Egyptian and 1 Tunisian.
- **Education**: 1 person went nine years in school, not reaching *thānawiyya* (Secondary Education certificate), 10 persons have *thānawiyya* certificate, and 9 persons have university degrees.

4.1.2 The texts

The six texts in the study were selected among texts with public information about the coronavirus published online by the Swedish Public Health Agency, The World Health Organization, The Swedish Red Cross and The Swedish Institute. The selection criteria were that the texts should be thought-provoking and easy to understand. Texts overburdened with facts or with bureaucratic content were avoided. Moreover, a variation of content between the selected texts was sought after, in order to maintain the reader's interest over the course of the reading. The texts give instructions on hygienic measures to avoid infection, debunk popular myths about the virus, give advice on how to cope mentally in a time of crisis, and provide information about the Swedish state's handling of the pandemic.

1. ***FM*** *- The Public Health Agency (Folkhälsomyndigheten)*
   Issued by the Public Health Agency, the text gives advice on how to avoid being infected or infect others. This includes an instruction to wash one's hand frequently, to use hand disinfectant and to cough and sneeze into one's elbow. The text contains 135 words in total, of which 32 words are tested.

2. ***WA*** *- The World Health Organization on spraying one's body with alcohol*
   This is the first of three selected texts from WHO's page "Advice for public: Myth busters". It denounces the misconception that spraying alcohol or chlorine on one's

body would help against the coronavirus. The text contains a total of 114 words, of which 35 words are tested.

3. ***WS*** *- The World Health Organization on applying sesame oil to one's skin*
   The second WHO text informs that applying sesame oil to one's skin does not prevent the virus from entering the body, but that certain chemical disinfectants can kill the virus if applied to surfaces. The text contains a total of 82 words, of which 28 words are tested.

4. ***WP*** *- World Health Organization on receiving parcels*
   The third WHO text affirms that it is safe to touch parcels and letters received by mail from areas with reported cases of Covid-19. The text contains a total of 66 words, of which 24 words are tested.

5. ***RK*** *- Swedish Red Cross (Röda Korset) - advice on how to cope mentally in a crisis*
   The text from the Swedish Red Cross gives a list of advice on how to reduce one's fears and anxieties in times of crisis. Among the advice are talking to friends about one's worries and not to consume news reports in an excessive manner. The text contains a total of 135 words, of which 38 words are tested.

6. ***SI*** *- The Swedish Institute (Svenska Institutet) on Sweden's handling of the crisis*
   The Swedish Institute is a state agency that serves to encourage interest and trust in Sweden around the world. The text gives brief information about how Sweden's government deals with the coronavirus crisis, for example that independent governmental bodies present recommendations the government which take decisions. The text contains a total of 88 words, of which 27 words are tested.

Several factors contribute to a focus on content rather than linguistic form in these texts. First of all, the subject itself can be assumed to attract attention to the content, since the interviews were carried out during the height of the pandemic in Sweden, in April to June 2020. Sweden lived through an unprecedented health crisis with parts of society shut down, and with several thousand deaths in Covid-19. There was also an ongoing and sometimes heated debate about Sweden's response to the pandemic, which many deemed too lax. The coronavirus crisis was probably one of the most discussed topics in all parts of society, and it had a considerable impact on the everyday life of virtually every inhabitant during the time of the study. Labov (1972, p. 92) states that one way to make a person use casual speech while being observed during an interview in a linguistic study, is to direct the conversation to the danger of death. When his interviewees talked about a situation where they thought they had been at risk of getting killed, they immediately changed to casual speech. Labov's study related to speech, and not to reading aloud, and it could be argued that short information texts about the coronavirus are not exactly conjuring up feelings of danger of death. Nevertheless, there is good reason to think that texts about the pandemic trigger associations to fears and worries about potential threats to different life projects. Labov's observation, then, supports the claim that the coronavirus pandemic as a choice of subject for texts appeals to emotional and intellectual responses that help directing focus to content rather than form.

During the interviews, it was noticeable that the texts' subject was at the centre of attention throughout the interviews. A few participants immediately aired some thoughts about the ongoing crisis. More than one person expressed concerns about their family members, one said his wife was worried he would bring the virus home, and she had him take many and

long showers. Another expressed worries about having experienced symptoms that no general practitioner would take time to check. Some expressed spontaneous comments even during the reading aloud of certain passages. One example was from the text WA which asks the question whether it helps against the coronavirus to spray the body with alcohol. Several participants halted after reading this claim aloud and laughingly exclaimed "Of course not!". One person noted with a smirk that the president of the USA had speculated that some similar measure would indeed help against the coronavirus. Another example was the text RK which gives advice on how to cope mentally in a crisis. This text was much appreciated by many of the participants, who spontaneously expressed this during, or after, reading it aloud. Especially the advice not to follow news excessively was commented upon in a positive way, with several persons referencing to certain friends or family members who would do best to follow this advice.

Furthermore, the objective of the selected texts is to transmit practical information to a broad public, not to exhibit a particular linguistic style. It is the kind of texts that could be found in brochures or on noticeboards, and one could imagine them being read aloud to a friend. All these factors back up the assumption that the selected texts will trigger a spontaneous focus on content, and not on linguistic form.

4.1.3 The interview procedure

The need for focus on content had one major implication for the design of the setting for the recording of the participant's reading the texts aloud: It was necessary that the participants be unaware of the fact that their use of case markers was being observed. Participants were invited to take part in a research interview about texts in Arabic dealing with the coronavirus pandemic. They were told that the research was part of a magister thesis in the Arabic language at the Department of Languages and Literatures at the University of Gothenburg, and that the topic of the research was the Arabic language used in texts about the coronavirus pandemic. In the interviews, they were asked to read the texts aloud and then answer some questions about each text, after it had been removed from their sight. The questions were of two kinds: Checking if the participants had understood the content and could recall some crucial parts of the information given in the texts ("What did the text say about such and such?"), and asking for the participants' opinions about the appropriateness of the Arabic language used in the text ("Do you think the Arabic in the text is appropriate?") The first category of questions served to direct focus to content. Before starting to read each text aloud, the participants knew they would need to answer questions about the given information, which served as an incentive to concentrate on the content rather than the linguistic form. Rather than a pressure to perform well according to formal grammar, there was a certain pressure to perform well in terms of giving correct answers to the questions asked. The second category of questions, about the participants' views on the appropriateness of the language used in the text, were meant to back up the credibility of the overall interview situation. Why would a language researcher from the institution of the Arabic language at the University of Gothenburg care to record the audio of an interview that only consisted of the participant reading aloud and answering fact questions? An interest in the participants' views on the language served to direct attention away from the part where the texts were read aloud, and towards an engaged conversation about the texts. It would signal that there was indeed some interest in the language from the test leader, but no sign that this interest would be focused on something like case markers. The reason given for the reading aloud, and for recording the sound of the interviews, was that it would help in order to analyse the interviews afterwards. It was stressed that for methodological reasons, it was important to prove afterwards that all participants had read exactly the same information before answering

the questions. It was also remarked to all participants that answers could be given in either Swedish or Arabic, even if the test leader does not understand everything in Arabic, since the audio recording would enable a later translation.

During the course of this study, the coronavirus crisis in Sweden took on unexpected proportions that had an impact on the study itself. The original plan was to meet participants in the school building where they study SFI and carry out the interviews there. However, in mid-March schools for adult education in Sweden closed down classroom teaching. Therefore, the interviews were, with six exceptions, carried out through online video calls rather than in real life. The texts were presented by screen-sharing, and the sound was recorded on my computer. Six interviews were carried out on location in the school building, due to the persons' expressed wish to participate in spite of the fact that they did not have the software needed to make video calls. These interviews were carried out as though they would have been done online. Sitting in the same room, I video-called a mobile phone from my computer and handed over the phone to the participant and then carried out the interview in the same way as the other online interviews.

As mentioned in the theory section, a desired target situation would be a person who in everyday life happens to see a thought-provoking headline and spontaneously reads it aloud to a friend. To some degree, this is an ideal goal that can never be fully reached in an interview situation where, after all, the whole setting signals that the participant is being observed. However, several factors indicate that the method was successful in encouraging a focus on content rather than linguistic form. Most participants seemed to quickly accept the premises of the interview. Typically, they would ask about the objective of the study at the very beginning of the conversation, after which I would read the introductory information, as described in Appendix X. Most participants then listened to my instructions, read through the consent form, confirmed their consent and waited for me to start the interview. Two persons did not fully follow this pattern. One asked some questions making sure that the anonymity promised in the consent form would really be guaranteed. One other person remarked that I was speaking a mix between Standard Arabic and spoken dialect and wanted a clear instruction as to what register to use when speaking back to me. I answered that it did not matter and that it was up to the participant, an answer that was met with some bewilderment. This person thus expressed some awareness of linguistic form, and it cannot be excluded that the participant kept some of that awareness alive while reading the texts aloud. The rest of the participants, however, did not show any hesitations about the interview situation.

It should be noted here that the method to conduct the interviews via video calls had some unexpected beneficial consequences. Often, the participants were at home in their familiar environment, in a highly informal setting. It would be common that children were present playing in the background. One participant took the interview as an excuse to his family to have a cigarette break on the balcony, where he read the texts aloud on his mobile screen while smoking. Another was sitting in his window smoking while reading. A third person was busy renovating his home, but took a break and put down the tools to talk to me and read the texts while standing in the garden. Hence, the setting might well have played out in favour of creating a relaxed atmosphere which diverted attention away from linguistic form. It is difficult to imagine that the persons just mentioned had a strong focus on pronouncing case markers according to prescriptive grammar.

Before commencing the reading aloud of the six texts, a short trial text from Wikipedia about the coronavirus crisis was given to the participants, as a brief preparation. The participant was told it was a trial text for practice, and asked to read it aloud and answer some questions about the content. This served to clarify the alleged purpose of the study, and consolidate a focus on content. Several participants said after reading the trial text that they now understood exactly what they were expected to do. Some expressed surprise when the trial

text was removed from sight before they were asked questions about the content, making them realize that they must try to focus on the content and be able to answer questions after reading. The reading of the trial text also created a natural starting point when the "real" reading aloud started after the trial text. The effective reading time for each of the six texts was between one, and one and half, minute for each text, totalling between six and eight minutes of reading aloud. The entire interviews typically lasted around 40 minutes, so the bigger part consisted of answering questions, and of small talk before and after reading the texts. When asked questions about each text, participants answered in correct ways in most cases, confirming that they had understood the content of the texts. Throughout the interviews they then seemed fully engaged in understanding the texts and providing as correct answers as possible. The questions about the appropriateness of the language used in the texts were usually answered by affirming that the texts were written in a very good and understandable language. At no point did any participants seem to find it strange that they were asked questions, nor show any signs of suspecting that their reading performance was being observed. The small talk taking place before and after reading aloud was usually about the struggles involved in learning Swedish and finding a good future in Sweden. The atmosphere of this small talk was rather informal and, in most cases, there was laughing and friendly joking going on already before commencing the reading aloud. Thus, the overall atmosphere was not very formal.

The interviews were carried out mostly in Arabic, sometimes mixed with Swedish. This was not the result of a deliberate language choice, but out of necessity. Most of the participants had limited knowledge of Swedish and my Arabic is limited to a mix between Standard Arabic and the Levantine dialect, both of which I speak with many mistakes. The overall atmosphere was such that we all used any skills we had in Standard Arabic, dialect or Swedish, in order to communicate as well as possible. This seemed to affect the participants choice of variety, and most of them spoke back to me with a similar mix of dialect and Standard Arabic. It is hard to say if this also affected the extent to which they pronounced case markers while reading aloud. My assumption is that my language use had the effect of creating a somewhat informal situation, since I used quite a lot of spoken dialect, and that this benefited the desired focus on content rather than linguistic form.

The video calls were done with the video conferencing software Google Meet. This choice of software was determined by the fact that it was already in use by many of the participants, since it had been chosen by schools to conduct online teaching during the coronavirus crisis. Most participants used mobile phones, while I used a laptop computer during the calls. The texts were shared during the video calls using Google Meet's screen sharing function. One text at a time was thus presented on the participant's mobile phone screen, and read aloud by him or her. The text was only shared during the reading aloud, and not during the following questions to the participant. The audio was recorded using the software recording my computer's sound. It was necessary to make a recording of the computer's internal sound of the video call, since only this gives a sound quality that is good enough to be played back at reduced speed with maintained quality.

4.1.4 Preparation of the texts

Before the interviews were carried out, the texts were prepared in a way that enables a subsequent analysis with the objective of determining what factors might cause a use, or non-use, of case markers. The method used for this is the one developed by Hallberg, (2016, p. 124), and it is followed closely, with only minor modifications. This method consists of a code system, where tested words are tagged with specific codes, representing grammatical features. The code system is made up of five different tags assigned to each word, representing five grammatical categories: Declension category, case, case governor,

headedness and definiteness. Each of these categories encompasses several optional grammatical features, as will be explained in detail in the coming paragraphs. For example, under the category definiteness there is one tag for words with definite article, and one for indefinite words. This makes it possible to test which features correlate to case marking, allowing for a detailed analysis of the use of case markers. In addition to these five tags, all words in the texts were given unique id-numbers, as well as a code for the correct diacritic case marker according to prescriptive grammar.

ID-number	Tags for grammatical features	Word in Arabic	Diacritic case marker
FM91	TRI-GEN-PRE-HED-ART	بالسعالِ	KAS

Example X shows word number 91 in the text FM. It belongs to the declension category triptotes, stands in the genitive, was governed into that case by a preposition, is a head noun rather than an attribute, and has a definite article. The word in Arabic is *bi-s-suʿāli*, (by the cough/while coughing), and the correct case marker according to prescriptive grammar is *-i* marked by the diacritic sign *kasra*.

The following is a detailed presentation of how the words in this study were labelled with the five tags for grammatical features. The presentation follows Hallberg's method.

4.1.4.1 Declension category

Ryding (2005, p 182) lists eight different declension categories, three of which are represented without orthographic case markers in the texts in the present study: triptotes, diptotes and sound feminine plural. Since words without visible case markers are the only ones that can be assumed to leave it up to the reader to pronounce a case marker or not, words in other declension categories are excluded from the study. (There is also one word in the texts from the category knows as the five nouns, but it is not included in the tested words, due to its belonging in the *waqf*-category, which will be explained in section X.) As is shown in Figure X, the vast majority of words belong to the triptote category. The frequency for SFP and DIP is almost twice as high as in Hallberg's (2016, p. 201) study, which can be explained by the manipulations of the texts in the present study, which serves to increase the number of words in those two categories.

Declension category	Tag	*n*	%
Triptote	TRI	145	78,8%
Sound feminine plural	SFP	18	9,8%
Diptote	DIP	21	11,4%

Hallberg (2016, p. 204) refines his analysis of diptotes by distinguishing those which are indefinite, since only these differ from the normal triptote declension. The same distinction will be made in this study, not only to diptotes but also a similar distinction regarding the sound feminine plural, where only words in the accusative deviate from a normal triptote declension. I label the words *active* DIP and SFP. They are not given special tags, since they are defined by a combination of already existing tags. However, the irregularity of the case marking for these words justifies analysing them as special categories.

4.1.4.2 Case

As described in the background section, there are three cases in Arabic: nominative, accusative and genitive. The words in the texts are labelled with a tag indicating which case the word stands in. It is

Case	Tag	*n*	%
Nominative	NOM	24	13,0%
Accusative	ACC	41	22,3%
Genitive	GEN	119	64,7%

worth noting that the same grammatical case not always takes the same case marker, because of the irregularities in the categories DIP and SFP.

4.1.4.3 Case governance

Case governance refers to the reason for a word to be in a certain case. Eleven different case governors are present in the texts. They are described here as formulated by Hallberg (2016, p. 131). Nominative is caused by a word being subject in a phrase with the word order verb-subject-object (VSO) or subject-verb-object (SVO) or as topic in a verbless equational phrase (TOP) consisting of a topic and a comment (Hallberg, 2016, p. 133). Accusative is caused by five different governors in the texts: The word being an object (OBJ), being governed by the particle *ʾinna* (INN) or *ʾanna* (ANN), by the verb *kāna* (KAN), or the word being an adverb (ADV). Genitive has two different governors: The word being annexed by a genitive construction (ABS), or being complement to a preposition (PRE).

Case governance	Tag	*n*	%
Subject of VSO	VSO	9	4,9%
Subject of SVO	SVO	1	0,5%
Subject of eq. clause	TOP	14	7,6%
Object	OBJ	23	12,5%
ʾinna	INN	8	4,3%
ʾanna	ANN	2	1,1%
kāna and her sisters	KAN	1	0,5%
Adverb	ADV	7	3,8%
Annexed genitive	ABS	26	14,1%
Preposition	PRE	93	50,5%

4.1.4.4 Headedness

The two different categories in the tag for headedness are head noun (HED) and attribute (ATT), which in the texts are always adjectives. Attributes harmonize with head nouns, and usually take the same case marker. However, when the head noun and the following attribute belong to different declension categories, they might take different case markers.

Headedness	Tag	*n*	%
Head noun	HED	149	81,0%
Attribute	ATT	35	19,0%

4.1.4.5 Definiteness

There are four types of definiteness: Indefinite words (IND), words with definite article (ART), words in construct state as the first part of a genitive construction annexing another noun (CON), and words with a possessive suffix pronoun (SPN), as described in the background section.

Definiteness	Tag	*n*	%
Definite article	ART	73	39,7%
Suffix pronoun	SPN	8	4,3%
Construct state	CON	34	18,5%
Indefinite	IND	69	37,5%

4.1.5 Categories of words excluded from analysis

Three categories of words have been excluded from analysis, in spite of the fact that they otherwise meet the criteria for inclusion. They are words that according to prescriptive grammar take case markers, but would be inappropriate to include in the study.

The first such category includes words that occur before a pause (*waqf*), described in the background section. Since case markers should not be pronounced if a word stands in the *waqf* position, all words in this position are excluded from this study. Here, *waqf* is defined as words that are followed by punctuation, or that are the last word in a headline. Hence, only words where the written text indicates a pause are counted. This choice of method has its shortcomings, of course, since normal reading aloud naturally includes pronounced pauses also in other instances. However, the alternative, to define and count these pronounced pauses, would be too cumbersome for this study. It is also questionable whether it would have a big enough impact on the results to be justified. A total of 36 words have been defined as standing in the *waqf* position in the texts in this study. They have been coded with tags in the same way as other words, and have also been marked with a yellow colour in the text sheets, and have been excluded from the tested words. What can, and will be done, however, is to look at the words in waqf position as a separate category in order to see to what extent case markers are being used.

The second category of words that are excluded from the study are those that are followed by *hamzat al-waṣl*. These words assimilate sounds in a way that makes it difficult or impossible to distinguish between helping vowels and expressions of grammatical case. Ryding (2005, p. 19) describes how a liaison occurs between the preceding and succeeding letters and likens it to the way "we are" is assimilated into "we're", or "she is" into "she's" in English. A total of 51 words are excluded from the texts in this study because they are followed by *hamzat al- waṣl*. The most common instance of this assimilation occurs when a word is succeeded the definite article *al-* making up for all but three of the excluded words in this study. All words in the category *hamzat al-waṣl* have been coded with tags, and also marked with a pink colour in the text sheets.

The third category of excluded words are possessive suffix pronouns in the second person singular, and third person masculine singular. As mentioned in the background section, SPN words take an internal case marker, rather than a word final ending. However, in spoken dialect the same position is occupied by a vowel marking gender. It is then, for example, impossible to know if a person pronouncing a *kasra* in the word *baytik* (your house) means to express a genitive case marker, or simply use the dialectal vowel for feminine form, addressing a woman. This ambiguity makes it necessary to filter these words away from the result. A total of 19 words in the category SPN are excluded for this reason.

After excluding the 36 words in the category *waqf*, the 51 words in the category *hamzat al-waṣl*, and the 19 words in the category SPN, a total number of 184 words are left. It is these words that makes up the corpus for this study.

4.1.6 Manipulations of the texts

The six texts in the study have in some instances been slightly manipulated in order to increase the number of active DIP and SFP, and decrease the number of words in the category *waqf*. The higher number of DIP and SFP words was desired in order to obtain more data for these categories, since they can be considered more challenging in terms of correct case marking.

In eleven instances, manipulations were made in order to "activate" diptotes; that is, changing the text so that a diptote stand in the indefinite rather than the definite. As noted in the background section, only indefinite diptotes deviate from the normal triptote declension. The manipulation thus consisted of changing already present diptotes from the definite into the indefinite. These changes typically implied minor changes of the surrounding words, as seen in example X, where red words are those that have been manipulated, and underlined words are the resulting active diptotes.

Original text:	لا تكرر متابعة التقارير الإخبارية أو الأحاديث المتعلقة بما حدث كل الوقت.
Manipulated text:	لا تتابع **تقارير** إخبارية و**أحاديث** متعلقة بما حدث كل الوقت.

In seven instances, words in the sound feminine plural were activated by changing them into the accusative case. As described in the background section, only sound feminine plurals in the accusatives differ from the normal triptote declension. As with diptotes, already existing words in the texts were manipulated.

Original text:	عند سعالك وعطاسك، تقوم بنشر قطرات صغيرة
Manipulated text:	عند سعالك وعطاسك، تنشر **قطرات** صغيرة

In five instances words were added to the end of a sentence in order to avoid that a case marked word would come last, and thus be excluded from the study because it stands in the *waqf* position.

Original text:	هي مخاطر ضئيلة.
Manipulated text:	هي مخاطر ضئيلة جدًا.

The altered texts were edited into documents to look like the original documents published online, including logotypes and other layout features (see Appendix X). When copying the texts into these documents, attention was paid so that line breaks did not occur just before any tested words, the idea being that this would have risked affecting the reading aloud of those words.

During the preparation of these manipulations, I had several native Arabic speakers who don't participate in the study read the manipulated texts aloud for me. The purpose was to see if they would stop or hesitate at manipulated parts. I also asked them after reading if they had found any parts strange or unnatural. After this, two inserted manipulations were removed, since I noticed that several persons seemed to stop or hesitate before reading them. When asked about it, however, nobody reacted to those particular places, but to other parts of the texts that had not been manipulated, such as the use of a wrong verb form in the text SI. After the real interviews with the participants had been carried out, I did a sample test of five interviews and marked every time a person stopped or hesitated before reading a word. I could not see any pattern indicating that the readers stopped before manipulated parts more than they would do before other parts of the texts. It doesn't seem, then, that the manipulations affected the reading aloud in any significant way.

		FM	WA	WS	WP	RK	SI	FM	FM	FM	WA	WA	WA	WS	WS	WS	WP	WP	WP	RK	RK	RK	SI	SI	SI	SUM	%
Number of words		135	114	82	66	135	88																			620	
Case marked words		56	56	40	31	64	43	-13	-7	-4	-10	-9	-2	-6	-6	0	-6	-1	0	-6	-8	-12	-10	-5	-1	184	100%
Declension category																											
Triptote	TRI	51	47	36	22	52	27	-12	-7	-4	-7	-8	-2	-5	-6	0	-4	-1	0	-6	-8	-9	-7	-3	-1	145	78,8%
Sound feminine plural	SFP	3	2	4	5	1	11	0	0	0	-2	0	0	-1	0	0	-1	0	0	0	0	0	-3	-1	0	18	9,8%
Diptote	DIP	2	6	0	4	11	5	-1	0	0	-1	0	0	0	0	0	-1	0	0	0	0	-3	0	-1	0	21	11,4%
Five nouns	FIV	0	1	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,0%
Active SFP		3	0	1	2	1	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	6,5%
Active DIP		1	3	0	3	7	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	9,2%
Case																											
Nominative	NOM	2	7	5	5	7	9	-1	0	0	-3	0	0	-1	0	0	0	0	0	-2	0	0	-4	0	0	24	13,0%
Accusative	ACC	14	4	8	3	16	8	-1	0	0	-1	0	-1	-1	-1	0	-1	0	0	-2	-1	-3	0	0	0	41	22,3%
Genitive	GEN	40	45	27	23	41	26	-11	-7	-4	-6	-9	-1	-4	-5	0	-5	-1	0	-2	-7	-9	-6	-5	-1	119	64,7%
Case governance																											
Subject of VSO	VSO	1	3	2	0	3	8	-1	0	0	-1	0	0	-1	0	0	0	0	0	-1	0	0	-4	0	0	9	4,9%
Subject of SVO	SVO	0	1	0	0	2	0	0	0	0	-1	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	1	0,5%
Subject of eq. clause	TOP	1	3	3	5	2	1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	7,6%
Object	OBJ	3	2	6	1	12	7	-1	0	0	-1	0	-1	-1	0	0	0	0	0	0	-1	-3	0	0	0	23	12,5%
Inna	INN	4	2	0	2	0	1	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	8	4,3%
Anna	ANN	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1,1%
Kaana and her sisters	KAN	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0,5%
Adverb	ADV	6	0	1	0	3	0	0	0	0	0	0	0	0	-1	0	0	0	0	-2	0	0	0	0	0	7	3,8%
Annexed genitive	ABS	11	14	13	9	6	4	-2	-5	-2	-3	-5	-1	-4	-2	0	-1	-1	0	0	-4	0	-1	0	0	26	14,1%
Preposition	PRE	29	31	14	14	35	22	-9	-2	-2	-3	-4	0	0	-3	0	-4	0	0	-2	-3	-9	-5	-5	-1	93	50,5%
Headedness																											
Head noun	HED	48	46	35	26	51	33	-13	-7	-4	-10	-3	-2	-5	-4	0	-6	-1	0	-6	-5	-11	-8	-4	-1	149	81,0%
Attribute	ATT	8	10	5	5	13	10	0	0	0	0	-6	0	-1	-2	0	0	0	0	0	-3	-1	-2	-1	0	35	19,0%
Definiteness																											
Definite article	ART	22	29	21	8	23	23	-3	-5	0	-5	-9	0	-3	-4	0	-2	-1	0	-3	-6	0	-8	-4	0	73	39,7%
Suffix pronoun	SPN	5	6	2	0	13	1	0	0	-4	0	0	-2	0	0	0	0	0	0	0	0	-12	0	0	-1	8	4,3%
Construct state	CON	16	13	10	11	5	6	-10	0	0	-5	0	0	-3	0	0	-4	0	0	-3	0	0	-2	0	0	34	18,5%
Indefinite	IND	13	8	7	12	23	13	0	-2	0	0	0	0	0	-2	0	0	0	0	0	-2	0	0	-1	0	69	37,5%

= Active SFP
= Active DIP
= *hamzat al-waṣl*
= *waqf*
= SPN 3m.s or 2m./f.s.

Figure 1 Overview of the texts included in the study.

4.1.7 Ethical considerations

A precondition for this study was the possibility to record audio when the participants read the texts aloud. It would not have been possible to note the use of case markers while listening, since normal reading speed is far too fast to permit this. A consent form was used to obtain consent to recording and storing audio from the interviews (Appendix X). Each participant read and confirmed this by a written approval. The consent form was designed according to guidelines for consent forms under the European General Data Protection Regulation legislation (GDPR), published by the University of Linköping (University of Linköping, 2018, p. 9). These guidelines state that all participant must be informed in a clear and understandable way:

- That active consent is a legal ground for storing the information
- What information will be collected and stored
- What the information will be used for
- Who will be able to take part of the information
- How long time the information will be stored
- That it is possible to recall the consent at any time, but that results that have already been obtained can still be used by the researcher
- That it is possible to contact a data protection representant with complaints

As part of the procedure for each interview (Appendix X), the consent form was sent to the participants by e-mail or SMS. After reading it, the participants sent a written confirmation back.

Another ethical consideration concerned the fact that the exact purpose of the interviews was withheld from the participants; they did not know that their use of case markers was being observed. Instead, they were only informed in general terms that the Arabic language in the texts were the object of study. In light of the highly revered role of case markers, it is possible that a participant might be critical of this after being informed about the results of the study, perhaps getting the feeling that he or she was underperforming in an important arena of educational skills. However, this risk is negligible and a natural part of many linguistic studies. The possible ethical problem does not motivate refraining from the study.

It is also worth mentioning that the six interviews that were carried out on location in a school building took place with full respect to adequate measures regarding the ongoing coronavirus pandemic. Social distancing was upheld, and hand disinfectant was made available for the participants. Although class room teaching had been cancelled, the school was open certain hours per week for students to come and talk to teachers and staff. The interviews took place during those hours.

4.2 Analysis of data

This section will present how the recorded data was coded in order to enable a quantitative analysis after the interviews were done. It will then describe how these data was used to calculate different results, and a third point will describe technical aspects of the methodology.

4.2.1 Coding of recorded audio from the interviews

The recorded interviews were analysed by listening to them at a speed reduced down to 25%, making it possible to hear phonemes also when the participants were reading at a high speed.

Any pronounced case marker, correct or incorrect, was noted in an excel document with predefined codes, in order to enable a subsequent quantification.

The codes used for the pronounced – or not pronounced – case markers are shown in figure X. These are not codes for what would be the correct case marker according to the prescriptive grammar, but for the phoneme that the persons actually pronounced in the recorded interviews. The term *kasra*, for example, has no grammatical function in and by itself, but is a name for the pronounced phoneme *-i*. Apart from the six different alternatives for case markers, there is also the code SUK for *sukūn* which means that there is no pronounced vowel at all, as described in the background section. This is thus the marker for absence of case marker, and it is the code that defines non-use. If a word stands in the *waqf* position, before punctuation or as the last word in a headline as described in the background section, *sukūn* is the correct alternative. Since all the words in the *waqf* position have been filtered out of the corpus before analysis, however, all instances of sukūn constitute an absence of a case marker that should have been pronounced according to prescriptive grammar.

There are two codes for instances where it has not been possible to determine which phoneme was pronounced by the speaker in the recordings. The first code, AMB, is for phonemes that have ambiguous sounds. It might, for example, be a vocal that lies between *-a* and *-i*. The second code, XXX, is for instances where it was not at all possible to hear what phoneme the speaker pronounced because it was inaudible. It might be due to technical reasons, such as the speaker accidentally scratching the microphone, or because the speaker skips reading a word, or part of a word.

Pronounced case marker	**Code**
ḍamma	DAM
fatḥa	FAT
kasra	KAS
ḍamma tanwīn	DAM-T
fatḥa tanwīn	FAT-T
kasra tanwīn	KAS-T
sukūn	SUK
Ambiguous	AMB
Inaudible	XXX

The codes described thus far serve to define and count the use and non-use of case-markers, as well as ambiguous and inaudible instances. The codes, however, do not permit any distinction between use of correct case markers and incorrect ones. Any given pronounced case marker might be correct or incorrect depending on the word's grammatical function. For a triptote definite word in the genitive, *kasra* is correct, but if the word stands in the accusative *kasra* is incorrect. These codes, then, do not permit a counting of correct and incorrect case markers. This problem is solved by adding the Swedish letters å and ö to the code for each pronounced case marker. The letter å is added to correct case markers, and ö to incorrect ones. A pronounced *kasra* that is correct according to prescriptive grammar is then coded as KASå. An incorrect *kasra* is coded as KASö. This permits not only counting all pronounced case markers, but also correct and incorrect ones respectively. Example X shows the word

istikhdāma from the text FM, as it was coded after listening to the audio from the interviews with the twenty participants. To the left of the Arabic word is the ID-number for the word, and the five tags for grammatical features. Immediately to the right of the word is the correct case marker according to prescriptive grammar, in this case a *fatḥa*. Further to the right follow the codes for how the twenty participants, P1 to P20, pronounced the word. Three persons pronounced a correct case marker, three pronounced an incorrect one, two persons pronounced an ambiguous phoneme, and twelve did not pronounce any case marker at all, which is marked as *sukūn*. The green and red colours are not part of the coding system, but serve to visualize the proportions between correct and incorrect case markers.

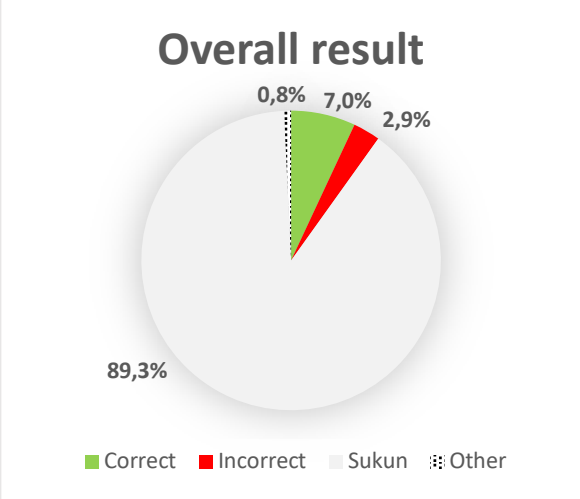
ID	Decl	Case	Gov	Head	Def	Word	Correct	P20	P19	P18	P17	P16	P15	P14	P13	P12	P11	P10	P9	P8	P7	P6	P5	P4	P3	P2	P1
FM75	TRI	ACC	INN	HED	CON	استخدام	FAT	SUK	SUK	KASö	FATå	KASö	SUK	SUK	SUK	SUK	AMB	FATå	SUK	FATå	SUK	SUK	KASö	SUK	SUK	AMB	SUK

5 Results

This chapter will first give the overall results with a total rate of use of case markers. These overall results will then be further presented in terms of categories of words that have unusually high rates of case marking, and how the rate of case marking shifted over time while the participants read the six texts aloud. After this, the results for each of the five categories of tags is presented, which reveals different patterns of use of case markers.

5.1 Overall result

The overall result shows the total use of case markers for all the 184 tested words in the study, and serves to answer the first part of the overarching question in the study: The *extent* to which native Arabic speakers use case markers when reading a text in Standard Arabic aloud, while focusing on content rather than on linguistic form. The overall rate of pronounced case markers, as shown in Figure X, is 9,9%, of which 7,0% are correct markers and 2,9% incorrect ones. A negligible number of words were coded as having inaudible or ambiguous case markers, together totalling 0,8% (in the following, these two categories will be excluded from the presentations of results). This means that an overwhelming majority, 89,3% of the tested words, were read aloud without any case marker. This result constitutes the main finding in the present study. Hence, the answer to the question posed above is that case markers are only used to a very small extent, compared to what is required by the prescriptive grammar. When texts in Standard Arabic are read aloud with focus on content rather than on linguistic form, it is for the most part a language of *sukūn*, that is, with words left unmarked for case.



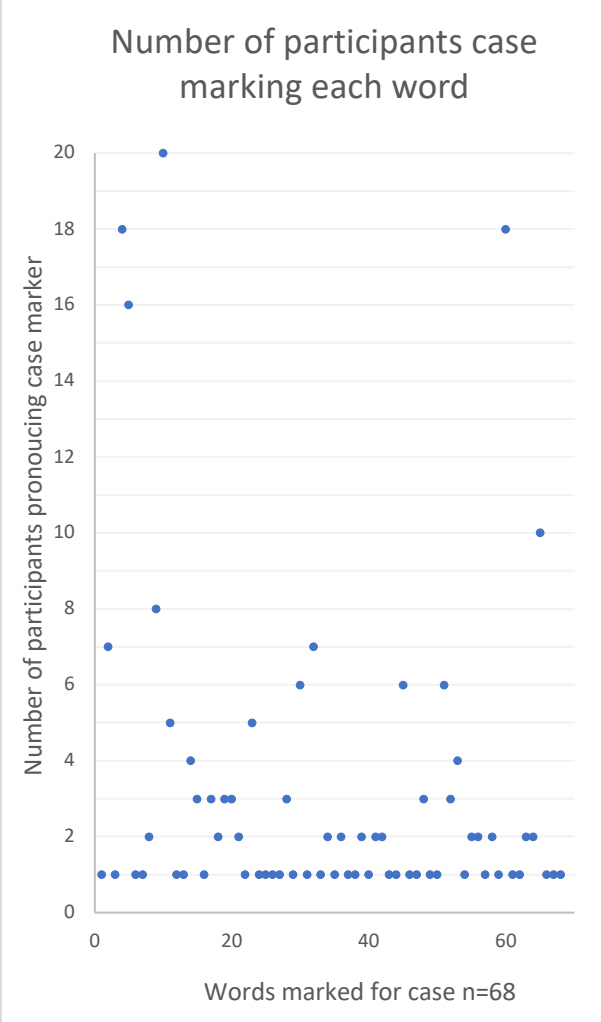
The variation between the participants is relatively limited, with a standard deviation of xx. Out of the 20 participants, 14 had a total rate of case marking (correct and incorrect) between 5,4% and 9,8%. One outlier had a significantly higher rate than all other participants, with a result of 30,4% (of which 27,7% were correct and 2,7% were incorrect).
The overall result does not give any information about patterns of use and non-use of case markers, but it constitutes the ground on which other results stand. Thus, when later results reveal certain patterns of use, these patterns all occur within an overall use of case markers that is very limited.

The overall results in this study can be compared to those in Hallberg (2016), which showed a 7,5% use of case markers, which is thus somewhat lower than in the present study. We need to note here, however, that 15% of the data points in Hallberg's study consist of words in the pause position (Hallberg, 2016, p. 168). As explained in the method section, words in pause position have been filtered out from the results in the present study. Their amount of data points corresponds to 16% of the total number, and if included the total result would be 8,2%. This figure is more correct to use in order to contrast the two studies, and as we can see the two results are now even closer, indicating that case markers are used to a similar extent in the two studies. There is, however, also one significant difference in the overall result between the two studies. Whereas the present study shows an almost 3% use of incorrect case markers, the same figure in Hallberg's study is only 0,3% (Hallberg, 2016, p. 165). This implies

that in the television interviews studied by Hallberg, people seemed to use case markers only when they were sure it would the correct thing to do, whereas in the present study participants used a slightly higher total rate of case markers, but with a much higher rate of markers that are wrong according to prescriptive grammar.

5.1.1 Number of participants marking each word for case

As the low overall rate of case marking suggests, not all of the tested words were marked for case by any of the participants, and for those that were, the number of persons pronouncing the word's case marker varies: Many words were marked by just one or a few participants, and a small number were marked by all, or almost all, participants. Out of the total number of 184 tested words, 108 words were left unmarked by all participants. Hence, the remaining 76 words were marked for case, correctly or incorrectly, by at least one person. A closer look reveals that words marked by a high number of participants tend to have one of two traits: They either belong to the category SPN, or the marked word is the word *kull*. In the following, a more detailed presentation of these two categories of words will be given. This is relevant for two reasons: First, it gives a more refined view of the overall rate of pronounced case markers, and secondly, it provides a tool which will later be used to analyse other results.



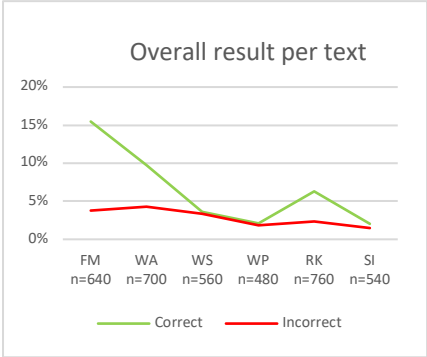
The category SPN belongs to the tag for definiteness, in accordance with the method used in this study. Results for use of case markers in the category will be presented in point X, as part of the results for each tag. Here, words in SPN are only analysed as proportion of the number of words tested. As can be expected, SPN words have a high rate of case markers. The 8 words in the category (corresponding to 11% of the total number of words) have almost 40% of the total number of case markers. An average of 18 participants pronounced a case marker for each of the words. If we remove the SPN words and look again at the number of participants pronouncing one or more case markers for each of the remaining 68 words, we can see the special role played by the word *kull*. Figure X illustrates the number of participants who pronounced a case marker for each word (after the SPN words have been excluded). The 68 words are shown as dots in the figure - the higher the number of participants pronouncing the word's case marker, the higher the dot's position in the figure. This visualizes a cluster of words at the lowest range, marked by only one or two persons. This is the vast majority of words. Only five words are marked by 10 or more persons. Every one of these five words is the word *kull*, which occurs five times in the texts. Four of the times the word has extremely high numbers, with between 16 and 20 participants pronouncing a case marker. This

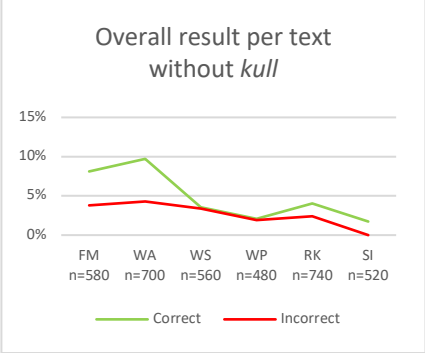
is relevant in that it indicates that there could be something about the word itself, rather than its grammatical role, that renders these high rates. Some evidence points at that direction: Participants consistently used the *fatḥa* ending for the word (with the exception of just 4 out of 82 case markers) even when the correct ending would be *ḍamma*. This indicates that the pronunciation *kulla* is a frozen form, rather than an expression of deliberate case marking. Moreover, three of five times the word occurs as part of the phrase *kulla yawm*, (every day) which could then be a fixed phrase.

In sum, the rate of participants marking words for case among the 184 tested words show that a majority of the words were left unmarked. Among the 76 words that were marked by at least one participant, words marked by a high number of participants were either words belonging to the category SPN, or the word *kull*. These two categories correspond to only 13 words, or 7% of the total number of tested words, but 62% of the total number of case markers. The reason for the high numbers for SPN words will be elaborated upon in point X. For the word *kull*, the results indicate that a possible reason is that this word, with a *fatḥa* ending, might be a frozen form, rather than a deliberate case marking according to prescriptive grammar. The total rate of case marking for all words, except SPN words and *kull*, is 4,0%, which is thus less than half compares to the rate for all words.

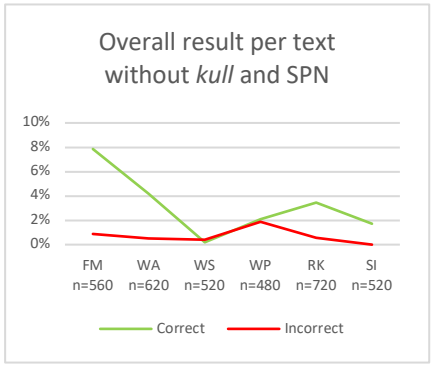
5.1.2 Consistency over time

As mentioned in the method section, the interviews lasted approximately 40 minutes, of which between six and eight minutes were effective reading time. By analysing the overall result over time, we can see whether participants increased or decreased their use of case markers during the course of the reading. It has been hypothesised that use of formal language markers are often waning over time. Schulz (1981, p. 156) reports that it is common wisdom that people in Egypt tend to speak at a higher stylistic level in the beginning of a discourse than in the end. He labels this a "discourse drift" (Schulz, 1981, p. 156), and theorizes that it is caused by the speakers' initial self-consciousness and desire establish their credibility and status. In his study of spoken monologues, Schulz confirmed this impressionistic observation and established that speakers used less features of Standard Arabic and more of spoken dialects, in the second half of their monologues (Schulz, 1981, p. 160). Schulz's study did not look at the use of case markers as a specific feature, but it is reasonable to assume that also they could be object of the same kind of discourse drift, since they are considered such a central part of prescriptively correct Standard Arabic. In contrast to Schulz's findings, Hallberg (2016, p. 175) did not find any general pattern of reduced case marking during the course of the studied interviews, but rather a consistent rate, be it high or low.





By looking at the overall result for each of the six texts in the order that they were read, we can see how the rate of case marking changes over the course of the reading. As figure X shows, there are clear signs of a discourse drift with a rate of pronounced case markers (correct and incorrect) gliding down a slope from almost 20% in the first text, to a mere 3,5% in the last one. The one exception is the fifth text RK. It is however motivated here to weigh in the special role played by the category SPN and the word *kull*. A closer look reveals that the first text contains a disproportionate number of the word *kull* (3 out of 5), which explains the sudden decrease in use of case markers. As figure X shows, results without the word kull instead show an increase between the first and second text. Figure X goes on to show the results without both *kull* and SPN, further nuancing the results. Figures X to X allow us to make a sounder assessment of possible discourse drift. As we can see, even when taking the differences into account, there is still an overall decrease from the first to the last text in all figures, albeit significantly smaller than what the first figure alone gives at hand. We can also see that the discourse drift seems to affect correct markers more; the rate for incorrect ones doesn't show the same amount of variation over time.



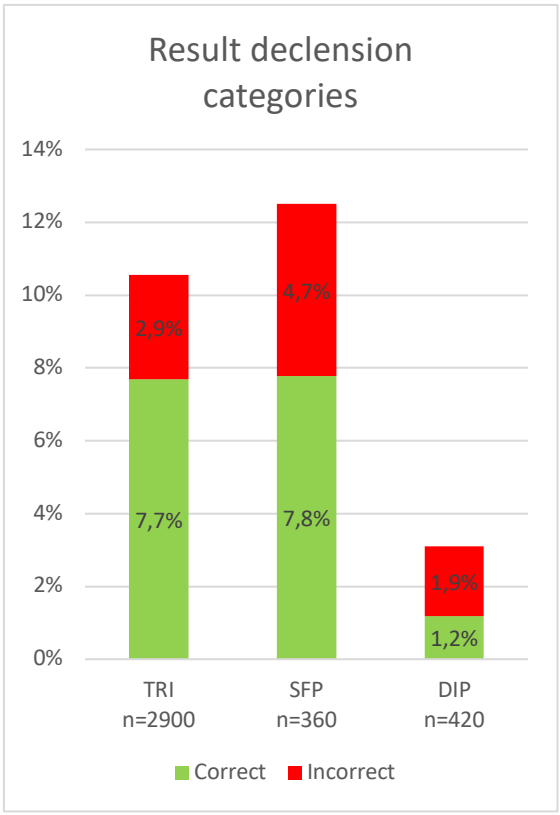
5.2 Results for the five tags

The results presented above were all based on the overall result. In the following points, results for each of the five categories for which the tested words were tagged will be presented. Whereas the overall results provided answers to the questions about the extent of use of case markers, the following results serve to answer the question what patterns this use follows.

5.2.1 Result declension categories

The results for the three declension categories show that TRI and SFP have a relatively similar rate of case marking, whereas DIP has a significantly lower rate, as shown in Figure X. This result differs from Hallberg's results in one significant way: The category SFP has a much higher rate. In Hallberg's study the rate for TRI was 4,1%, while both SFP and DIP had significantly lower rates with 2,7% and 2,0% (Hallberg, 2016, p. 203). The rate for the category DIP, on the other hand, follows the same pattern as in Hallberg's study, with its much lower rate.

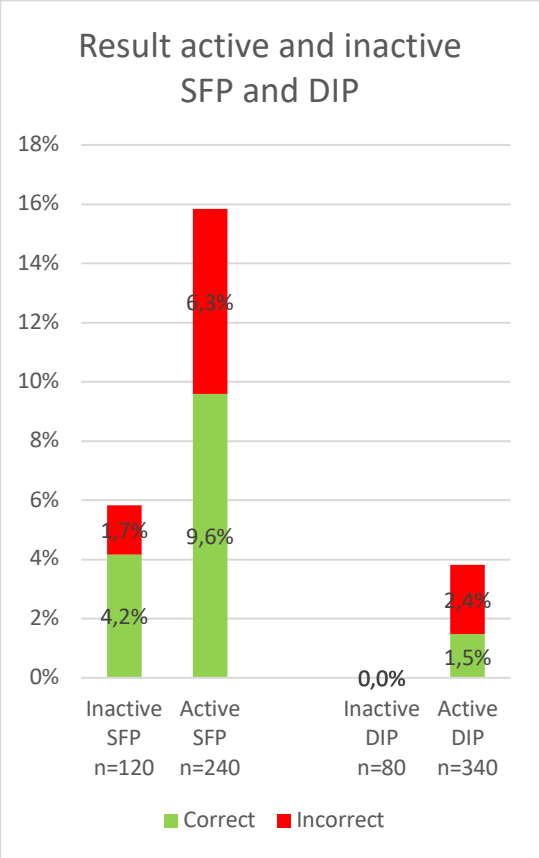
A more detailed look at the tested words in the category SFP shows no overrepresentation of SPN words, and there is for natural reasons no occurrence of the word *kull*, that could lead to a disproportionate rate of case marking. However, there are two adverbs in the text that have an unusually high rate of case marking, both of which are the same word: *marātin* (times), and both times part of the exact same phrase *marātin ʿadīdatan* (several times). These two instances of the word *marātin*, correspond to 15 case markers, or 33% of the markers for all SFP words (or 57% if SPN words are excluded). Indeed, the two instances of the word *marātin* have the highest number of persons marking them for case, after the word *kull*, with 7 and 8 persons respectively (See Figure X). The data is too limited to make more elaborate attempts at refining this result. What can be said, though, is that two instances of the same word lie behind a disproportionate number of case markers for SFP words. Possible reasons for the high rate for these words might be their grammatical function as adverbs, or that case markers might be treated by some participants as part of a frozen form for the word. It can however not be excluded that also the declension category SFP itself has an effect on the higher rate, though this claim has to be taken with caution.



For the category DIP, it can be noted that it contains no instances of neither SPN words or kull, which contributes to a lower rate.

When we isolate the rates for active SFP and DIP and compare them to inactive words in the same categories, some additional results come to light, as shown in Figure X. The significant difference between active and inactive SFP can be explained by the fact that all instances of SPN and the word *marātin* belong to active SFP. Still, without these words there is a rate of almost 6% for inactive SFP, which is more than the equivalent overall rate of 4,0% (when SPN and *kull* are excluded). Again, this suggests that a word's belonging to the category SFP might cause a significantly higher rate of case marking than diptotes, and even somewhat than the overall average.

When it comes to active and inactive DIP, we can see that inactive words have no pronounced case markers at all. This result needs to be taken with some caution, since the amount of data for inactive DIP is relatively low, corresponding to only four words. The rate for active DIP differs from Hallberg's (2016, p. 203) study where only a predicted 0,01% of indefinite diptotes were marked for case. Hallberg's study did not confirm the hypothesis that indefinite diptotes would be routinely inflected as triptotes, since they are considered particularly difficult to mark correctly. Rather, Hallberg's results (2016, p. 203) indicate that most speakers chose not to mark them at all. In the present study, however, a different result emerges. Active DIP words are indeed marked for case, at a total rate of 3,9%, which can be compared to the overall rate for indefinite words which is 5,0%, or the overall result of 4,0% excluding SPN and *kull*. The data is too limited to determine if speakers inflected words as if

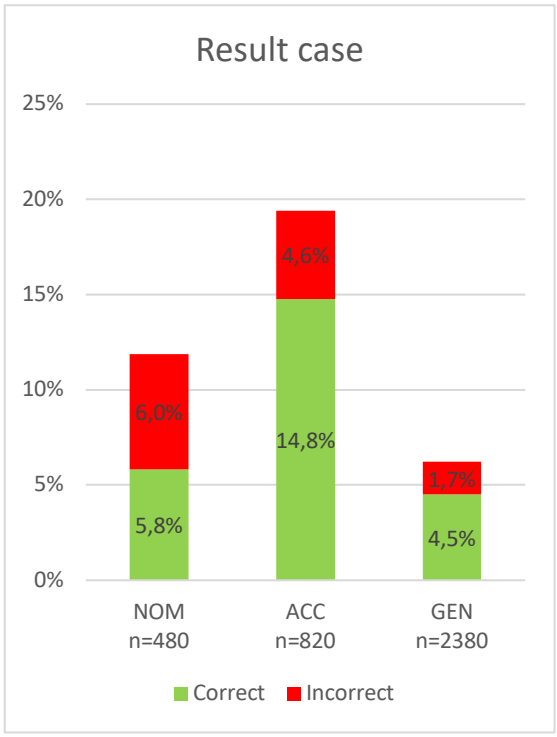


they were triptotes, in the sense that they would follow grammatical patterns that would be correct for triptotes. It can however be concluded that speakers don't chose to leave active DIP words unmarked, as was the result in Hallberg's study.

In sum, the results for declension categories show a similar rate of case marking for TRI and SFP words, with DIP having a much lower rate. The rate for SFP is dominated by one word occurring two times with an unusually high rate, but also after taking that into consideration, there is a higher rate for SFP than for DIP words. These results are to be taken with some caution, since it is not entirely clear what factors caused the different rates. What can be said with more certainty, is that neither active SFP or DIP words were systematically avoided for case marking by the participants to any large extent. Their irregular rules for case marking could be assumed to make them more difficult to mark for case, and thus cause an avoidance, but this is rejected by the results in the present study. This differs from Hallberg's study, where active DIP words were almost never marked for case.

5.2.2 Result case

As is shown in Figure X, the accusative case has a significantly higher rate of case markers then the other two cases. This differs from Hallberg's results, which showed no major difference between the three cases. However, this result too can be analysed in more detail by taking the category SPN and the word kull into consideration. The category ACC contains four of the five *kull*, leading to a higher rate of case markers. Without the word *kull*, the rate turns out to be more even, with 10% for NOM, 12% for ACC and 6% for GEN. There is no overrepresentation of SPN words in any of the three cases, but the significantly higher number of tested GEN words implies a less significant impact for the ones that occur. This might be a reason for the lower rate for GEN words.



The method, and the additional considerations of SPN words and the word *kull*, does not permit an exact assessment or provide detailed rates for the three cases. It is however enough to argue that there is no significant difference between them that would indicate that case has any major

influence on participants' use of case markers. This result then harmonizes with Hallberg's (2016, p. 229).

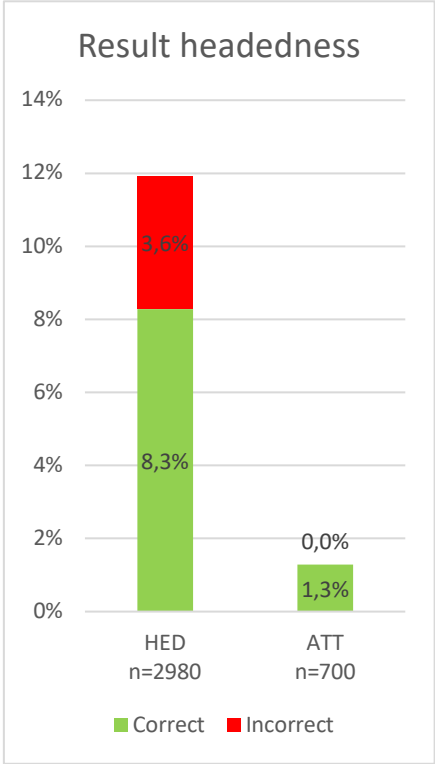
5.2.3 Result case governor

The rates of case marking for the ten different case governors provides only limited information in terms of patterns that could explain use of case markers. The rates are relatively similar, and there might be other factors in play that might explain the rates. It can be noted that ANN and ADV have significantly higher rates than the all other governors. It turns out, however, that ANN contains only two words, both of which are SPN words, which can be said with certainty to be the reason for the high rate. The ADV words contains four of the five *kull*, which similarly boosts the rate. In the case of ADV, however, it should be noted that there is also another word that occurs two times with unusually high rate of case marking; the word *marātin*, both of the times part of the phrase *marātin ʿadīdatan*, as mentioned above. This might indicate that the word's function as adverb does indeed play some role for the rate of case marking. Alternatively, the word might be perceived as a frozen form when part of the mentioned phrase.

In sum, the present study doesn't provide evidence of any clear patterns of use of case markers as a result of case governors. This is similar to the results in Hallberg's study (Hallberg, 2016, p. 235).

5.2.4 Result headedness

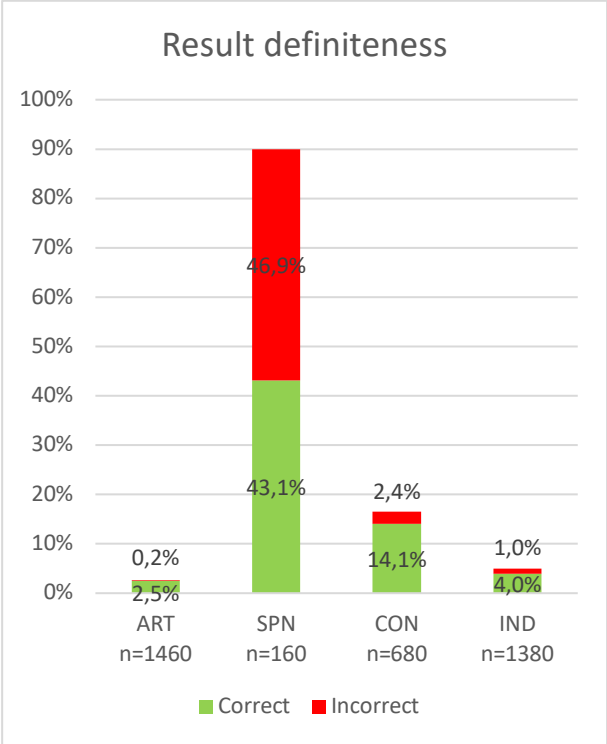
The result for headedness shows a distinctly higher rate for head nouns as compared to attributes. Figure X shows that HED words have a rate which is almost ten times higher than for ATT. All SPN words and *kull* are included in the HED words, contributing to the high rate. However, even a result without SPN and *kull* shows a rate of 4,7% for HED, which is still almost five times higher than for ATT. This proportion is similar to the result found by Hallberg (2016, p. 227). He argues that the lower rate for attributes might be caused by their longer distance from the governing element, and the fact that they can take a different case ending than the preceding head noun if they belong to a different declension category such as sound feminine plural. (Hallberg, 2016, p. 226). In the present study, there is an example of just that: the phrase *marātin ʿadīdatan,* which has been mentioned above. Both times this phrase occurs, a high number of participants pronounced a case marker for the head noun *marātin*, 7 and 8 participants respectively. This makes the word the most marked word in the study after the word *kull* (See figure X), if SPN words are excluded. In stark contrast, only 1 respectively 0 persons pronounced any case marker for the following attribute *ʿadīdatan*.



This confirms the results and conclusions in Hallberg's study, namely that there is a significant difference in use of case markers caused by headedness, and that attributes are marked for case significantly less often than head nouns.

5.2.5 Result definiteness

The results for definiteness show some of the clearest patterns of use of case markers in the study. As has been mentioned, SPN words are expected to have a have a high rate of case markers, due to their morphological trait with no possibility for pause position. The results for SPN words show a 90% use of case markers, 43,1% of which are correct markers. This is evidence that the presence of possessive suffix pronouns causes participants to pronounce case markers. There are no instances of the word *kull* among the SPN words, and no other underlying factor that could explain the high rate. Hallberg (2016, p 130) showed a rate of 43,9% for the same category, which is thus almost exactly the same rate as for correct case markers in the present study. However, this study also includes another 46,9% of incorrect case markers, while the rate for *sukūn* is a mere 7,5%. The overall use of incorrect markers, which is a main difference from Hallberg's study, is to a large part concentrated to SPN words; Their rate of incorrect markers corresponds to almost 70% of the total number of incorrect markers in the whole study. Hallberg, too, found that a disproportionally high rate of incorrect markers occurred on SPN words (Hallberg, 2016, p. 212), albeit on a much smaller scale than in the present study. His conclusion was that in extemporaneous speech, people seem to mark words for case only when they are certain of the right form, but that this "threshold of doubt" (Hallberg, 2016, p. 212) is lower for SPN words: it is worth risking to pronounce an incorrect marker for these words. The results in the present study point at the same direction, but on a much larger scaler, with significantly higher rates of incorrect markers for SPN words.



In contrast to SPN words, the result for words with definite article, ART, show a low rate of case marking, with only 2,7%. Similar to the use of case marker for SPN words, a non-use appears to be the norm for words with definite article. Hallberg's results show an even lower rate for words with definite article, with a mere 0,3%. It is worth noting, then, that the rate of case markers for these words, even if very low, is almost ten times higher than in Hallberg's study. The bigger picture, though, is the same, namely a very low use compared to SPN, and the other categories.

The category IND shows a rate of 5% which is similar to Hallberg's study, while CON words have a much higher rate, with 16,5% in the present study, compared to 2,8% in Hallberg's study. However, among the CON words are all five *kull*, boosting the rate. Without them, the rate for case marking is 5,2%. This follows the same pattern as in Hallberg's study, then, but at a higher rate.

In sum, the results strongly indicate that pronouncing case markers is a norm for words with possessive suffix pronoun when reading Standard Arabic aloud, while words with definite article are only rarely marked for case. The results for SPN also illustrate a significant difference between the present study and Hallberg's: In addition to a use of correct case markers, there is also a wide use of incorrect markers.

5.3 Summary

The overall result in this study is that almost 90% of the data points were coded as *sukūn*, which means that no case marker was pronounced. The rate of correct case markers was 7% while there was also an additional occurrence of incorrect case markers, with a rate of 2,9%. The results clearly show that when native speakers of Arabic read a text in Standard Arabic aloud, with focus on content rather than linguistic form, the use of case marker is low, around ten percent of the prescribed rate. A relatively small number of words with possessive suffix pronouns and the word *kull* (both categories together correspond to 7% of the total number of words) have exceptionally high rates of case marking, corresponding to 62% of all case markers. There is some evidence of a discourse drift while the participants are reading, with a decrease in use of case markers over the course of time: However, this result needs to be taken with some caution.

The results also reveal different patterns of use of case markers. Some of these results give clear indications, while some others need to be taken with caution. The following patterns of use can be identified with certainty:

- Words with possessive suffix pronouns are marked for case to a very large extent, with a rate around 90%
- An unusually high proportion of markers (almost 47%) for possessive suffix pronouns were incorrect.
- Words with definite article are marked for case to a small extent with a rate of 2,7%.
- Words that are attributes are only rarely marked for case, with a rate of 1,3%, indicating that case marking occurs mostly for head nouns.

Some other results give indications about patterns, but with less certainty. When words in the sound feminine plural stand in the accusative, and diptotes are indefinite, they don't seem to be avoided for case marking, despite being more difficult to mark due to irregular grammatic rules. Indefinite diptotes, seen as particularly difficult, are marked for case at almost as high rate as the average for indefinite words, in contrast to Hallberg's study where they were almost always left unmarked. There are signs that some words have become frozen with pronounced case markers when reading aloud in Standard Arabic, notably the word *kull* with a *fatḥa* ending.

6 ANALYSIS

The results in the present study reveals a wide gap between the codified norm and the empirical norm for reading aloud in Standard Arabic with focus on content rather than linguistic form: The codified norm is a 100% rate of case marking, while the empirical norm turns out to be 9,9% (of which around 2,9% are incorrect case markers). This confirms the presence of a standard language ideology, as described by Milroy (1991 and 2006), but as any standard language ideology, it is an ideal that is not achieved in the reality of language use (Milroy, 2006, p. 133).

It can be discussed whether the result really shows a limited use or not. Obviously, a 10% use is a very low rate if the set target is a 100% use, and it is therefore correct to conclude that the present study shows a very limited use of case markers. On the other hand, due to the lack of previous empirical studies of reading aloud, different people might have different expectations and preconceptions about what a normal use would be like. Perhaps some speakers would guess that a normal use of case markers would have a much higher rate (surely, many would say it sho*uld be* higher), and they might agree that the results show a limited use. At the same time, some other people might expect that there would be hardly any use at all of case markers in this kind of context, and they might argue that there is actually a substantial use, after all. What we can do is to compare the results to the empirical evidence at hand, in the form of Hallberg's study of extemporaneous speech. One of the purposes of this study is precisely to contrast the results to Hallberg's study, where it is hypothesized that case marking when reading aloud plays out in the same way as in extemporaneous speech. Indeed, at a first glance the results in the present study seem to confirm this hypothesis when it comes to the overall use of case markers, with a rate of 9,9% in the present study compared to 7,5% in Hallberg's study. The difference is even smaller when one weighs in the words in pause position, 8,2% in the present study compared to 7,5% in Hallberg's study. It is worth considering here, though, that in Hallberg's study, the participants were all highly educated, half of them had doctoral degrees and the rest had other university degrees, (three persons had unknown educational background). Likewise, it is noteworthy that they all spoke in the formal context of a television interview. Therefore, it might be speculated that if a study would observe extemporaneous speech in Standard Arabic under more comparable circumstances - spoken by participants comparable to the ones in the present study, and in a less formal context, similar to the present study - the rate would be lower than what Hallberg's results show. This is of course a speculation, but it is worth considering the possibility that the results in the present study in fact indicate a somewhat higher use of case markers for reading aloud, than for extemporaneous speech. It is outside the scope of this study to theorize about a possible reason for this, but it would be worth exploring the role played by the visual input from printed text, which is absent in extemporaneous speech. Could it be that this triggers a somewhat higher tendency to pronounce case markers? Or is it the fact that the language in the texts read aloud doesn't originate in the speaker's mind, as in extemporaneous speech? Could this create a distance that activates the codified norm to a somewhat higher degree?

The words with possessive suffix pronouns stand out in the results of this study, and the high rate of use of case markers for these words is indeed one of the main findings. What we see here is evidently how the codified norm of the standard language ideology is very much active, in spite of the low overall rate. Since the rate for SPN words is as high as 90%, it can be said that use of case markers for these words is part of an empirical norm when reading Standard Arabic aloud. In other words, their use is ordained both by prescriptive and descriptive grammar, and to read SPN words aloud without case markers would be to break both the codified and empirical norm. It seems probable that the reason for the high rate is the lack of potential pause position: Case markers are not word-final endings here, but occur inside the

word, making it more difficult to ignore them. Interestingly, as much as half of the case markers for SPN words in the present study were incorrect, indicating that speakers find it more important to pronounce a case marker, than being sure it is actually the correct one. As theorized by Hallberg, the threshold of doubt seems lower for SPN words. When a word-internal "gap" opens up and leaves room for the speaker to fill it by a case marker, it seems to increase the pressure to live up to the codified norm. One way of seeing this could be as a push-pull relation between the prescriptive grammar and an empirical norm of a predominant non-use of case markers. The norm of an overall non-use pushes the use of case markers down to very low rates, but the category of SPN words pulls it up to the surface, and reveals a use that is to a large extent incorrect, according to prescriptive grammar. This both contradicts and confirms Hallberg's results for extemporaneous speech. On the one hand, he too found a higher frequency of incorrect case markers for SPN words. This frequency, however, was much lower than in the present study, mostly leaving SPN words to be correctly marked, or not marked at all. The tension between the codified and empirical norms might also explain the slight signs of discourse drift in the present study, with a self-imposed pressure to live up to the codified norm at the beginning of the interview. The fact that there was no similar discourse drift in Hallberg's study could indicate, again, that the empirical norm for reading aloud means a somewhat higher rate of case markers.

Similar to the results for SPN words, the rate for active SFP and DIP indicate a somewhat higher use of case markers than in Hallberg's study. This is particularly clear for active DIP words, which had almost no case marking in Hallberg's study. The present study indicate that these words are not avoided for case marking, which could point at the same direction as the argument above, namely that there might be a slightly more frequent use of case markers in reading aloud than in extemporaneous speech. However, this is a speculation that needs to be taken with caution.

If we look at the different patterns of use and non-use of case markers - a use for SPN words, a non-use for words with definite article and attributes - we can see exactly the kinds of language variations that are out of reach for the prescriptive grammar to explain. Since it builds on a codified norm, this grammar is an expression of the prevailing standard language ideology and gives no clue as to which grammar is needed to live up to the empirical norm. The results in this study, with the patterns of use and non-use described, provides a contribution to answering the question what grammar is needed to read Standard Arabic aloud according to the empirical norm.

Appendix A - Script for interview procedure

Before conducting the interviews, a script with a detailed description of the planned interview procedure was written. This script was then followed closely during each interview. The following is a translation of the script into English.

- I am a student of Arabic at the university of Gothenburg, and conducting a study about the Arabic language in texts with information about the coronavirus from the Swedish state, the Red Cross, and the WHO.

- The coronavirus itself is not the object of research, and I don't have any medical competence. I chose this subject because people are familiar with it.

- I will do an interview with you. You will read some short texts aloud and I will ask questions about them. The sound will be recorded so that I can analyse the interview afterwards. There is a law saying I must have your consent for this, so I want you to read this consent form and confirm that you agree.

- (I send the consent form via e-mail or SMS and wait until the participant have read it and sent a written reply confirming consent.)

- I will now share texts on your screen. For methodological reasons, I must know after the interviews that all participants read exactly the same information before answering the questions. So, you must read the texts aloud in their entirety, including headlines.

- You can answer my questions in Arabic. If I don't understand everything, I can always listen to the recording afterwards.

- We will now do a trial before starting the real interview. Please read this text. (I share a very short Wikipedia text about the coronavirus pandemic, and ask three questions about it after the reading aloud)

- Now we start the interview. Please read text 1 which is from the Swedish Public Health Agency. (Reading aloud) Now answer these questions:
  How many times should one wash the hands every day?
  What did the text say about door handles?
  What happens when you cough and sneeze?
  What do you think about this way of giving information?

- Now please read text 2 which is from the WHO. (Reading aloud)
  Why doesn't it help against the coronavirus to spray the body with alcohol?
  When does it work to use alcohol and chlorine?
  Do you think it is necessary to give this information, or did everybody already know this?

- Now please read text 3 which is from the WHO. (Reading aloud)
  Why doesn't sesame oil protect against the coronavirus?
  Why isn't it good to apply chemicals onto the skin?
  Do you think a lot of people know the words *al-klūrūfūrm* and *ḥamaḍ al-bīrūksī āsitīk*? Is it appropriate to include such words in this text?

- Now please read text 4 which is from the WHO. (Reading aloud)
  Why is it safe to touch parcels from regions with reported cases of Covid-19?
  Do you think this information is useful for many Arabic speakers in Sweden?

- Now please read text 5 which is from the Swedish red Cross. (Reading aloud)
  Can you mention at least three of the advice given in the text?
  Is it necessary to have this text in Arabic, or would it suffice to write it in Swedish?

- Now please read text 6 which is from the Swedish Institute, which represents the Swedish government. (Reading aloud)

From where come recommendations to the government of Sweden?
What kinds of persons are represented in those bodies?
What do you think about the language used, is it appropriate?

- The interview is now over. Could you tell me your age, dialect and number of years in school?

Appendix B

Appendix B to D show the text FM in three steps: The way it was presented to the participants, the way it was prepared with codes before the experiment, and the result sheet for words from the text that was tested.

العناية بنظافة اليدين

تجنب العدوى من المصابين أو إصابة الآخرين بالعدوى

اغسل يديك مرات عديدة كل يوم!

إن كل ما يسبب العدوى يعلق بسهولة على اليدين وغيرها من الأسطح مثل مقابض الأبواب، و هي تنتقل كذلك عند المصافحة باليد.
ولذلك يجب غسل اليدين بالصابون والماء الدافئ مرات عديدة كل يوم.
قم دائماً بغسل اليدين قبل تناول الطعام، تحضير الطعام و كذلك بعد استخدام المرحاض.
إن استخدام كحول التنظيف اليدوي قد يكون خياراً إذا لم تكن لديك الفرصة لغسل اليدين.

قم بالسعال والعطس في فجوة ذراعك!

عند سعالك وعطاسك، تنشر قطرات صغيرة تحتوي على عناصر معدية جداً.
إن السعال والعطس في فجوة ذراعك أو في منديل ورقي يمنع العدوى من الانتشار في البيئة المحيطة بك و يحمي يديك من التلوث.

Folkhälsomyndigheten

God handhygien_arabiska 2015-12-03

Appendix C

The text FM is here shown in full, after it has been coded with tags for id-number, grammatical features and correct case marker according to prescriptive grammar. Colours have been added to visualise correct markers, and categories of words that have been excluded from the analysis: *hamzat al-waṣl, waqf,* and SPN in third person masculine and second person masculine/feminine. Moreover, active SFP and DIP have been marked with blue colours.

ID	Decl	Case	Gov	Head	Def		Correct
FM1	TRI	NOM	TOP	HED	ART	العناية	DAM
FM2	TRI	GEN	PRE	HED	CON	بنظافة	KAS
FM3						اليدين	
FM4						تجنب	
FM5						العدوى	
FM6						من	
FM7						المصابين	
FM8						أو	
FM9	TRI	ACC	OBJ	HED	CON	إصابة	FAT
FM10						الآخرين	
FM11						بالعدوى	
FM12						اغسل	
FM13						يديك	
FM14	SFP	ACC	ADV	HED	IND	مرات	KAS-T
FM15	TRI	ACC	ADV	ATT	IND	عديدة	FAT-T
FM16	TRI	ACC	ADV	HED	CON	كل	FAT
FM17	TRI	GEN	ABS	HED	IND	يوم	KAS-T
FM18						!	
FM19						إن	
FM20	TRI	ACC	INN	HED	CON	كل	FAT
FM21						ما	
FM22						يسبب	
FM23						العدوى	
FM24						يعلق	
FM25	TRI	GEN	PRE	HED	IND	بسهولة	KAS-T
FM26						على	
FM27						اليدين	
FM28						و	
FM29	TRI	GEN	PRE	HED	SPN	غيرها	KAS
FM30						من	
FM31	TRI	GEN	PRE	HED	ART	الأسطح	KAS
FM32						مثل	
FM33	DIP	GEN	PRE	HED	CON	مقابض	KAS
FM34	TRI	GEN	ABS	HED	ART	الأبواب	KAS
FM35						،	
FM36						و	
FM37						هي	
FM38						تنتقل	
FM39						كذلك	
FM40						عند	
FM41	TRI	GEN	PRE	HED	ART	المصافحة	KAS
FM42	TRI	GEN	PRE	HED	ART	باليد	KAS
FM43						.	

ID	Decl	Case	Gov	Head	Def		Correct
FM44						و	
FM45						لذلك	
FM46						يجب	
FM47	TRI	NOM	VSO	HED	CON	غسل	DAM
FM48						اليدين	
FM49	TRI	GEN	PRE	HED	ART	بالصابون	KAS
FM50						و	
FM51	TRI	GEN	PRE	HED	ART	الماء	KAS
FM52	TRI	GEN	PRE	ATT	ART	الدافئ	KAS
FM53	SFP	ACC	ADV	HED	IND	مرات	KAS-T
FM54	TRI	ACC	ADV	ATT	IND	عديدة	FAT-T
FM55	TRI	ACC	ADV	HED	CON	كل	FAT
FM56	TRI	GEN	ABS	HED	IND	يوم	KAS-T
FM57						.	
FM58						قم	
FM59						دائما	
FM60	TRI	GEN	PRE	HED	CON	بغسل	KAS
FM61						اليدين	
FM62						قبل	
FM63	TRI	GEN	PRE	HED	CON	تناول	KAS
FM64	TRI	GEN	ABS	HED	ART	الطعام	KAS
FM65						،	
FM66	TRI	GEN	PRE	HED	CON	تحضير	KAS
FM67	TRI	GEN	ABS	HED	ART	الطعام	KAS
FM68						و	
FM69						كذلك	
FM70						بعد	
FM71	TRI	GEN	PRE	HED	CON	استخدام	KAS
FM72	TRI	GEN	ABS	HED	ART	المرحاض	KAS
FM73						.	
FM74						إن	
FM75	TRI	ACC	INN	HED	CON	استخدام	FAT
FM76	TRI	GEN	ABS	HED	CON	كحول	KAS
FM77	TRI	GEN	ABS	HED	ART	التنظيف	KAS
FM78	TRI	GEN	ABS	ATT	ART	اليدوي	KAS
FM79						قد	
FM80						يكون	
FM81						خيارا	
FM82						إذا	
FM83						لم	
FM84						تكن	
FM85						لديك	
FM86	TRI	ACC	KAN	HED	ART	الفرصة	FAT
FM87	TRI	GEN	PRE	HED	CON	لغسل	KAS
FM88						اليدين	
FM89						.	

ID	Decl	Case	Gov	Head	Def		Correct
FM90						قم	
FM91	TRI	GEN	PRE	HED	ART	بالسعال	KAS
FM92						و	
FM93	TRI	GEN	PRE	HED	ART	العطس	KAS
FM94						في	
FM95	TRI	GEN	PRE	HED	CON	فجوة	KAS
FM96	TRI	GEN	ABS	HED	SPN	ذراعك	KAS
FM97						!	
FM98						عند	
FM99	TRI	GEN	PRE	HED	SPN	سعالك	KAS
FM100						و	
FM101	TRI	GEN	PRE	HED	SPN	عطاسك	KAS
FM102						،	
FM103						تنشر	
FM104	SFP	ACC	OBJ	HED	IND	قطرات	KAS-T
FM105	TRI	ACC	OBJ	ATT	IND	صغيرة	KAS-T
FM106						تحتوي	
FM107						على	
FM108	DIP	GEN	PRE	HED	IND	عناصر	FAT
FM109	TRI	GEN	PRE	ATT	IND	معدية	KAS-T
FM110						جدا	
FM111						.	
FM112						إن	
FM113	TRI	ACC	INN	HED	ART	السعال	FAT
FM114						و	
FM116	TRI	ACC	INN	HED	ART	العطس	FAT
FM117						في	
FM118	TRI	GEN	PRE	HED	CON	فجوة	KAS
FM119	TRI	GEN	ABS	HED	SPN	ذراعك	KAS
FM120						أو	
FM121						في	
FM122	TRI	GEN	PRE	HED	IND	منديل	KAS-T
FM123	TRI	GEN	PRE	ATT	IND	ورقة	KAS-T
FM124						يمنع	
FM125						العدوى	
FM126						من	
FM127	TRI	GEN	PRE	HED	ART	الانتشار	KAS
FM128						في	
FM129	TRI	GEN	PRE	HED	ART	البيئة	KAS
FM130	TRI	GEN	PRE	ATT	ART	المحيطة	KAS
FM131						بك	
FM132						و	
FM133						يحمي	
FM134						يديك	
FM135						من	
FM136	TRI	GEN	PRE	HED	ART	التلوث	KAS
FM137						.	

= Active SFP

= Active DIP

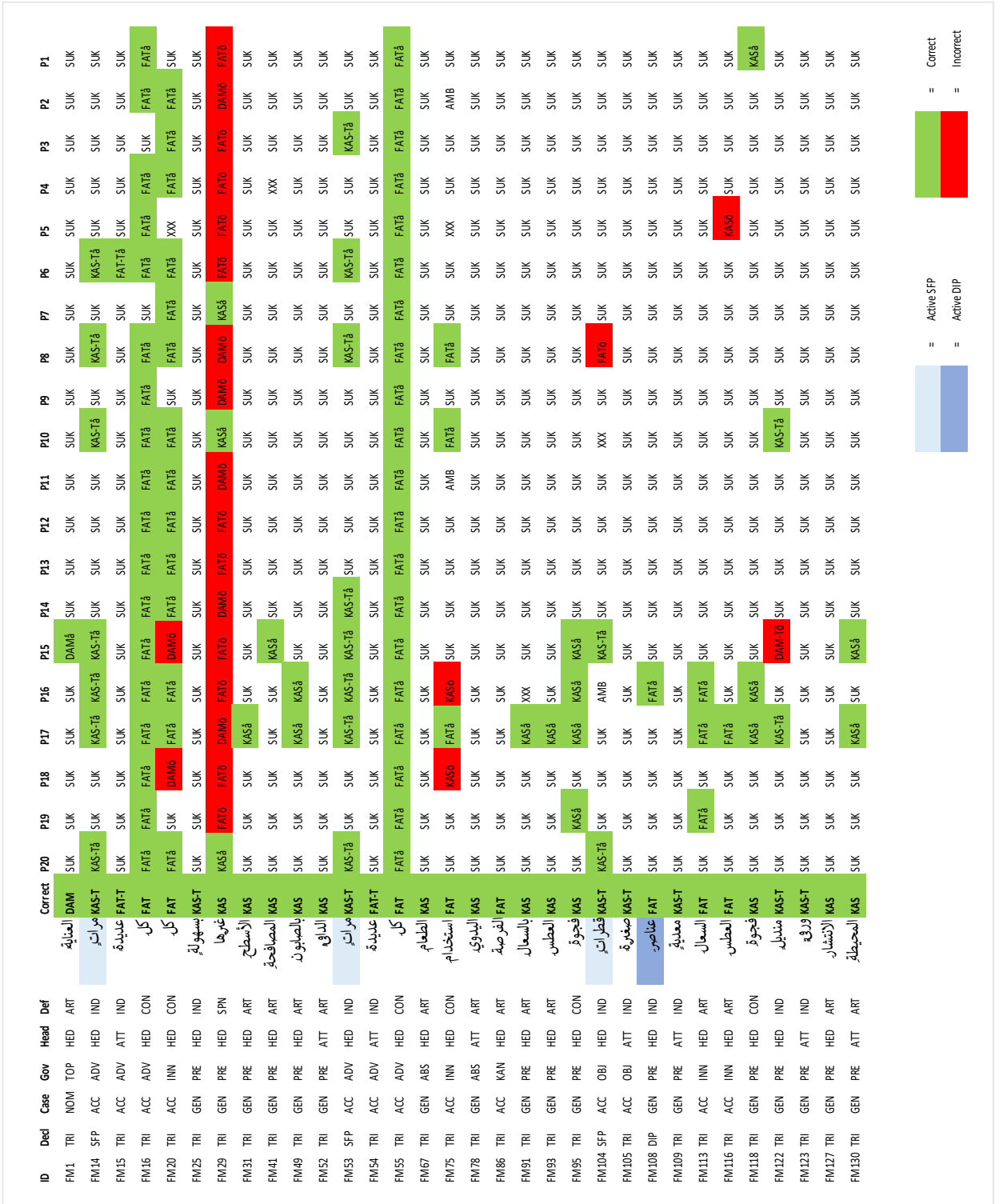
= *hamzat al-waṣl*

= *waqf*

= SPN 3m.s or 2m./f.s.

Appendix D

Words included in the analysis of text FM is shown here, together with results from the recorded audio of the reading aloud. The twenty participants are numbered P1 to P20. A code has been registered for the pronunciation of each participant, and coloured green if they are correct and red if they are incorrect. If the participant didn't pronounce any marker, SUK for *sukūn* has been registered without any added colour.



Appendix E - Consent form

This appendix contains the consent form that was given to the participants in the study. The original document was in Arabic only, the translation into English has been made for this appendix.

الموافقة بموجب النظام الأوروبي العام لحماية البيانات Dataskyddsförordningen

شكرا لمشاركتك في هذه الدراسة في اللغة العربية في جامعة يوتيبوري، قسم اللغات والأدب.

كجزء من الدراسة، أسجل الصوت عندما نتحدث عن النصوص. يتم التسجيل لمساعدتي في التحليل بعد الانتهاء من الدراسة.

يجب أن تتوافق جميع الدراسات في الجامعات السويدية مع النظام الأوروبي العام لحماية البيانات. هذا يعني أنه لا يسمح لي بحفظ التسجيل أو معلومات أخرى عنك دون موافقتك.

مشاركتك مجهولة. فقط الأشخاص الذين يعملون في الدراسة يستمعون للتسجيل. لن يستمع أي شخص آخر إلى التسجيل بدون إذنك. سيتم تخزين التسجيل على جهاز كمبيوتر في منزلي. لن يتمكن أي شخص آخر من الوصول إليها.

في الأطروحة سأستخدم رقمًا لكل شخص في التسجيل. لن تكون هناك معلومات شخصية عنك في الأطروحة.

بالموافقة على هذه المعلومات فإنك توافق على المشاركة. الموافقة صالحة حتى إشعار آخر. لديك الحق في تغيير رأيك وسحب موافقتك في أي وقت. أنت تفعل ذلك الكتابة إلى عنوان بريدي الإلكتروني.

ومع ذلك، فإن المعلومات المدرجة في النتائج التي تم الحصول عليها بالفعل لن تتأثر بسحب موافقتك. إذا كنت تريد تقديم شكوى حول كيفية معالجة المعلومات، فيمكنك الاتصال بممثل حماية البيانات بجامعة يوتيبوري.

عنوان بريدي الإلكتروني guswemag@student.gu.se Magnus Wennerholm

الممثل حماية البيانات بجامعة يوتيبوري dataskydd@gu.se

Consent in accordance with the European General Data Protection Regulation Dataskyddsförordningen
Thank you for participating in this study of the Arabic language at the Department of Languages and Literature at the University of Gothenburg.
As part of the study, I record audio when we talk about the texts. The recording is made in order to help me with the analysis after completing the study.
All studies at Swedish universities must comply with the European General Data Protection Regulation. This means that I am not allowed to store the recording or other information about you without your consent.
Your participation is anonymous. Only persons working with the study will listen to the recording. No one else will listen to the recording without your permission. The recording will be stored on a computer in my home. No one else will have access to it.
For the thesis, I will use a number for each person in the recording. There will be no personal information about you in the thesis.
By agreeing to this information, you agree to participate. The approval is valid until further notice. You have the right to change your mind and withdraw your consent at any time. You do this by writing to my email address.
However, information included in results already obtained will not be affected by your withdrawal of consent. If you would like to file a complaint about how the information is handled, you can contact a data protection representative at the University of Gutenberg.

My email address is guswemag@student.gu.se
Magnus Wennerholm
Data Protection Representative at the University of Gothenburg: dataskydd@gu.se

7 References

Bohas, G., Guillaume, J.-P., & Kouloughli, D. E. (1990). *The Arabic Linguistic Tradition.* London: Routledge.

Ferguson, Charles A. (1959). *Diglossia*. WORD, 15:2.

Göteborgs Stad. (2020). *Att läsa sfi*. Viewed 2020-06-06. https://goteborg.se/wps/portal/start/forskola-och-utbildning/vuxenutbildning-komvux-larvux/svenska-for-invandrare/att-lasa-sfi/

Göteborgs Stad. (2020). *Så här ansöker du till sfi.* Viewed 2020-06-06. https://goteborg.se/wps/portal/start/forskola-och-utbildning/vuxenutbildning-komvux-larvux/svenska-for-invandrare/for-dig-som-ska-ansoka-till-svenska-for-invandrare

Haeri, Niloofar. (2003). *Sacred Language, Ordinary People: Dilemmas of Culture and Politics in Egypt*. New York: Palgrave Macmillan.

Hallberg, Andreas. (2016). *Case Endings in Spoken Standard Arabic: Statistics, Norms, and Diversity in Unscripted Formal Speech*. Studia Orientalia Lundensia. Nova Series 4. Lund: Lund University.

Hallberg, Andreas. (2019). *Towards a grammar of the (non-)use of case marking in Standard Arabic.* (Unpublished)

Isaksson, Bo. *Transkription av arabiska.* https://www.studentportalen.uu.se/uusp-webapp/auth/webwork/filearea/download.action?nodeId=2626929&toolAttachmentId=645280&uusp.userId=guest

Linköpings Universitet. (2018). *Personuppgiftsbehandling vid examinerande moment och examensarbete – vägledning för studenter med kommentarer för handledare och examinatorer*. Linköping: DNR LIU-2018-01721

Milroy, James. 2006. "The ideology of the standard language" In: The Routledge Companion to Sociolinguistics. Edited by Carmen Llamas, Louise Mullany, Peter Stockwell, 133-139.

Milroy, James & Milroy, Lesley. (1991). Authority in language: investigating language prescription and standardisation. (2. ed.) London: Routledge.

Parkinson, D.B. (1993). Knowing Standard Arabic: Testing Egyptians' MSA Abilities. *Perspectives on Arabic Linguistics V Papers from the fifth annual symposium on Arabic linguistics*, *101*, 47-73.

Parkinson, D.B. (1991). Searching for Modern fuṣḥā: Real-Life Formal Arabic. *Al-Arabiyya* (24) 31-64.

Ryding, Karin C. (2005). *A Reference Grammar of Modern Standard Arabic.* New York: Cambridge University Press.

Schulz, David Eugene. (1981). *Diglossia and Variation in Formal Spoken Arabic in Egypt*. Ann Arbor: Unpublished Ph.D. thesis.

Wennerholm, Magnus. (2019). *Användning av kasusmarkörer i undervisning i Standardarabiska vid fyra svenska universitet*. Bachelor thesis. Gothenburg: University of Gothenburg.