Case and sequence in coordination

A corpus-based study of English and Danish

Bang Lauridsen, Freja

2022

Document Version:
Publisher's PDF, also known as Version of record

AARHUS UNIVERSITY

Case and sequence in coordination:

A corpus-based study of English and Danish

Freja Bang Lauridsen

Student ID: 201704932

Master's Thesis

Supervisor: Sten Vikner

Department of English

15.09.2022

INSTITUT FOR
KOMMUNIKATION OG KULTUR
AARHUS UNIVERSITET

Table of contents

Summary

Although much research has been done on case and sequence within coordination structures separately, few studies have investigated the impact the two have on each other. Moreover, very little research has dealt with the striking similarity of English and Danish regarding case and sequence within CoDPs. This thesis, therefore, aimed to explore this matter to uncover the complex system which determines what pronoun cases can be used in what sequences in CoDPs in the two languages. Through the study of four corpora from two periods in time, the behaviour of English and Danish pronouns was examined to determine the special properties that allow for unexpected use of case forms – e.g., the use of the first person singular nominative *I*/*jeg* in non-subject contexts and the use of its oblique counterpart *me*/*mig* in subject contexts. Contributing to the complexity of the behaviour of personal pronouns within CoDPs is the inextricable intertwining of case and the sequence of conjuncts making *Him and I* significantly more acceptable than *I and him.*

From the investigations of the four corpora, it was found that the use of oblique in unexpected syntactic environments is a widespread phenomenon in both Present Day English and Present Day Danish. The high number of occurrences of unexpected oblique indicates that it is not the standard, generative case system alone that determines what morphological case is used where. Another important syntactic mechanism, namely default case, also plays a role in the distribution of case. It is the alternation between these two mechanisms that causes the unexpected and unpredictable use of oblique form in subject position in English and Danish.

Unexpected nominative, also known as hypercorrection, is also apparent in the data of the four corpora, although much less prominent than unexpected oblique. Unexpected nominative has to do with the absence of a syntactic case specification, just as unexpected oblique. Those instances where case specifications have not been given prior to spell-out and where default oblique case has been supplied are more easily overridden than 'regular' oblique. Hence, prescriptive rules and viruses – in this case in the form of nominative – can easily replace the default case resulting in the nominative case form being used unexpectedly.

The corpora investigations of pronouns in CoDPs in English and Danish furthermore demonstrated tendencies that helped uncover the implicit rules which decide what sequences are acceptable in coordination structures and what are not. The primary tendency in the formation of coordination structures is the almost categorical tendency to have *I/jeg* positioned in the second conjunct, while the remaining pronouns in the vast majority of cases are placed in the first conjunct. As is seen from comparisons with other numbers from similar studies, this

so-called politeness norm, which stipulates that the first person pronoun should occur in the last conjunct of the coordination, has only increased in popularity during the past 400 years.

As becomes evident from the current project, it simply cannot be that speakers of English and Danish are wishy-washy when it comes to case and sequence in coordination structures. We are not dealing with random speaker variation but with systematised variation occurring due to certain syntactic properties. The current project has accounted for this systematised variation in case and sequence in English and Danish coordination structures, ascribing the variation primarily to default case and politeness norms.

Abbreviations

Nom.	Nominative
Obl.	Oblique
DP	Determiner phrase
CoDPs	Coordinate determiner phrases
D°	The minimal projection of a determiner phrase (DP)
IP	Inflection phrase
I°	The minimal projection of an inflection phrase (IP)
VP	Verb phrase
V°	The minimal projection of a verb phrase (VP)
NP	Noun phrase
N°	The minimal projection of a noun phrase (NP)
PP	Preposition phrase
P°	The minimal projection of a preposition phrase (PP)
1sg	First person singular
3sg	Third person singular

Freja Bang Lauridsen
201704932

Chapter 1: Introduction

The English and Danish languages provide an interesting backdrop for a discussion of case and sequence within coordinated determiner phrases. Both languages are so-called case-impoverished languages, where case only remains on a limited number of pronouns including the personal pronouns (Jensen 2019, 72; McFadden 2017, 2; Parrott 2013, 213). Compared to languages such as German, Icelandic, and Faroese which have "'rich' inflectional case morphology on a range of elements comprising nominal phrases, including articles, determiners, demonstratives, nouns, pronouns, wh-words, and more" (Parrott 2013, 213), the case systems of English and Danish seem impoverished and uninteresting. Nevertheless, the sparse case systems of English and Danish have several special and unique features, often behaving in unexpected and unexplainable ways. One of the features unique to the English and Danish case systems – and something which is never seen in languages like German and Faroese – is the ability to mix case (Parrott 2009a, 166-169, 180-182). This ability makes (1) and (2) possible and common in everyday English and everyday Danish speech and writing:

(1) That is a matter for <u>him</u> (obl.) and <u>I</u> (nom.).
Det er en sag for <u>ham</u> (obl.) og <u>jeg</u> (nom.).

(2) <u>Him</u> (obl.) and <u>I</u> (nom.) went for a walk.
<u>Ham</u> (obl.) og <u>jeg</u> (nom.) gik en tur.

The ability to mix case is not the only phenomenon worth noting in (1) and (2). There seems to be a connection between the choice of case and the sequence within the coordinate determiner phrases (from here on, referred to as CoDPs), making (2) more acceptable than (3).

(3) <u>I</u> (nom.) and <u>him</u> (obl.) went for a walk.
<u>Jeg</u> (nom.) og <u>ham</u> (obl.) gik en tur.

This complex system which involves both case and sequence and the connection between them, deserves closer investigation. Although much research has been done on case and sequence within CoDPs separately, very few scholars have investigated the impact the two have on each other. Moreover, very little research has dealt with the striking similarity of English and Danish regarding case and sequence within CoDPs.

This thesis seeks to investigate the behaviour of English and Danish pronouns within CoDPs with respect to case and sequence. Through the study of four corpora from two periods in time, this project examines the special properties of personal pronouns in coordination over the last 500 years. One focus is the, at times, unexpected use of case forms – e.g., the use of the first person singular nominative *I*/*jeg* in non-subject contexts (also known as hypercorrection) and the use of its oblique counterpart *me*/*mig* in subject contexts. Another focus of this project is the sequence of the two conjuncts within the CoDPs, including the rising tendency to favour the first person singular nominative pronoun as the second conjunct, not the first. Additional phenomena will be addressed as well for a more complete picture of the behaviour of pronouns within CoDPs in English and Danish.

Before setting out to investigate the behaviour of CoDPs within the four chosen corpora, the groundwork for researching on case and sequence in CoDPs will be laid. Fundamental theory and research on the field will be given, and the framework for the present study will be established. Central and important phenomena such as *abstract Case* and *morphological case* will furthermore be accounted for. Afterwards, in chapter 3, the methodology of the four corpora investigations will be given one by one, and it will be thoroughly explained how numbers and percentages are calculated. In chapter 4, I present the results of the four corpora investigations before highlighting the five most interesting findings within the results. After having stated the five points of interest, the project moves into chapter 5, in which these five findings are examined, explained, and discussed exhaustively. Chapter 5 also provides the reader with a discussion of the possible problems which arise when doing work on historical and comparative syntax. Finally, chapter 6 gives a summary of the findings presented in this project along with suggestions for future research within the field.

Chapter 2: Theoretical framework

2.1 The dismantling of the English and Danish case systems

Although only the last vestiges remain today, English and Danish were once case-rich languages with phonologically distinctive case morphology on a wide range of elements. Descending from the common ancestor Proto-Germanic and before that Proto-Indo-European, the two languages derive from "highly inflecting languages with rich systems of morphological case" (McFadden 2020, 282). Based on the earliest attested Indo-European languages, Proto-Indo-European is thought to have distinguished eight cases: nominative, vocative, accusative, instrumental, dative, ablative, genitive, and locative (Fritz 2003, 264-272). As part of the Proto-Indo-European languages developed into Proto-Germanic, the two cases locative and ablative were lost, resulting in Proto-Germanic only distinguishing six cases (Ringe 2017, 261). Branching off into East-, West-, and North-Germanic, cases were reduced even further. All languages within the Germanic family "have reduced the extent to which cases are distinguished morphologically, though with significant differences in the details of how much has been lost and when" (McFadden 2020, 282). This has led to a contemporary situation where some languages within the Germanic family have retained most of the old, Indo-European case distinctions while others have "reduced morphological case to a few vestiges in the pronominal system" (McFadden 2020, 282).

Both English and Danish belong to the last-mentioned category of Germanic languages, which have lost most of their case distinctions. For English, the levelling and loss of the inflectional endings were traceable even at the earliest stages of Old English (450-1150). In early Old English, distinctions were still made according to four cases, nominative, accusative, genitive, and dative, however, a fair amount of syncretism was already present in the case morphology of Old English nouns and adjectives (Quinn 2005, 8-9). The deterioration of the English case system accelerated in Late Old English and Early Middle English (1150-1350) as the remaining nominal case inflections gradually disappeared (Quinn 2005, 14). The rearrangement of the grammatical system of English took place gradually, being finally established by the Early Modern English period (1450-1700) (Rissanen, Kytö, and Heikkonen 1997, 2).

As in Old English, four morphological cases, nominative, accusative, genitive, and dative, were distinguished in Old Danish (800-1100) (Skautrup 1968, 135). Towards the end of the Old Danish period, the spoken language started levelling out its inflectional endings – a process that did not, however, become evident in written language before the Early Middle Danish period (1100-1350) (Skautrup 1968, 137). During the Early Middle Danish period and the Late Middle Danish period (1350-1500), the loss of the inflectional endings accelerated as the case systems

for nouns, adjectives, and pronouns were reduced (Skautrup 1968, 266). By the end of the Early Modern Danish period (1500-1750), the rearrangement of the Danish grammatical system had been completed.

2.2 Case in Present Day English and Present Day Danish

Only the last vestiges of once rich inflectional systems remain in Present Day English and Present Day Danish. In both languages, case only remains on a limited number of pronouns, including the personal pronouns. The personal pronouns differ morphologically from nouns in that they generally have distinct nominative, accusative, and genitive case forms, whereas nouns have a common nominative/accusative form and a distinct genitive form (Radford 2004, 45). In the current project, the genitive form is left out of any considerations of case. In Present Day English and Present Day Danish, the genitive form is realised either as the possessive ending *'s* in English or as *s* in Danish or as the possessive pronoun forms *his/hans*, *hers/hendes, their/deres*, etc. Thus, the genitive case has undergone a structural change resulting in it no longer functioning as an inflection, but merely as a clitic and a determiner (Rosenbach 2002, 273). For this reason, the genitive is no longer viewed as a 'real' case in Present Day English and Present Day Danish, and thus for the remainder of this project, it is only the nominative, accusative, and dative cases which will be addressed.

Disregarding the genitive forms, Present Day English and Danish distinguish two cases in the personal pronouns: nominative and accusative. For the remainder of this thesis, the accusative form will be referred to as the oblique form. Oblique is a term for the cases which are not nominative (Skautrup 1968, 347) (also sometimes referred to as objective form, due to them being primarily used in object position). I distinguish between nominative and oblique instead of nominative and accusative to avoid any confusion with the Old English and Old Danish accusative form. Several of the present day accusative/oblique pronoun forms do not, in fact, descend from the old accusative form, which is why it seems misleading to refer to the case as accusative. For example, the oblique form of the Present Day Danish third person singular feminine *hende* descends from the dative form of the Early Middle Danish third person singular feminine *henne* (see figure 4). Thus, from now on I distinguish between two cases in Present Day English and Present Day Danish: nominative and oblique. The Present Day English and Present Day Danish personal pronouns and their case forms are given below in figure 1 and 2:

	Singular	
	Nominative	**Oblique**
First person	I	me
Second person	you	you
Third person masculine	he	him
Third person feminine	she	her
Third person neuter	it	it
	Plural	
	Nominative	**Oblique**
First person	we	us
Second person	you	you
Third person	they	them

Figure 1: The paradigm of the Present Day English personal pronouns.

	Singular	
	Nominative	**Oblique**
First person	jeg	mig
Second person	du	dig
Third person masculine	han	ham
Third person feminine	hun	hende
Third person neuter	den/det	den/det
	Plural	
	Nominative	**Oblique**
First person	vi	os
Second person	I	jer
Third person	de	dem

Figure 2: The paradigm of the Present Day Danish personal pronouns.

In the current project, any considerations of the third person singular neuter are left out. One reason for this choice is that the third person neuter is easily confused with other words, such as the Present Day Danish determiner articles *den* and *det*. Another reason is that the third person singular neuter only has one form – a shared nominative/oblique form – in both Present Day English and in Present Day Danish (*it* in Present Day English and *den*/*det* in Present Day Danish). We thus have no way of knowing which case form we are dealing with, making it unnecessary to investigate the third person singular neuter further.

Syncretisms are also found in Present Day English between the second person singular nominative and oblique and the second person plural nominative and oblique. Because they are

all *you*, no distinctions can be made between these four forms, and we are thus compelled to leave these pronouns out of any further investigations as well.

2.3 Case theory

Typologically speaking, all the Germanic languages – including both the present day variants and the older variants of English and Danish – uniformly show nominative-accusative patterns of case marking (McFadden 2020, 283). Nominative-accusative languages (as opposed to ergative-absolutive languages) treat subjects of transitive verbs and subjects of intransitive verbs alike and distinguish them from objects of transitive verbs. Ergative-absolutive languages, on the other hand, treat subjects of intransitive verbs and objects of transitive verbs alike distinguishing them from subjects of transitive verbs (Valin 2001, 35-36).

Case in nominative-accusative languages can be accounted for by the theory of abstract Case (written with a capital C) proposed by Chomsky (1981, 170) as part of his Government and Binding framework:

(4) Nominative is assigned by a finite I° to the NP/DP which it governs.

(5) Accusative is assigned by a V° to the NP/DP which it governs.

(6) Dative is assigned by a P° to the NP/DP which it governs.

(7) Genitive is assigned by some X° in N-projections, say Q, to the NP/DP which it governs (Chomsky 1981, 170; Emonds 2010, 100).

Abstract Case is found in all the world's languages, even in those languages which do not have morphological case. If a language has morphological case (if morphological case features are present in the form of inflectional endings), it is assumed to agree with the abstract Case (Crystal 2008, 67).

In practice, this means the following:

(8) Nominative is the form taken by a DP when it is the subject of a finite clause.

(9) Accusative is the form taken by a DP when it is the object of a verb.

(10) Dative is the form typically taken by a DP to express an indirect object relationship or similar notions.

(11) Genitive is the form typically taken by a DP to express a possessive relationship or some similar 'close' connection (Crystal 2008, 6, 129, 210, 328; McFadden 2017, 10).

Languages such as German, Icelandic, Faroese – and of course Old English and Old Danish – distinguish(ed) all four abstract Cases morphologically. Morphological case in the Germanic languages "generally take[s] the form of inflectional suffixes, usually fusional markers also indicating number, information about inflectional class and (indirectly) gender, sometimes complemented by semi-regular mutations of the stem to which they are applied (i.e., Umlaut)" (McFadden 2020, 283). Case in Germanic pronouns is often realised as "a mixture of suffixation and stem suppletion" (McFadden 2020, 283.) The figures below show the paradigms for the Old English and Old Danish pronouns, including the four abstract Cases and their morphological realisations:

	Singular			
	Nominative	**Accusative**	**Dative**	**Genitive**
First person	ic	mec, mē	mē	mīn
Second person	þū	þec, þē	þē	þīn
Third person masculine	hē	hine	him	his
Third person feminine	hēo	hīe	hire	hire
Third person neuter	hit	hit	him	his
	Dual			
	Nominative	**Accusative**	**Dative**	**Genitive**
First person	wit	uncit, unc	unc	uncer
Second person	git	incit, inc	inc	incer
	Plural			
	Nominative	**Accusative**	**Dative**	**Genitive**
First person	wē	ūsic, ūs	ūs	ūre
Second person	gē	ēowic, ēow	ēow	ēower
Third person	hīe	hīe	him	hira, heora

Figure 3: The paradigm of the Old English personal pronouns (Based on McFadden 2017, 7-8; Bergs and Brinton 2012, 285-286; Lass 1992, 117).

	Singular			
	Nominative	**Accusative**	**Dative**	**Genitive**
First person	iak, jak, ek, iac, jac, iæc, æc, iæk, eg, ieg	mik, mæk, mic	mer, mér, mær	min, mín, miin, mina
Second person	thu, þú, þū, þu, tu	thik, þik, tik, thic, þek	ther, þér, þær	thin, þín, þin
Third person masculine	han, hann	han, hann	hanum, hánum, hānum, honum	hans
Third person feminine	hun, hon	hana	henne, henni, hænni, hennæ, hænnæ, hænæ, hæne, hane, hanæ, hænnj	henna, hennar, hænnar, hænna
	Plural			
	Nominative	**Accusative**	**Dative**	**Genitive**
First person	wi, vér, wir, ui	os, oss, oos	os, oss, oos	war, wār, vár, vār
Second person	i, ér, ír, ir	ither, yðr, iðr, iþær	ither, yðr, iðr, iþær	ithar, yðvar, iðar
Third person masculine	the, þe, thæ	them, thæm	them, thæm	therra, theræ, therræ, there, thærræ, thera, thæræ, thærre
Third person feminine	the, þe, thæ	the, þe, thæ	them, thæm	therra, theræ, therræ, there, thærræ, thera, thæræ, thærre

Figure 4: The paradigm of the Old Danish personal pronouns (Based on Jensen 2018, 61; Bertelsen 1939; Nielsen and Stoklund 2019, 36-37).

As has already been established, and as appears in figures 3 and 4, Old English and Old Danish were highly inflecting languages. Apart from distinguishing the four morphological cases above, distinctions were also made according to gender, number, and person.

In Present Day English and Present Day Danish, most of these inflections are lost – the only remnants being the sparse case system of the personal pronouns, as seen in figure 1 and figure 2. Hence, the rules for assignment of the four cases found on pages 6 and 7 apply in this form in neither English nor Danish since the two languages simply do not distinguish all four cases. Instead, case assignment in English and Danish is typically taken to follow two elementary rules:

(12) Nominative is used when a pronoun is the subject of a finite clause.

(13) Oblique is used when a pronoun is in any other structural context (Parrott 2009b, 276; Jensen, Kragh, and Strudsholm 2018, 75).

Because Present Day English and Present Day Danish only have case distinctions left in their pronouns (distinctions, which even here, are degraded to the bare minimum due to syncretisms), the languages need to rely on other features to express their grammatical information. Many

case-impoverished languages, including Present Day English and Present Day Danish, instead rely on constituent order to express the functional relationships between constituents in a clause (Sapir 2014, 135; Jensen 2002, 161). The question that arises is: How did English and Danish go from being highly inflecting, synthetic languages to case-impoverished and analytical languages where case only remains in pronouns?

2.4 From case-rich to case-impoverished

The reduction of the morphological case system has been quite well studied by a number of scholars – both with regard to English (see Kemenade 1987; Allen 1995; Pintzuk 2002; Lass 1992) and with regard to Danish (see Jensen 2002; 2003; Jørgensen 2002; Brøndum-Nielsen 1935; Heltoft and Nielsen 2019). Traditionally, the dismantling of the case system is accounted for in more or less in the same way in English and Danish: As "the direct result of certain phonological changes whereby the pronunciation of the vowel in unstressed syllables became less distinct (...) thereby causing a massive loss of inflectional case markings" (Jensen 2003, 221). As it became harder to distinguish between the various inflectional suffixes, the languages were forced to rely on other features to express the functional relationships between the constituents in a clause. It is traditionally believed that English and Danish gained a more fixed constituent order to be able to discriminate between the different syntactic relations of the sentence (Jensen 2002, 163; 2003, 221; Pintzuk 2002, 382). Thus, one option is that the fixed constituent order was introduced as a form of repair mechanism to maintain the order of the languages as the inflectional suffixes were lost.

As Pintzuk (2002, 382) points out, several scholars view the development the opposite way around suggesting that "the fixing of constituent order permitted the gradual loss of overt morphology." Jensen (2003, 221) supports this perception arguing that "rather than being the result of the change in stress pattern, a more fixed word order might have been a participant facilitating the dismantling of the case system." With a fixed constituent order, there is no longer a need to maintain distinct inflectional case endings for the purpose of determining the syntactic functions[1]. In the case of English and Danish, this resulted in a loss of inflectional case endings (Jensen 2002, 163-164).

[1] It should be noted that, although there is a tendency for fixed constituent order and a sparse inflectional case system to go hand in hand in Germanic languages, this is not always the case. Consider, for example, Icelandic, which distinguishes four cases but at the same time has a rather fixed constituent order (Thráinsson 2007, 2, 21), or Dutch, which oppositely has a relatively free constituent order but where case to all intents and purposes is dead (Donaldson 2017, 23).

Despite the varying explanations as to how the loss of overt case morphology came about, there is general agreement about how the specific loss of inflectional endings took place. In English, a number of major transformations occurred within the pronoun system from the Old English period to the Late Middle English period: the dual was lost; dative and accusative merged so that there were at most three case distinctions – nominative, oblique, and genitive; the old, third person singular nominative *hēo* was replaced by the new form *sche* (an early form of *she*); the genitive case forms was 'detached' from the pronoun paradigm, and came to function rather as adjectives and determiners than true case forms; the old, third person plural forms beginning with *h* began to yield to a new (Scandinavian) type beginning with *th* – the originals of *they*, *their*, *them* (Lass 1992, 117). After undergoing these transformations, English was left with a pronoun paradigm similar to the present day paradigm (see figure 1).

Similarly to English, the Danish pronoun system has undergone a number of transformations between the Old Danish period and the present day: dative and accusative have merged so that there are only three case distinctions – nominative, oblique, and genitive; the genitive case forms have been 'detached' from the pronoun paradigm, and has come to function rather as adjectives and determiners than true case forms; the third person plural feminine, masculine, and neuter have merged into one form covering all grammatical genders.

Apart from all these structural changes, the English and Danish case systems also changed with respect to how they were used. Traditionally, case was distributed as described in section 2.3, where nominative = subjects of finite clauses, accusative = objects of verbs, dative = indirect objects of verbs, and genitive = a possessive relationship. However, at some point in the development from Old English and Old Danish to Present Day English and Present Day Danish, as these four case distinctions were narrowed down to two (nominative and oblique), it became possible to use case in new ways. This phenomenon will be further addressed in the next section.

2.5 Unexpected use of case in Present Day English and Present Day Danish

Following the standard conventions of case in generative linguistics (described in section 2.3), we expect the following distribution of case in Present Day English and Present Day Danish: Subjects of finite clauses are assigned nominative case and everything else (most often that is objects (of verbs) and complements of prepositions) is assigned oblique case (e.g., Crystal 2008, 6, 129, 210, 328; Quinn 2005, 26; McCreight 1988, 2; Chomsky 1981, 170; 1993; Pollard and Sag 1994; Burzio 2000).

Thus, in Present Day English and Present Day Danish, we would expect the pronoun in (14) to be marked for nominative case because it is part of the CoDPs which make up the subject of the finite clause:

(14) Peter and I (nom.) went for a walk.
Peter og jeg (nom.) gik en tur.

Likewise, we expect the personal pronoun in (15) to be marked for oblique case because it is part of the CoDPs whose syntactic function is the complement of the preposition.

(15) That is a matter for Peter and me (obl.).
Det er en sag for Peter og mig (obl.).

Consider now the examples below, which are extracted from the British National Corpus (BNC) and KorpusDK:

(16) Me (obl.) and my boyfriend play cards, and watch telly, and that sort of thing (*British National Corpus*, n.d.).

(17) Mig (obl.) og Jeppe var begyndt at komme sammen igen.
"Me (obl.) and Jeppe had started seeing each other again" (*KorpusDK*, n.d.).

(18) It has given my wife and I (nom.) a lot of comfort (*British National Corpus*, n.d.).

(19) Da vi kom, omfavnede han både min kone og jeg (nom.).
"When we arrived, he embraced both my wife and I (nom.)" (*KorpusDK*, n.d.).

Contrary to what we would expect based on the standard case conventions in generative linguistics, the CoDPs which make up the subjects in (16) and (17) (*Me and my boyfriend* and *Mig og Jeppe*) contain the oblique pronoun *me/mig*. In the same way, the CoDPs which make up the objects in (18) and (19) (*my wife and I* and *min kone og jeg*) contain the nominative pronoun *I/jeg*. Thus, we have a situation where the first person singular oblique case, which according to the standard case conventions, only appears in complement position, is used in a subject context. Likewise, it turns out that the first person singular nominative case, which, according

to the standard case conventions, is solely used in subjects, can appear in object contexts. As shown in examples (16)-(19), this unexpected case situation is present in both Present Day English and Present Day Danish.

Not only can the nominative and oblique case forms appear in unexpected contexts, as shown above, but the case forms can also occur mixed within CoDPs in both English and Danish (Parrott 2009a, 166-169, 181-182):

(20) Him (obl.) and I (nom.) got on very well together (*British National Corpus*, n.d.).

(21) I dag er det en sød hemmelighed mellem hende (obl.) og jeg (nom.).
"Today it is a sweet secret between her (obl.) and I (nom.)" (*KorpusDK*, n.d.).

In (20), the CoDPs *Him and I* form the subject. Thus, we would expect both pronouns to be in their nominative form: *He and I*. In (21), the CoDPs *hende og jeg* are the complement of the preposition; thus, we would expect both pronouns to be in their oblique form: *hende og mig*. Despite what is expected based on the standard case conventions in generative linguistics, both (20) and (21) are perfectly acceptable and intelligible, and both constructions are well attested in both written and spoken English and Danish (Parrott 2009b, 285). At the same time, the standard case distribution exemplified in (22) and (23) is still acceptable:

(22) He (nom.) and I (nom.) got on very well together.

(23) I dag er det en sød hemmelighed mellem hende (obl.) og mig (obl.).
"Today it is a sweet secret between her (obl.) and me (obl.)".

The situation is thus that speakers of English and Danish are free to choose between (20) and (22) and between (21) and (23). Other combinations of case are also socially salient, e.g., *He* (nom.) *and me* (obl.) and *hun* (nom.) *og mig* (obl.) – although significantly less common.

Based on the above phenomena, the following questions arise:

- How is it possible for two pronouns within CoDPs to have different case?
- Why is it possible for nominative pronouns to occur in complement position, and oppositely, why is it possible for oblique pronouns to occur in subject position?
- How can this unexpected use of case exist simultaneously with the standard case system (nominative = subjects of finite clauses and oblique = all other)?

I will return to address these questions in section 5.2.

2.6 Case and sequence in CoDPs

Above, we saw that both *Him* (obl.) *and I* (nom.) *got on very well together* and *He* (nom.) *and I* (nom.) *got on very well together* were socially salient. *He* (nom.) *and me* (obl.) *got on very well together,* on the other hand, is far less common and less acceptable – in fact, there are no results of this constellation at all in the BNC or in KorpusDK when translated into Danish (*British National Corpus*, n.d.; *KorpusDK*, n.d.). For some reason, this combination of case with the nominative pronoun *he* in first position followed by the oblique pronoun *me* appears way less acceptable than any of the other combinations.

A similar phenomenon is seen when we take (20) and (22) and switch around the pronouns, giving us the following two sentences:

(24) I (nom.) and him (obl.) got on very well together.

(25) I (nom.) and he (nom.) got on very well together.

The only thing that has changed from (20) to (24) and from (22) to (25) is the sequence of the pronouns within the CoDPs. Despite this, and even though they contain the exact same pronouns, (24) and (25) are significantly less acceptable than (20) and (22). In fact, there are no occurrences of neither *I and him* nor *I and he* in the entire BNC (*British National Corpus*, n.d.), confirming exactly how unpopular these sequences are. How come sequences of pronouns such as *him and I* and *he and I* be much more acceptable and popular as subjects than *I and him*, *I and he*? And likewise, how come *he and me* be so unacceptable in subject position among English speakers? It seems as if there are some unspoken and implicit rules which we follow when choosing the case and the sequence of the personal pronouns within the clauses that we utter.

With respect to the complement position, the same phenomenon is observable. (21) and (23) are perfectly acceptable, whereas examples such as (26) and (27) are very uncommon and unacceptable among native speakers:

(26) I dag er det en sød hemmelighed mellem hun (nom.) og jeg (nom.).
"Today it is a sweet secret between she (nom.) and I (nom.)".

(27) I dag er det en sød hemmelighed mellem hun (nom.) og mig (obl.).
"Today it is a sweet secret between she (nom.) and me (obl.)".

For some reason, the sentence becomes significantly less acceptable when the third person singular feminine pronoun is in its nominative form *hun* opposed to when it is in its oblique form *hende* as seen in (21) and (23). How come the nominative *hun* is so unpopular in complement position when (21), which also has a nominative pronoun in the CoDPs in complement position (*hende og jeg*), is completely acceptable?

Just as was the case in the subject position above, the acceptability of the sentence also decreases if we switch around the two pronouns of the CoDPs. Consider for example (28) and (29), which are identical to sentences (21) and (26) apart from the fact that they have the pronouns of the CoDPs in the opposite sequence.

(28) I dag er det en sød hemmelighed mellem jeg (nom.) og hende (obl.).
"Today it is a sweet secret between I (nom.) and her (obl.)".

(29) I dag er det en sød hemmelighed mellem jeg (nom.) og hun (nom.).
"Today it is a sweet secret between she (nom.) and I (nom.)".

Examples (28) and (29) are very rarely used in Danish (nor in English), which is emphasised by the lack of occurrences of these sequences of pronouns in the KorpusDK (*KorpusDK*, n.d.).

Thus, what can be deduced from all the above is that there seems to be some sort of rules that decide what case can be used in what sequence in CoDPs. Based on the acceptability and lack of acceptability of examples (14) to (29), the following rough observations are made:

- Pronouns within CoDPs can be mixed with respect to case and can appear in positions which are unexpected according to standard case theory in generative linguistics.
- Some combinations of case are more acceptable than others, and some sequences are better than others.

As has become evident from the above six sections, English and Danish are remarkably similar with respect to case and sequence within CoDPs. Unlike most other Germanic languages, "pronominal case form variation in CoDPs and other structures is salient to native speakers" of both English and Danish (Parrott 2009b, 284-285), meaning that the case distribution in the two languages often diverges from the standard case conventions as was described in section 2.5.

Later in the present thesis, more observations regarding the relationship between case and sequence will be recorded. Together with the rough observations noted above, these will be investigated, discussed, and systematised to understand better the special properties and the similarities between the two languages and their case systems.

Chapter 3: Methodology

To understand the behaviour of English and Danish pronouns within CoDPs with respect to case and sequence, it was decided to work diachronically engaging a total of four corpora from two periods in time: the Early Modern English period (1450-1700) and Early Modern Danish period (1500-1750) as well as present day (from approx. 1980 and until present day). The four corpora that were employed in the collection of data were: a corpus consisting of the Danish playwright Ludvig Holberg's complete works, a corpus consisting of the British playwright William Shakespeare's complete works, KorpusDK, and the British National Corpus (BNC).

The Holberg corpus and the Shakespeare corpus are limited to the works of the aforementioned playwrights, and this is reflected in the relatively small size of the corpora. The Holberg corpus consists of 708,558 words (*Holberg Corpus* 2022), and the Shakespeare corpus consists of 958,268 words (*Shakespeare Corpus* 2022). In comparison, KorpusDK consists of 56 million words (Det Danske Sprog- og Litteraturselskab, n.d.), and the BNC consists of 100 million words (Burnard 2015). Below I will account for how the investigation of pronouns in CoDPs took place in the different corpora, beginning with the Holberg corpus. It should be noted that the investigation of the Shakespeare corpus was conducted using the exact same methods as in the Holberg corpus, and to avoid repetition, it is mainly the Holberg corpus whose methods will be covered. A brief note on the Shakespeare corpus is nevertheless also included with the purpose of describing any discrepancies from the methods used in the Holberg investigation. Hereafter, the KorpusDK investigation and BNC investigation are both covered exhaustively.

3.1 Early Modern Danish: Holberg Corpus

The works of the Danish playwright Ludvig Holberg were chosen for the Early Modern Danish period part of the pronoun investigation. Some scholars (see Jensen 2017, 7-13; Levin 1844; Petersen 1858, 759-760; Dyrlund 1895, 203) have argued that Holberg's works cannot be used as evidence of the language situation in the first half of the 18th century. This claim is based on the fact that Holberg was heavily influenced by languages other than Danish – including Latin and his native language, Norwegian – and thus that his written language is not reliable as an example of Early Modern Danish.

I, however, did choose to use the works of Holberg in the current project due to the lack of alternatives. Unfortunately, there exists a limited number of sources from the Early Modern Danish period, and many of the existing texts are written in verse making them unreliable as evidence for everyday speech at that time. Most of Holberg's texts are written in prose, and

furthermore, his works include a large number of plays which imitates everyday dialogue. Another reason for choosing Holberg's works for the Early Modern Danish part of the research was that these texts are highly comparable with the works of the British playwright William Shakespeare, which I intended to use for the Early Modern English part of the research.

The complete works of Ludvig Holberg were compiled and arranged into a corpus (*Holberg Corpus* 2022). The Holberg corpus (in TXT-file format) was then uploaded to AntConc – a corpus analysis toolkit for concordancing and text analysis (Anthony 2022). Using the search strings in figure 5 below, I searched for CoDPs containing at least one nominative pronoun and CoDPs containing at least one oblique pronoun. Searches were conducted with the pronouns as both the first and the second conjunct, as seen below.

Search strings used in the Holberg corpus

Nominative forms		Oblique forms	
jeg og \w	\w og jeg	mig og \w	\w og mig
du og \w	\w og du	dig og \w	\w og dig
han og \w	\w og han	ham og \w	\w og ham
hun og \w	\w og hun	hende og \w	\w og hende
vi og \w	\w og vi	os og \w	\w og os
i og \w	\w og i	jer og \w	\w og jer
de og \w	\w og de	dem og \w	\w og dem

Figure 5: AntConc search strings for all Danish nominative and oblique personal pronoun forms. Searches were conducted with the pronoun in both first and second position. \w is the symbol for an alphanumeric wildcard.

All 28 searches in AntConc (*jeg og \w, \w og jeg*, *du og \w, \w og du, han og \w, \w og han*, etc.) were run separately using the 'regex' (regular expression) setting. Each search triggered a number of hits, which indicate the number of occurrences of the used search string in the Holberg corpus. Take, for example, *mig og \w*, which gives us exactly 60 hits – meaning that in the entire Holberg corpus, there is 60 occurrences of *mig* followed by an alphanumeric character.

All these 60 results (each result including a few surrounding words for context purposes) were extracted from AntConc, arranged into a table, and uploaded to Microsoft Excel. In Excel, the table containing all the results of the given search was examined manually – one result at a time – with the purpose of judging whether that result was, in fact, a case of CoDPs or another syntactic constellation, which had been impossible to rule out during the AntConc search.

Relevant results are examples of CoDPs containing at least one pronoun as either the first or the second conjunct as in the examples below:

(30) Beed et got Ord for mig og min Herre.
"Pray a good word for me and my lord" (*Holberg Corpus* 2022, "Mascarade").

(31) Hans gode Moer har stillet min Papa og mig tilfreds igien.
"His good mother has satisfied my father and me again" (*Holberg Corpus* 2022, "Mester Gert Westphaler eller Den meget talende barber").

Irrelevant results are sentences which contain *mig og \w*, but where *mig og \w* are not CoDPs. In the example below, the alphanumeric wildcard \w is not a noun, an adjective, a pronoun, or any other word class that can form a coordination. Instead, \w is a verb which cannot be coordinated with *mig* and hence cannot form CoDPs:

(32) Nu rører Samvittigheden mig og nøder mig at bekiende…
"Now the conscience gets to me and forces me to confess…" (*Holberg Corpus* 2022, "Hexerie eller Blind Allarm").

Irrelevant results can also be examples where the alphanumeric wildcard \w is, in fact, a noun, an adjective, a pronoun, or any other word class that would be able to make up CoDPs, but where \w and the following word do not form a constituent due to another syntactic constellation. Consider (33), where the Present Day English translation demonstrates that *mig og mange andre* are not CoDPs, but instead two separate DPs, which do not form a constituent:

(33) Byens Øvrighed viste mig hvordan jeg kunde hævne mig og mange andre trængte paa for at faa mig til at gaa til Makronerne.
"The city's authorities showed me how I could take my revenge and many others exerted pressure on me to get me cracking" (*Holberg Corpus* 2022, "Første Brev til en højvelbaaren Herre. 1728").

All the irrelevant occurrences, which the searches in AntConc (*jeg og \w, \w og jeg, du og \w, \w og du, han og \w, \w og han*, etc.) generated, were sorted out manually. In this process, a large number of results were discarded. Ideally, there had been a way of narrowing down the results of the AntConc searches even more so that all the irrelevant results resembling (32) and

(33) would be sorted out automatically. Unfortunately, there is no way for AntConc to distinguish between the relevant example in (30) and the irrelevant example in (33). This would require a more complex software that could distinguish different syntactic structures from each other and identify only the correct ones; that is, the ones that constitutes CoDPs. However, because no such software exists, the sorting was done manually. Luckily this did not constitute a major problem for the research because the Holberg corpus is a relatively small corpora (708,558 words) with relatively few instances of CoDPs.

After all results for each of the 28 searches had been examined manually and all irrelevant instances such as example (32) and example (33) had been discarded, I was left with 28 tables consisting of occurrences of CoDPs containing at least one pronoun in either first or second conjunct. For the full overview of the 28 tables constituting the Early Modern Danish corpora investigation, see appendix 1. For comparison purposes, the number of relevant results for each of the 28 categories (*jeg og \w, \w og jeg, du og \w, \w og du, han og \w, \w og han*, etc.) was counted. The total number of relevant results can be found in the top bar of each table.

After having sorted and arranged all searches into tables, the results were examined one by one once more, this time with the purpose of analysing each instance and determine the use of case (nominative or oblique) in relation to the surrounding sentence. Each instance of *pronoun og \w* or *\w og pronoun* was colour-coded with either green, red, or yellow. Green indicating that the use of case is expected, red indicating that the use of case is unexpected, and yellow indicating that there is doubt about the use of case. The instances that are marked with yellow are typically followed by a short explanation describing why there is doubt about the case (see appendix 1).

3.2 Early Modern English: Shakespeare Corpus

The works of the British playwright William Shakespeare were chosen as the empirical foundation for the Early Modern English part of the project. The complete works of Shakespeare were compiled and arranged into a corpus (*Shakespeare Corpus* 2022) and uploaded to AntConc. From here on, the same searches that were run in the Holberg corpus were also run in the Shakespeare corpus – just in English (see figure 6 below). Note that the obsolete English case distinction *thou/thee* was used when searching for CoDPs containing second person singular pronouns. Had it not been for this distinction, one would not be able to tell the second person singular pronouns from the second person plural pronouns, and furthermore one would not be able to tell the nominative case from the oblique case due to syncretisms between these pronoun categories. In the second person plural pronouns (marked with red in figure 6), no case

distinction is found (both the nominative and the oblique form is *you*), and thus we cannot tell the two cases apart. For this reason, the second person plural category is left out of the English corpora investigations (this goes for both the Shakespeare corpus and the present day investigations in the BNC).

Search strings used in the Shakespeare corpus

Nominative forms		Oblique forms	
i and \w	\w and i	me and \w	\w and me
thou and \w	\w and thou	thee and \w	\w and thee
he and \w	\w and he	him and \w	\w and him
she and \w	\w and she	her and \w	\w and her
we and \w	\w and we	us and \w	\w and us
you and \w	\w and you	you and \w	\w and you
they and \w	\w and they	them and \w	\w and them

Figure 6: AntConc search strings for all English nominative and oblique personal pronoun forms. Searches were conducted with the pronoun in both first and second position. \w is the symbol for an alphanumeric wildcard.

After having run all of the English searches in the Shakespeare corpus and sorted out the irrelevant examples following the same method as with the Holberg searches, I was left with 24 tables (as described above, the second person plural searches have been left out, since it was impossible to distinguish between the nominative and the oblique form in this category). For the full overview of the 24 tables constituting the Early Modern English corpora investigation, see appendix 2. For comparison purposes, the number of relevant results for each category (*I and \w, \w and I, thou and \w, \w and thou, he and \w, \w and he*, etc.) were counted. The total number of relevant results can be found in the top bar of each table.

Just as in the Early Modern Danish investigations, the final results of the Shakespeare corpus search were examined once more, this time with the purpose of analysing each instance and determining the expectedness of case in relation to the surrounding sentence. Each instance of *pronoun and \w* and *\w and pronoun* was colour-coded with either green, red, or yellow, indicating the expectedness of case.

3.3 Present Day Danish: KorpusDK

The investigations of pronouns in KorpusDK and the BNC took place somewhat similarly to the investigations of the Holberg corpus and the Shakespeare corpus. The main difference between the 17th and 18th century playwright corpora and the present day corpora is that I have compiled the Holberg and the Shakespeare corpus myself for the purpose of this project, whereas KorpusDK and the BNC are highly advanced corpora compiled by professional corpus linguists and computational linguists over a time span of several decades. Thus, KorpusDK and the BNC are significantly more advanced than the playwright corpora – both with respect to size and tagging. Apart from containing respectively 70 and 125 times as many words, both KorpusDK and the BNC have been tagged for grammatical information using sophisticated tagging systems such as CLAWS and have online interfaces with advanced search engines to increase the user-friendliness of the corpora.

For the investigation of pronouns within CoDPs in Present Day Danish, I consulted KorpusDK's online interface. As part of its online interface, KorpusDK has several different search functions, including 'standard søgning' (standard search), 'udvidet søgning' (expanded search), and 'formel søgning' (formal search). The 'formel søgning' (formal search) setting of the KorpusDK search engine was chosen since it enabled me to execute the most confined and restricted searches. The formal search strings were written in the search language from the corpus processor CQP (Corpus Query Processor), also known as CQL (Corpus Query Language), which KorpusDK uses, and followed the formulas found in figure 7 below.

Search string formula for searching in the online KorpusDK search engine

jeg og
```
[word="jeg" & pos="PERS"] []{0,0} [word="og"] []{0,0} ([pos="N|ADJ|PROP|DET|ART|PERS|INDP|NUM"])
```

og jeg
```
([pos="N|ADJ|PROP|DET|ART|PERS|INDP|NUM"]) []{0,0} [word="og"] []{0,0} [word="jeg" & pos="PERS"]
```

Figure 7: Search strings formulas for searching in the online KorpusDK search engine. KorpusDK uses the search language from the corpus processor CQP (Corpus Query Processor), also known as CQL (Corpus Query Language) (KorpusDK, n.d.).

By means of these search string formulas, I was able to restrict my results to a personal pronoun followed by either a noun, an adjective, a proper name, a determiner, an article, a personal pronoun, another type of pronoun, or a numeral. Or the opposite way around: a noun, an adjective, a proper name, a determiner, an article, a personal pronoun, another type of pronoun, or a numeral followed by a personal pronoun. Thus, instead of having the pronoun followed or

preceded by the alphanumeric wildcard \w, which would give us all words beginning with an alphanumeric character and thus give us a great number of hits which were not examples of CoDPs, KorpusDK's formal search setting and the above search string formula enable us to narrow down the search significantly. Essentially, this means that we end up with significantly fewer irrelevant results compared to a standard AntConc search.

As in the AntConc searches described in sections 3.1 and 3.2, I conducted a search for each pronoun in either first or second position – a total of 28 searches (see figure 5). Each search in KorpusDK was run separately, each search triggering a number of hits indicating the number of occurrences of the used search string in the entire KorpusDK. In some instances, the search strings triggered a large number of hits. The *og jeg* search string found in figure 7, for example, triggered 16,916 hits. Unfortunately, due to lack of engine capacity, the KorpusDK interface is not able to show more than 5,000 results of a given search, meaning that if a search string triggers more than this number of hits, the results are randomly narrowed down to 5,000 results. This issue is, of course, unfortunate since one would like as much data as possible to solidify one's findings. However, an upper limit of 5,000 does not constitute a major problem for this exact investigation. As in both the Holberg corpus searches and the Shakespeare corpus searches, all results need to be evaluated manually, and no matter the upper limit of results from KorpusDK, it would be too time-consuming to evaluate and analyse more than 5,000 occurrences manually. The only way we would have been able to evaluate more than 5,000 results would be if the evaluation of results were conducted automatically. However, this would require highly advanced software, which does not exist at the moment.

Having run the *og jeg* search in KorpusDK, I was left with 5,000 hits, which have been narrowed down from 16,916 hits. All 5,000 results of the search (each result including a few surrounding words for context purposes) were extracted from KorpusDK, converted from a DOCX-file into a TXT-file, and uploaded to the corpus analysis toolkit, AntConc. From here on, the methodology resembles the playwright corpora investigations to the letter. Using the search strings found in figure 5, I searched for *\w og jeg* in the KorpusDK results, which have just been extracted from the KorpusDK interface. This process was repeated with all KorpusDK searches *(jeg og \w, \w og jeg, du og \w, \w og du, han og \w, \w og han,* etc.).

Why would I need to upload the KorpusDK results to AntConc and conduct another round of searches, one might ask? Unfortunately, the KorpusDK search engine does not take punctuation into account, meaning that the KorpusDK search strings cannot rule out examples as in (34) and (35):

(34) Jeg var ny elev, og jeg troede, jeg skulle gå den slagne vej.
"I was a new student and I assumed I would follow the usual way" (KorpusDK, n.d.).

(35) Men det er virkelig en god ting for os. Og jeg tror, publikum finder det langt mere uformelt.
"But it is in fact a good thing for us. And I think the audience finds it far more informal" (KorpusDK, n.d.).

Examples such as (34) and (35) that have some sort of punctuation immediately before or immediately after the coordinating conjunction can under no circumstances constitute CoDPs. Thus, all such examples must be discarded since they make up irrelevant examples.

Fortunately, AntConc's search engine does take punctuation into account – as opposed to KorpusDK's search engine, which does not. Thus, by running all the results extracted from KorpusDK through AntConc using the search strings found in figure 5, I can eliminate those examples that contain some sort of punctuation, and which are not relevant example of CoDPs. This manoeuvre lessens the burden of having to manually sort out all these irrelevant examples – a process that would be extremely time-consuming considering the size of KorpusDK.

After having run my KorpusDK results through AntConc – repeating for all 28 search strings – the results from AntConc were extracted, arranged into a table, and uploaded to Microsoft Excel. Here all results were examined manually to sort out the occurrences which were not CoDPs but another syntactic constellation. Although still very time-consuming, this process was made somewhat more manageable by carrying out the second round of searches in AntConc, discarding many irrelevant examples. Finally, after having examined all the remaining results one by one, all results were analysed to determine the use of case in relation to the surrounding sentence and colour-coded accordingly, just as was done in the Early Modern Danish and Early Modern English investigations. For the full overview of the 28 tables constituting the Present Day Danish corpora investigation, see appendix 3. For comparison purposes, the number of relevant results for each of the 28 categories (*jeg og \w, \w og jeg, du og \w, \w og du, han og \w, \w og han*, etc.) was counted. The total number of relevant results can be found in the top bar of each table.

3.4 Present Day English: The British National Corpus

To investigate Present Day English pronouns within CoDPs, I accessed the BNC online interface hoping that it could perform the same searches as in KorpusDK, providing me with results comparable to those extracted from KorpusDK. Unfortunately, I quickly realised that the online version of the BNC only had the capacity to display 500 results of a given search. Compared to KorpusDK's upper limit of 5,000 results, it seemed insufficient to draw conclusions based on only 500 results. Furthermore, it would significantly complicate the comparisons between the two present day corpora if the two samples were of such different sizes. This issue constituted a major problem for my project, and I had to consider alternative options for the Present Day English part of the investigations. Among other considerations, I explored the potential of working with another present day Anglophone corpus, such as the Corpus of Contemporary American English (COCA) or Brown University Standard Corpus of Present-Day American English. However, both these corpora are collections of American English texts, and although British and American English are indeed very similar, the choice of an American corpus would still complicate the comparison with the Early Modern English Shakespeare results significantly.

Unfortunately, I found that there does not exist a great British English alternative to the BNC. Either the corpora would be too small – and thus incomparable with KorpusDK, which contains 56 million words – or they would be too specific, containing only texts from one genre, such as the British Academic Written English corpus, which consists only of academic works written at universities in the UK.

Thus, my only real option was finding some way to access the BNC without using the online interface. One opportunity was to download the entire BNC (that is, all 100 million words) to my computer and conduct the searches manually. Not only would this method be extremely user-hostile, but it would also, in fact, require me to encode my own search machine that would take tagging etc., into account. Due to a lack of coding skills and time, this option was discarded.

Luckily, during my considerations of how to go about the searches in the BNC, I encountered the computer software LancsBoxX. Developed at Lancaster University, LancsBoxX is a corpus analysis tool developed with advanced XML (Extensible Markup Language) capabilities designed for very large corpora (Brezina and Platt 2022). Because of these features, the programme is suitable for working with the BNC. In fact, the software comes with both the British National Corpus 1994 and the newer and updated version of the corpus, British National Corpus 2014, releasing me from all speculations of having to code my own search engine to be able to work with the BNC.

I chose to run my searches in the British National Corpus 1994 because this version of the corpus fits better timewise with KorpusDK, which was compiled between 1983 and 2002. This decision was taken to make the two corpora even more comparable. After having accessed the BNC through LancsBoxX, the searches were conducted more or less similar to the searches carried out in KorpusDK. Using the search string formulas in figure 8, I was able to carry out just as confined and restricted searches as in the KorpusDK search engine.

Search string formula for searching in the British National Corpus through LancsBoxX

i and
[word="i"][word="and"][pos="J.*|NN.*|X.*|A.*|D.*|P.*|AP.*|MC.*"]

and i
[pos="J.*|NN.*|X.*|A.*|D.*|P.*|AP.*|MC.*"][word="and"][word="i"]

Figure 8: Search strings formulas for searching in the British National Corpus through LancsBoxX. LancsBoxX uses the search language from the corpus processor CQP (Corpus Query Processor), also known as CQL (Corpus Query Language) (Brezina and Platt 2022).

The search strings are formulated in the same search language that KorpusDK uses, namely the language from the corpus processor CQP (Corpus Query Processor), also known as CQL (Corpus Query Language). There are, however, a few minor differences between how the KorpusDK search strings and the LancsBoxX search strings are written up, but the basic syntax is the same.

As in the three other corpus investigations described above, each search in the BNC was run separately, each search triggering a number of hits indicating the number of occurrences of that search string in the entire BNC. Due to syncretisms between the second person singular nominative and oblique and the second person plural nominative and oblique, these four searches have been left out of the investigations. No distinction is found between the four categories (all forms are *you*), meaning that we cannot tell them apart and thus cannot use them in our examinations of case in Present Day English pronouns. For this reason, we are left with the 20 searches found in figure 9.

Searches carried out in the British National Corpus

Nominative forms		Oblique forms	
i and	and i	me and	and me
you and	and you	you and	and you
he and	and he	him and	and him
she and	and she	her and	and her
we and	and we	us and	and us
you and	and you	you and	and you
they and	and they	them and	and them

Figure 9: Searches carried out in the British National Corpus. Searches were conducted with the pronoun in both first and second position. The forms that have been left out of the searches are marked with red.

Just as in KorpusDK, some search strings triggered a large number of hits while others triggered only a few. The *and us* search, for example, triggered only 67 results, whereas the *and I* search triggered 31,530 results. Whereas KorpusDK, due to lack of engine capacity, would reduce this high number of hits to its upper limit of 5,000, the BNC – when accessed through LancsBoxX – has no upper limit of the number of results it can show. Thus, our issue with the BNC only wanting to show 500 results has been resolved through LancsBoxX.

However, 31,530 results were quite a handful when having to sort all instances manually. For this reason and to increase comparability between the BNC results and the KorpusDK results, the BNC searches that triggered a large number of hits were deliberately reduced to a sample of 5,000 results – just as the KorpusDK search engine had done automatically.

The results of each search *(I and, and I, he and, and he, she and, and she,* etc.) were extracted from LancsBoxX, one at a time. The searches which triggered more than 5,000 hits were inserted into a Microsoft Excel document. Using the Excel RAND function, I randomised all the hits of a given search (e.g., the *and I* search triggering 31,530 results) so that all results were listed in entirely random order. I then extracted the first 5,000 results and inserted them into a DOCX-file before converting the file into a TXT-file. The TXT-file were then uploaded to AntConc. The searches which triggered less than 5,000 results did not have to undergo the randomisation in Excel and, thus, they were extracted from LancsBoxX and directly inserted into a DOCX-file which was then converted into a TXT-file, which was uploaded to AntConc.

As with the KorpusDK searches, I uploaded the BNC searches to AntConc to carry out a second round of searches. This was done because the LancsBoxX search engine – just as the KorpusDK search engine – cannot take punctuation into account and thus provided me with a

large number of irrelevant results, which are not examples of CoDPs but of some other syntactic constellation. The second round of searches was conducted in the same manner as described in section 3.3. Adding the alphanumeric wildcard \w either before or after the searches found in figure 9, I searched for *I and \w, \w and I, he and \w, \w and he, she and \w, \w and she,* etc., repeating for all the BNC searches. By doing so, I avoided having to manually discard all those examples that contain some sort of punctuation, and which are not relevant examples of CoDPs.

Having run my BNC results through AntConc – repeating for all 20 search strings – the results from AntConc were extracted, arranged into a table, and uploaded to Microsoft Excel. In Excel, all results were examined manually to sort out the remaining occurrences, which were not examples of CoDPs but another syntactic constellation. After having examined all the remaining results one by one, all results were analysed once again to determine the use of case in relation to the surrounding sentence and colour-coded accordingly, just as was done with both the Early Modern Danish, the Early Modern English, and the Present Day Danish investigations. For the full overview of the 20 tables constituting the Present Day English corpora investigation, see appendix 4. For comparison purposes, the number of relevant results for each of the 28 categories *(I and \w, \w and I, he and \w, \w and he, she and \w, \w and she*, etc.) were counted. The total number of relevant results can be found in the top bar of each table.

3.5 Calculation of percentages for frequency

3.5.1 Preparation of data and elimination of discrepancies

After having extracted all results from my four corpora, sorted them manually, and arranged them into columns, I was left with four major tables: a table for Early Modern Danish containing all results from the Holberg corpus, a table for Early Modern English containing all results from the Shakespeare corpus, a table for Present Day Danish containing all results from KorpusDK, and finally a table for Present Day English containing all results from the BNC (see appendices 1, 2, 3, and 4).

With all the raw data arranged into tables, I was almost ready to begin the comparisons across the different pronouns and the different corpora. However, one thing needed to be considered before the percentages could be calculated, namely all those searches which had triggered a high number of hits and had been narrowed down to 5,000 results. In KorpusDK, this reduction was executed automatically by the search engine because the software does not have the capacity for more. The BNC results, on the other hand, were reduced to 5,000 deliberately for two reasons: 1) so that the sample size would fit the KorpusDK sample size, and 2) because examining more than 5,000 results manually would be too time-consuming a job. In the corpora

from the Early Modern Period, this does constitute an issue since there are no searches that trigger more than 5,000 results.

The fact that I have reduced some search results from a greater number of hits to 5,000 hits while leaving other search results (those below 5,000 hits) as they were means there is a discrepancy in my data. A search such as the BNC search for *and us* triggered only 67 relevant results. Because this number is lower than the maximum limit of 5,000, no reductions have been made. Thus, we can conclude that there are exactly 67 relevant examples of CoDPs containing *and us* in the entire BNC. Considering now the BNC search for *and I*, which triggered 31,530 results. Due to the upper limit of 5,000, the results of this search have been narrowed down from 31,530 to 5,000, meaning that there is, in fact, 26,530 examples (31,530-5,000 = 26,530) of *and I* in the BNC that we are not able to consider. In other words, we have scaled down the actual number of occurrences of *and I* 6.3 times (31,530/5,000 = 6.306). After having manually examined all the 5,000 occurrences of *and I* (the scaled-down number of occurrences), I found that only 239 of these were relevant examples – that is, examples of CoDPs containing at least one pronoun (see appendix 4). The rest were examples of other syntactic constellations of no interest to the current project.

Out of 5,000 search results, 239 were relevant examples. However, this does not mean that there are only 239 relevant examples of *and I* in the entire BNC. To find the actual number of relevant instances of *and I*, we need to multiply this number by 6.3 since this was the number of times, we had narrowed down our original number of hits. Thus, to find the actual number of relevant instances, the following calculation has been carried out: 239*6.306 = 1,507 (number of relevant results out of 5,000*the reduction factor = the actual number of relevant examples). This means that in the entire BNC, there are, in fact, 1,507 instances of *and I,* not 239 as anticipated at first glance. For this calculation to prove correct, the data needs to be completely randomised, so that the sample of 5,000 results is representative of all 31,530 results. As described on pages 22 and 26, this is the case for both KorpusDK and the BNC.

The above calculation was repeated for all those searches that triggered more than 5,000 results and which had thus been narrowed down earlier in the process. In the four major tables for each corpus investigation (appendices 1, 2, 3, and 4), it is noted when this calculation has been made and, thus what the actual number of relevant results is (see the top bar of each column). After these calculations had been carried out, there were no longer any discrepancies in my data. Thus, the data was now ready to be converted into percentages making the numbers easier to compare to one another.

3.5.2 Percentages for frequency

Percentages were calculated for a total of 35 comparisons across the different numbers extracted from each of the four corpora. Comparisons were only made corpus-internal, meaning no percentages were calculated across the different corpora. Percentages were calculated for the following comparisons:

Comparisons made across the different numbers

Danish *(Used for Holberg Corpus and KorpusDK)*	**English** *(Used for Shakespeare Corpus and the BNC)*
Jeg versus *mig*	*I* versus *me*
Du versus *dig*	*Thou/~~you~~* versus *thee/~~you~~*
Han versus *ham*	*He* versus *him*
Hun versus *hende*	*She* versus *her*
Vi versus *os*	*We* versus *we*
I versus *jer*	*~~You versus you~~*
De versus *dem*	*They* versus *them*
First conjunct *jeg* versus *mig*	First conjunct *I* versus *me*
First conjunct *du* versus *dig*	First conjunct *thou/~~you~~* versus *thee/~~you~~*
First conjunct *han* versus *ham*	First conjunct *he* versus *him*
First conjunct *hun* versus *hende*	First conjunct *she* versus *her*
First conjunct *vi* versus *os*	First conjunct *we* versus *us*
First conjunct *i* versus *jer*	*~~First conjunct you versus you~~*
First conjunct *de* versus *dem*	First conjunct *they* versus *them*
Second conjunct *jeg* versus *mig*	Second conjunct *I* versus *me*
Second conjunct *du* versus *dig*	Second conjunct *thou/~~you~~* versus *thee/~~you~~*
Second conjunct *han* versus *ham*	Second conjunct *he* versus *him*
Second conjunct *hun* versus *hende*	Second conjunct *she* versus *her*
Second conjunct *vi* versus *os*	Second conjunct *we* versus *us*
Second conjunct *i* versus *jer*	*~~Second conjunct you versus you~~*
Second conjunct *de* versus *dem*	Second conjunct *they* versus *them*
First versus second conjunct *jeg*	First versus second conjunct *I*
First versus second conjunct *du*	First versus second conjunct *thou/~~you~~*
First versus second conjunct *han*	First versus second conjunct *he*
First versus second conjunct *hun*	First versus second conjunct *she*
First versus second conjunct *vi*	First versus second conjunct *we*
First versus second conjunct *i*	*~~First versus second conjunct you~~*
First versus second conjunct *de*	First versus second conjunct *they*
First versus second conjunct *mig*	First versus second conjunct *me*
First versus second conjunct *dig*	First versus second conjunct *thee/~~you~~*
First versus second conjunct *ham*	First versus second conjunct *him*
First versus second conjunct *hende*	First versus second conjunct *her*
First versus second conjunct *os*	First versus second conjunct *us*
First versus second conjunct *jer*	*~~First versus second conjunct you~~*
First versus second conjunct *dem*	First versus second conjunct *them*

Figure 10: Comparisons made across the results of the corpora investigations. Comparisons are only made corpus-internally meaning that comparisons are made between the different results in each corpora - not across the four corpora.

In the first section of comparisons, it was determined whether the nominative or the oblique form was used the most when considering all CoDPs found in the given corpora. In the second section of comparisons, it was determined whether the nominative or the oblique form was used the most in the first conjunct of all CoDPs found in the given corpora. Likewise, in the third section of comparisons, it was determined whether the nominative or the oblique form was used the most in the second conjunct of all CoDPs found in the given corpora. In the fourth section of comparisons, all CoDPs containing a nominative pronoun (*I/jeg, thou/du, he/han*, etc.) were examined to see how many of that pronoun (e.g., *jeg*) occur in the first conjunct and how many occur in the second conjunct. Likewise, in the fifth and last section of comparisons, all CoDPs containing an oblique pronoun (*me/mig, thee/dig, him/ham*, etc.) were examined to see how many of that pronoun (e.g., *mig*) occur in the first conjunct and how many occur in the second conjunct.

The five sections of comparisons given above were calculated following the formulas below. The investigations of the first person singular *I* and *me* in the BNC are given as examples.

The first section of comparisons:

$$\frac{\text{Number of occurrences of } I \text{ in CoDPs in BNC}}{(\text{Number of occurrences of } I \text{ in CoDPs in BNC} + \text{Number of occurrences of } me \text{ in CoDPs in BNC})}$$

$$\frac{\text{Number of occurrences of } me \text{ in CoDPs in BNC}}{(\text{Number of occurrences of } I \text{ in CoDPs in BNC} + \text{Number of occurrences of } me \text{ in CoDPs in BNC})}$$

The second section of comparisons:

$$\frac{\text{Number of occurrences of } I \text{ in the first conjunct of CoDPs in BNC}}{(\text{Number of occurrences of } I \text{ in the first conjunct of CoDPs in BNC} + \text{Number of occurrences of } me \text{ in the first conjunct of CoDPs in BNC})}$$

$$\frac{\text{Number of occurrences of } me \text{ in the first conjunct of CoDPs in BNC}}{(\text{Number of occurrences of } I \text{ in the first conjunct of CoDPs in BNC} + \text{Number of occurrences of } me \text{ in the first conjunct of CoDPs in BNC})}$$

The third section of comparisons:

$$\frac{\text{Number of occurrences of } I \text{ in the second conjunct of CoDPs in BNC}}{(\text{Number of occurrences of } I \text{ in the second conjunct of CoDPs in BNC} + \text{Number of occurrences of } me \text{ in the second conjunct of CoDPs in BNC})}$$

$$\frac{\text{Number of occurrences of } me \text{ in the second conjunct of CoDPs in BNC}}{(\text{Number of occurrences of } I \text{ in the second conjunct of CoDPs in BNC} + \text{Number of occurrences of } me \text{ in the second conjunct of CoDPs in BNC})}$$

The fourth section of comparisons:

$$\frac{\text{Number of occurrences of } I \text{ in the first conjunct of CoDPs in BNC}}{\text{Number of occurrences of } I \text{ in CoDPs in BNC}}$$

$$\frac{\text{Number of occurrences of } I \text{ in the second conjunct of CoDPs in BNC}}{\text{Number of occurrences of } I \text{ in CoDPs in BNC}}$$

The fifth sections of comparisons:

$$\frac{\text{Number of occurrences of } me \text{ in the first conjunct of CoDPs in BNC}}{\text{Number of occurrences of } me \text{ in CoDPs in BNC}}$$

$$\frac{\text{Number of occurrences of } me \text{ in the second conjunct of CoDPs in BNC}}{\text{Number of occurrences of } me \text{ in CoDPs in BNC}}$$

After having calculated all 35 comparisons (see figure 10) into percentages, repeating for each of the four corpora, the final percentages were arranged into tables to create the best possible overview. These tables can be found in chapter 4.

3.6 Calculation of percentages for expected and unexpected case

After having calculated the percentages for frequency, I moved on to calculating the percentages for expected and unexpected case. Taking my starting point in the colour-coding which had been carried out during the examination of the corpora search results, I calculated the percentages of the results, which, based on the standard case conventions of generative linguistics (see section 2.3), had the expected case according to its syntactic environment. I furthermore calculated the percentages of the results which had the unexpected case according to its syntactic environment as well as the percentages of the results where the case cannot be judged to be either expected or unexpected. Percentages were calculated for all searches (all pronouns as both the first and second conjunct) (see figure 5, figure 6, and figure 9) and were calculated following the below formula:

$$\frac{\text{Number of relevant hits with the expected case in e.g., the } I \textit{ and} \text{ search in BNC}}{\text{Total number of relevant hits in e.g., the } I \textit{ and} \text{ search in the BNC}}$$

By inserting the number of hits from the relevant search from the relevant corpus into the above formula, one gets the percentages for the results with unexpected case, the results with the expected case, and the results where the expected case cannot be determined. The results can be found in chapter 4 below.

Chapter 4: Results

Below, in figure 11 and figure 12, the frequency percentages from the comparisons in figure 10 are given (see section 3.5).

Frequency percentages - Early Modern Danish	**Frequency percentages - Early Modern English**
Jeg vs *mig*: 58.7% - 41.3%	*I* vs *me*: 55.3% - 44.7%
Du vs *dig*: 59.3 - 40.7 %	*You* (*thou*) vs *you* (*thee*): 46.7% - 53.3%
Han vs *ham*: 66.7% - 33.3%	*He* vs *him*: 41.7% - 58.3%
Hun vs *hende*: 20.0% - 80.0%	*She* vs *her*: 45.7% - 54.3%
Vi vs *os*: 0.0 % - 100.0%	*We* vs *us*: 28.0% - 72.0%
I vs *jer*: 44.4% - 55.6%	*You* vs *you*: -
De vs *dem*: 66.7% - 33.3%	*They* vs *them*: 25.0% - 75.0%
First conjunct *jeg* vs *mig*: 51.4% - 48.6%	First conjunct *I* vs *me*: 37.9% - 62.1%
First conjunct *du* vs *dig*: 58.3% - 41.7%	First conjunct *you* (*thou*) vs *you* (*thee*): 56.6% - 43.4%
First conjunct *han* vs *ham*: 61.5% - 38.5%	First conjunct *he* vs *him*: 37.5% - 62.5%
First conjunct *hun* vs *hende*: 20.0% - 80.0%	First conjunct *she* vs *her*: 52.2% - 47.8%
First conjunct *vi* vs *os*: 0.0% - 0.0%	First conjunct *we* vs *us*: 33.3% - 66.7%
First conjunct *i* vs *jer*: 44.4% - 55.6%	First conjunct *you* vs *you*: -
First conjunct *de* vs *dem*: 80.0% - 20.0%	First conjunct *they* vs *them*: 22.2% - 77.8%
Second conjunct *jeg* vs *mig*: 81.8 % - 18.2%	Second conjunct *I* vs *me*: 64.3% - 35.7%
Second conjunct *du* vs *dig*: 66.7% - 33.3%	Second conjunct *you* (*thou*) vs *you* (*thee*): 22.7% - 77.3%
Second conjunct *han* vs *ham*: 100.0% - 0.0%	Second conjunct *he* vs *him*: 50.0% - 50.0%
Second conjunct *hun* vs *hende*: 0.0% - 0.0%	Second conjunct *she* vs *her*: 33.3% - 66.7%
Second conjunct *vi* vs *os*: 0.0% - 100.0%	Second conjunct *we* vs *us*: 23.1% - 76.9%
Second conjunct *i* vs *jer*: 0.0% - 0.0%	Second conjunct *you* vs *you*: -
Second conjunct *de* vs *dem*: 0.0% - 100.0%	Second conjunct *they* vs *them*: 33.3% - 66.7%
First vs second conjunct *jeg*: 67.0% - 33.0%	First vs second conjunct *I*: 23.4% - 76.6%
First vs second conjunct *du*: 88.0% - 12.0%	First vs second conjunct *you* (*thou*): 85.7% - 14.3%
First vs second conjunct *han*: 80.0% - 20.0%	First vs second conjunct *he*: 60.0% - 40.0%
First vs second conjunct *hun*: 100.0% - 0.0%	First vs second conjunct *she*: 75.0% - 25.0%
First vs second conjunct *vi*: 0.0% - 0.0%	First vs second conjunct *we*: 57.1% - 42.9%
First vs second conjunct *i*: 100.0% - 0.0%	First vs second conjunct *you*: -
First vs second conjunct *de*: 100.0% - 0.0%	First vs second conjunct *they*: 66.7% - 33.3%
First vs second conjunct *mig*: 89.0% - 11.0%	First vs second conjunct *me*: 47.4% - 52.6%
First vs second conjunct *dig*: 91.0% - 9.0%	First vs second conjunct *you* (*thee*): 57.5% - 42.5%
First vs second conjunct *ham*: 100.0% - 0.0%	First vs second conjunct *him*: 71.4% - 28.6%
First vs second conjunct *hende*: 100.0% - 0.0%	First vs second conjunct *her*: 57.9% - 42.1%
First vs second conjunct *os*: 0.0% - 100.0%	First vs second conjunct *us*: 44.4% - 55.6%
First vs second conjunct *jer*: 100.0% - 0.0%	First vs second conjunct *you*: -
First vs second conjunct *dem*: 50.0% - 50.0%	First vs second conjunct *them*: 77.8% - 22.2%

Figure 11: Percentages calculated from the results of the searches in the Holberg Corpus and in the Shakespeare Corpus.

Frequency percentages - Present Day Danish	**Frequency percentages - Present Day English**
Jeg vs *mig*: 75,0% - 25,0%	*I* vs *me*: 70.9% - 29.1%
Du vs *dig*: 49,3% - 50,7%	*You* vs *you*: -
Han vs *ham*: 67,8% - 32,2%	*He* vs *him*: 69.3% - 30.7%
Hun vs *hende*: 64,2% - 35,8%	*She* vs *her*: 62.4% - 37.6%
Vi vs *os*: 42,3% - 57,7%	*We* vs *us*: 37.6% - 62.4%
I vs *jer*: 24,0% - 76,0%	*You* vs *you*: -
De vs *dem*: 32,3% - 66,7%	*They* vs *them*: 37.2% - 62.8%
First conjunct *jeg* vs *mig*: 36,1% - 63,9%	First conjunct *I* vs *me*: 31.6% - 68.4%
First conjunct *du* vs *dig*: 51,7% - 48,3%	First conjunct *you* vs *you*: -
First conjunct *han* vs *ham*: 67,7% - 32,3%	First conjunct *he* vs *him*: 70.4% - 29.6%
First conjunct *hun* vs *hende*: 64,0% - 36,0%	First conjunct *she* vs *her*: 63.7% - 36.3%
First conjunct *vi* vs *os*: 43,7% - 56,3%	First conjunct *we* vs *us*: 38.0% - 62.0%
First conjunct *i* vs *jer*: 21,1% - 78,9%	First conjunct *you* vs *you*: -
First conjunct *de* vs *dem*: 34,0% - 66,0%	First conjunct *they* vs *them*: 37.2% - 62.8%
Second conjunct *jeg* vs *mig*: 82,3% - 17,7%	Second conjunct *I* vs *me*: 84.3% - 15.7%
Second conjunct *du* vs *dig*: 23,7% - 76,3%	Second conjunct *you* vs *you*: -
Second conjunct *han* vs *ham*: 67,8% - 32,2%	Second conjunct *he* vs *him*: 44.6% - 55.4%
Second conjunct *hun* vs *hende*: 65,6% - 34,4%	Second conjunct *she* vs *her*: 46.1% - 53.9%
Second conjunct *vi* vs *os*: 40,6% - 59,4%	Second conjunct *we* vs *us*: 36.2% - 63.8%
Second conjunct *i* vs *jer*: 33,3% - 66,7%	Second conjunct *you* vs *you*: -
Second conjunct *de* vs *dem*: 29,2% - 70,8%	Second conjunct *they* vs *them*: 37.1% - 62.9%
First vs second conjunct *jeg*: 7,6% - 92,4%	First vs second conjunct *I*: 11.4% - 88.6%
First vs second conjunct *du*: 95,9% - 4,1%	First vs second conjunct *you*: -
First vs second conjunct *han*: 93,7% - 6,3%	First vs second conjunct *he*: 97.3% - 2.7%
First vs second conjunct *hun*: 87,7% - 12,3%	First vs second conjunct *she*: 94.6% - 5.4%
First vs second conjunct *vi*: 58,4% - 41,6%	First vs second conjunct *we*: 79.0% - 21.0%
First vs second conjunct *i*: 66,7% - 33,3%	First vs second conjunct *you*: -
First vs second conjunct *de*: 88,5% - 11,5%	First vs second conjunct *they*: 90.1% - 9.9%
First vs second conjunct *mig*: 40,5% - 59,5%	First vs second conjunct *me*: 59.8% - 40.2%
First vs second conjunct *dig*: 87,3% - 12,7%	First vs second conjunct *you*: -
First vs second conjunct *ham*: 93,7% - 6,3%	First vs second conjunct *him*: 92.5% - 7.5%
First vs second conjunct *hende*: 88,4% - 11,6%	First vs second conjunct *her*: 89.4% - 10.6%
First vs second conjunct *os*: 55,3% - 44,7%	First vs second conjunct *us*: 77.7% - 22.3%
First vs second conjunct *jer*: 78,9% - 21,1%	First vs second conjunct *you:* -
First vs second conjunct *dem*: 86,1% - 13,9%	First vs second conjunct *them*: 90.1% - 9.9%

Figure 12: Percentages calculated from the results of the searches in KorpusDK and in the BNC.

Some of the percentages in figure 11 and figure 12 are of more interest for the current project than others. Especially, the percentages calculated from the findings of the first person singular pronouns *I/jeg* and *me/mig* are of crucial interest since they, at times, behave interestingly and differently from the rest of the pronouns. The questions that these percentages raise will be discussed briefly later in this chapter before being accounted for thoroughly in chapter 5.

Below in figure 13 and figure 14, the percentages for expected and unexpected case are given (see section 3.6).

Percentages for expected and unexpected case - Early Modern Danish

	Unexpected case	Expected case undeterminable	Expected case		Unexpected case	Expected case undeterminable	Expected case
jeg og	2.8%	5.6%	91.6%	mig og	2.9%	0.0%	97.1%
og jeg	0.0%	0.0%	100.0%	og mig	0.0%	0.0%	100.0%
du og	7.1%	7.1%	85.7%	dig og	0.0%	0.0%	100.0%
og du	0.0%	0.0%	100.0%	og dig	0.0%	0.0%	100.0%
han og	12.5%	0.0%	87.5%	ham og	0.0%	0.0%	100.0%
og han	0.0%	0.0%	100.0%	og ham	----------	----------	----------
hun og	0.0%	0.0%	100.0%	hende og	0.0%	0.0%	100.0%
og hun	----------	----------	----------	og hende	----------	----------	----------
vi og	----------	----------	----------	os og	----------	----------	----------
og vi	----------	----------	----------	og os	0.0%	0.0%	100.0%
i og	0.0%	0.0%	100.0%	jer og	0.0%	0.0%	100.0%
og i	----------	----------	----------	og jer	----------	----------	----------
de og	0.0%	0.0%	100.0%	dem og	0.0%	0.0%	100.0%
og de	----------	----------	----------	og dem	0.0%	0.0%	100.0%

Percentages for expected and unexpected case - Early Modern English

	Unexpected case	Expected case undeterminable	Expected case		Unexpected case	Expected case undeterminable	Expected case
i and	0.0%	22.7%	77.3%	me and	0.0%	0.0%	100.0%
and i	1.4%	2.8%	95.8%	and me	2.5%	0.0%	97.5%
thou and	3.3%	6.7%	90.0%	thee and	0.0%	0.0%	100.0%
and thou	0.0%	0.0%	100.0%	and thee	5.8%	0.0%	94.2%
he and	0.0%	5.5%	94.5%	him and	0.0%	0.0%	100.0%
and he	8.3%	0.0%	91.7%	and him	0.0%	0.0%	100.0%
she and	8.3%	8.3%	83.3%	her and	0.0%	0.0%	100.0%
and she	25.0%	0.0%	75.0%	and her	0.0%	0.0%	100.0%
we and	0.0%	25.0%	75.0%	us and	0.0%	0.0%	100.0%
and we	0.0%	33.3%	66.7%	and us	0.0%	10.0%	90.0%
you and	----------	----------	----------	you and	----------	----------	----------
and you	----------	----------	----------	and you	----------	----------	----------
they and	0.0%	0.0%	100.0%	them and	0.0%	0.0%	100.0%
and they	0.0%	0.0%	100.0%	and them	50.0%	0.0%	50.0%

Figure 13: Early Modern Danish and Early Modern English percentages for expected and unexpected case.

Percentages for expected and unexpected case - Present Day Danish

	Unexpected case	Expected case undeterminable	Expected case		Unexpected case	Expected case undeterminable	Expected case
jeg og	1.5%	3.3%	95.2%	mig og	8.7%	4.1%	87.2%
og jeg	1.6%	5.3%	93.1%	og mig	8.6%	7.2%	84.2%
du og	2.1%	7.7%	90.2%	dig og	2.3%	4.8%	92.9%
og du	5.6%	5.6%	88.8%	og dig	13.8%	1.7%	84.5%
han og	0.2%	1.2%	98.6%	ham og	1.6%	3.1%	95.3%
og han	0.0%	4.9%	95.1%	og ham	15.2%	6.5%	78.3%
hun og	0.4%	2.3%	97.3%	hende og	3.2%	2.3%	94.5%
og hun	1.2%	6.2%	92.6%	og hende	7.7%	13.5%	78.8%
vi og	0.0%	0.7%	99.3%	os og	0.0%	5.5%	94.5%
og vi	4.8%	1.6%	93.6%	og os	2.0%	2.0%	96.0%
i og	0.0%	37.5%	62.5%	jer og	0.0%	0.0%	100.0%
og i	0.0%	0.0%	100.0%	og jer	0.0%	0.0%	100.0%
de og	0.9%	0.0%	99.1%	dem og	1.0%	1.0%	98.0%
og de	0.0%	11.1%	88.9%	og dem	29.4%	2.9%	67.7%

Percentages for expected and unexpected case - Present Day English

	Unexpected case	Expected case undeterminable	Expected case		Unexpected case	Expected case undeterminable	Expected case
i and	0.5%	2.6%	96.9%	me and	18.9%	2.6%	78.5%
and i	1.7%	2.9%	95.4%	and me	13.9%	10.7%	75.4%
you and	----------	----------	----------	you and	----------	----------	----------
and you	----------	----------	----------	and you	----------	----------	----------
he and	0.2%	0.6%	99.2%	him and	1.9%	0.3%	97.8%
and he	0.0%	0.0%	100.0%	and him	27.5%	2.0%	70.5%
she and	1.3%	0.3%	98.4%	her and	2.0%	0.0%	98.0%
and she	0.0%	12.5%	87.5%	and her	9.7%	0.0%	90.3%
we and	0.0%	2.5%	97.5%	us and	0.0%	0.0%	100.0%
and we	0.0%	0.0%	100.0%	and us	0.0%	5.4%	94.6%
you and	----------	----------	----------	you and	----------	----------	----------
and you	----------	----------	----------	and you	----------	----------	----------
they and	0.0%	1.4%	98.6%	them and	0.0%	0.0%	100.0%
and they	0.0%	0.0%	100.0%	and them	2.6%	0.0%	97.4%

Figure 14: Present Day Danish and Present Day English percentages for expected and unexpected case.

Although the percentages for expected and unexpected case do not show much discrepancy – most occurrences having the expected case compared to its surroundings – they do raise a few questions. The most apparent question being the high percentages for unexpected use of the first person singular oblique *me/mig*. The following sections will briefly discuss the questions that the percentages in figures 11, 12, 13, and 14 raise. The following sections do not examine the problems thoroughly or try to solve them. I will proceed to do so in chapter 5.

4.1 Increased popularity of the first person singular nominative compared to the first person singular oblique

We begin by considering figures 11 and 12, which gives the frequency percentages from the comparisons in figure 10 (see subsection 3.5.2). More precisely, we begin by considering the *I*

vs me/jeg vs mig calculations in the top row of the first section of comparisons which determine the ratio of the nominative form to the oblique form in the first person singular (figure 11 and figure 12, first section, first row). In other words, these calculations determine which one of the nominative and the oblique form of the first person singular pronoun is more common as part of CoDPs in the given corpora.

Here we would expect all four corpora, Holberg corpus, Shakespeare corpus, KorpusDK and the BNC, to behave similarly and thus have the same nominative to oblique ratio. The findings from the corpora from the Early Modern period (Holberg and Shakespeare corpora) support this initial expectation. The Early Modern Danish corpora have a 58.7% - 41.3% ratio, meaning that of all first person singular pronouns in CoDPs in the Holberg corpus, 58.7% were nominative, and 41.3% were oblique. Likewise, the Early Modern English corpora have a 55.3% - 44.7% ratio, meaning that of all first person singular pronouns in CoDPs in the Shakespeare corpus, 55.3% were nominative, and 44.7% were oblique. This parallel distribution is expected.

However, this expectation does not prove correct when considering the ratios of nominative to oblique in the two present day corpora. For Present Day Danish, the ratio is 75.0% - 25.0%, and for the Present Day English corpus, the ratio is 70.9% - 29.1%. Thus, the situation is that the two corpora from the present day period have almost the same distribution of case and the two corpora from the Early Modern period have almost the same distribution of case. Compared to each other, however, the two language periods have very different case distributions, the nominative form rising by 16 percentage points from ≈ 57.0% to ≈ 73.0%, and the oblique form dropping by 16 percentage points from ≈ 43.0% to ≈ 27.0% between the two time periods. Has something caused the nominative form to rise so drastically in popularity, or can this development be explained by other means? I will attempt to answer these questions in section 5.1.

4.2 Nominative pronoun forms and the sequence of the conjuncts

Considering now the fourth section of comparisons in figures 11 and 12 (the section beginning with *first vs second conjunct jeg/I*). This section of comparisons determines how many of a given nominative pronoun (e.g., *I/jeg*) occur in the first conjunct and how many occur in the second conjunct of the CoDPs they are part of.

Looking across the four corpora investigations, there seems to be a pattern with respect to these *first vs second conjunct nominative* investigations. In three of the corpora – the Early Modern English corpus, the Present Day Danish corpus, and the Present Day English corpus –

the first person singular pronoun, *I/jeg*, is far more common in the second conjunct than in the first conjunct (see the blue markings in figure 15). Oppositely, the remaining nominative pronouns (second and third person singular as well as first, second, and third person plural) all have the opposite distribution with a clear tendency to have the highest frequency in the first conjunct, not in the second (see the red markings in figure 15 below).

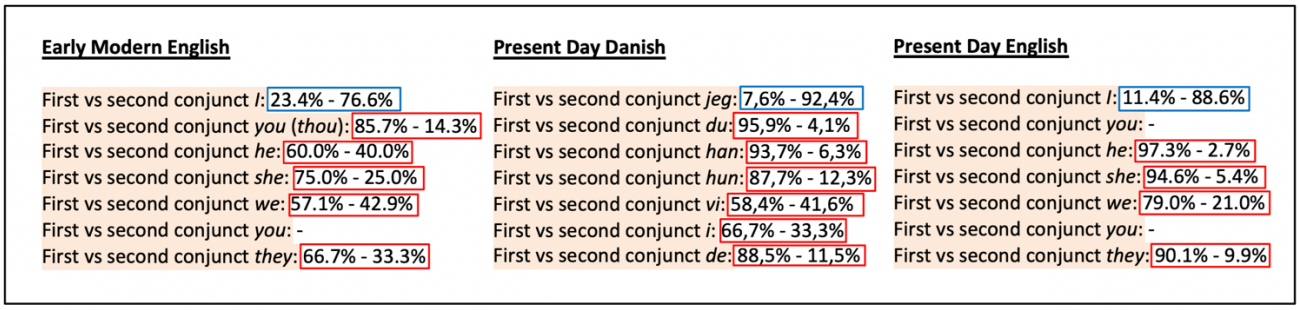


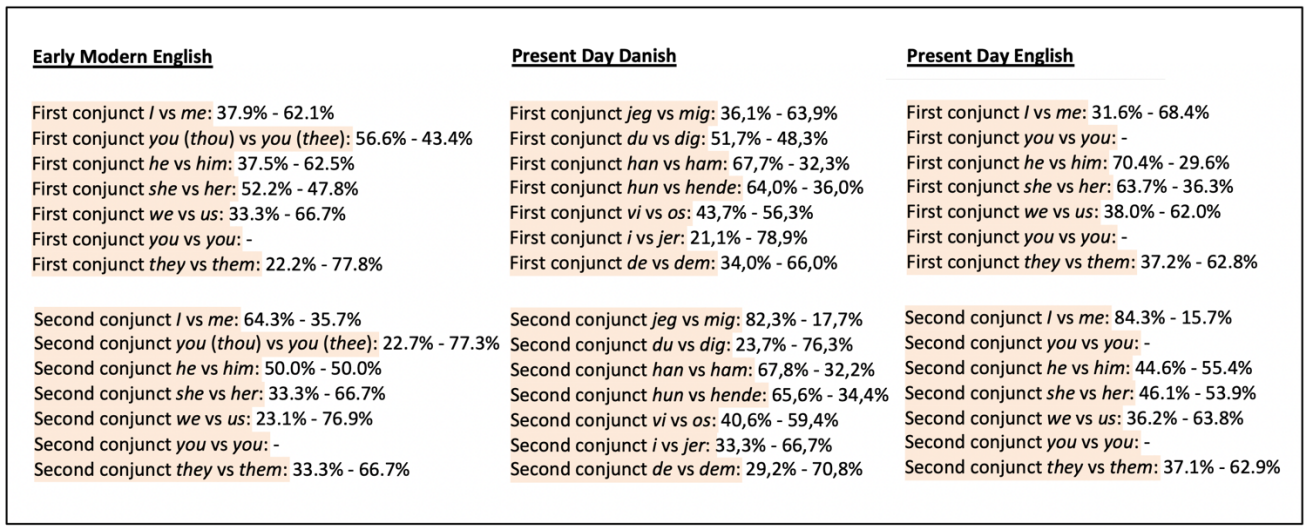
Figure 15: Extract from figure 11 and figure 12. The distribution of first person singular nominative pronouns in first and second conjunct in CoDPs.

Taking Present Day English as an example, this means that out of all CoDPs containing *I* in the BNC, only 11.4% of the cases have *I* appearing in the first conjunct, whereas 88.6% of the cases have *I* appearing in the second conjunct. The rest of the nominative pronouns have the opposite distribution being significantly more common in the first conjunct compared to the second conjunct – consider, for example, *he*, which has a ratio of 97.3% - 2.7%, meaning that only 2.7% of all occurrences of *he* in Present Day English appear in the second conjunct. This pattern recurs in Early Modern English and Present Day Danish: In Early Modern English, the distribution of the first person singular nominative pronoun is 23.4% to 76.6%, and in Present Day Danish the distribution is 7.6% to 92.4%. In both corpora investigations, the remaining nominative pronouns display the complete opposite picture being significantly more popular in the first conjunct than in the second.

As the only one of the four corpora investigations, Early Modern Danish does not conform to this pattern. In Early Modern Danish, the first person singular nominative distribution is 67.0% to 33.0%, meaning that out of all occurrences of *jeg*, 67.0% have *jeg* appearing in the first conjunct, whereas 33.0% have *jeg* appearing in the second conjunct. Thus, the Early Modern Danish distribution of the first person singular nominative pronoun is diametrically opposed to the distribution of the same pronoun in the other three corpora investigations. Possibly, this difference can be explained by the fact that we have so few occurrences of CoDPs in the Early Modern Danish corpus. In fact, there are only 54 occurrences of the first person singular nominative *jeg* as part of CoDPs, making our percentages quite vague and unreliable (see appendix

1). Thus, the Early Modern Danish Holberg corpus will be left out of any further considerations of this specific question.

The tendency for the first person singular nominative form to be most popular in the second conjunct and the rest of the nominative pronoun forms to be most popular in the first conjunct is backed up by the percentages from the second and third sections of comparisons (see figure 16, which is an extract from figure 11 and figure 12). The second section of comparisons beginning with *first conjunct I/jeg vs me/mig* determines whether it is the nominative or the oblique form of the first person singular pronoun that is most popular in the first conjunct of CoDPs. As expected, based on the findings above, the oblique form *me/mig* is significantly more common in the first conjunct than the nominative form *I/jeg* in all three corpora investigations (see figure 16). In the same way, the third section of comparisons, which determines which of the nominative or the oblique form is more prevalent in the second conjunct, demonstrates that the nominative form *I/jeg* is significantly more common in the second conjunct compared to the oblique form *me/mig*. This pattern, once again, recurs in each of the three corpora investigations (see figure 16).



Early Modern English	Present Day Danish	Present Day English
First conjunct *I* vs *me*: 37.9% - 62.1%	First conjunct *jeg* vs *mig*: 36,1% - 63,9%	First conjunct *I* vs *me*: 31.6% - 68.4%
First conjunct *you* (*thou*) vs *you* (*thee*): 56.6% - 43.4%	First conjunct *du* vs *dig*: 51,7% - 48,3%	First conjunct *you* vs *you*: -
First conjunct *he* vs *him*: 37.5% - 62.5%	First conjunct *han* vs *ham*: 67,7% - 32,3%	First conjunct *he* vs *him*: 70.4% - 29.6%
First conjunct *she* vs *her*: 52.2% - 47.8%	First conjunct *hun* vs *hende*: 64,0% - 36,0%	First conjunct *she* vs *her*: 63.7% - 36.3%
First conjunct *we* vs *us*: 33.3% - 66.7%	First conjunct *vi* vs *os*: 43,7% - 56,3%	First conjunct *we* vs *us*: 38.0% - 62.0%
First conjunct *you* vs *you*: -	First conjunct *i* vs *jer*: 21,1% - 78,9%	First conjunct *you* vs *you*: -
First conjunct *they* vs *them*: 22.2% - 77.8%	First conjunct *de* vs *dem*: 34,0% - 66,0%	First conjunct *they* vs *them*: 37.2% - 62.8%
Second conjunct *I* vs *me*: 64.3% - 35.7%	Second conjunct *jeg* vs *mig*: 82,3% - 17,7%	Second conjunct *I* vs *me*: 84.3% - 15.7%
Second conjunct *you* (*thou*) vs *you* (*thee*): 22.7% - 77.3%	Second conjunct *du* vs *dig*: 23,7% - 76,3%	Second conjunct *you* vs *you*: -
Second conjunct *he* vs *him*: 50.0% - 50.0%	Second conjunct *han* vs *ham*: 67,8% - 32,2%	Second conjunct *he* vs *him*: 44.6% - 55.4%
Second conjunct *she* vs *her*: 33.3% - 66.7%	Second conjunct *hun* vs *hende*: 65,6% - 34,4%	Second conjunct *she* vs *her*: 46.1% - 53.9%
Second conjunct *we* vs *us*: 23.1% - 76.9%	Second conjunct *vi* vs *os*: 40,6% - 59,4%	Second conjunct *we* vs *us*: 36.2% - 63.8%
Second conjunct *you* vs *you*: -	Second conjunct *i* vs *jer*: 33,3% - 66,7%	Second conjunct *you* vs *you*: -
Second conjunct *they* vs *them*: 33.3% - 66.7%	Second conjunct *de* vs *dem*: 29,2% - 70,8%	Second conjunct *they* vs *them*: 37.1% - 62.9%

Figure 16: Extract from figure 11 and figure 12. The distribution of nominative and oblique form in first and second conjunct of all CoDPs.

The tendency for the first person singular nominative to be most popular in the second conjunct and thus stand out from the rest will be examined further in section 5.3. Here, I will also attempt to explain what caused this divergence between the first person singular nominative and the rest of the nominative pronouns and determine what this divergence tells us about CoDPs and the sequence within them.

4.3 Oblique pronoun form in subject position

Let us now proceed to consider figure 13 and figure 14, which display the percentages for expected and unexpected case in all the corpora searches of all four corpora. All the results from each search (see appendices 1, 2, 3, and 4 for the full picture of all results) are divided into three categories: expected case, unexpected case, and instances where the expected case cannot be determined. Hereafter, the percentages for each of these categories were calculated to see how many, out of all results, had the expected case, the unexpected case, etc. (this procedure is described in-depth in section 3.6).

A few searches stand out from the rest by having a significantly high percentage of results with unexpected case. These searches are the Present Day English and Present Day Danish CoDPs which contain the first person singular oblique pronoun *me/mig* (see figure 14). For reasons yet unclear, the searches *mig og*, *og mig*, *me and*, *and me* trigger a large number of results which have the unexpected case according to the standard case theory in generative linguistics. In other words: A large number of the *me/mig* search results have the CoDPs containing *me/mig* appearing in subject contexts where the oblique form is highly unexpected. The percentages are as follows: 8.7% of all Present Day Danish CoDPs in which *mig* is the first conjunct *(mig og*) have *mig* appearing in unexpected contexts. 8.6% of all Present Day Danish CoDPs in which *mig* is the second conjunct (*og mig*) have *mig* appearing in unexpected contexts. Likewise, 18.9% of all Present Day English CoDPs in which *me* is the first conjunct *(me and*) have *me* appearing in unexpected contexts. 13.9% of all Present Day English CoDPs in which *me* is the second conjunct (*and me*) have *me* appearing in unexpected contexts. The fact that ≈ 10-20% of all occurrences of *me/mig* in CoDPs have the unexpected case is quite astonishing as it indicates that Present Day English and Present Day Danish speakers do not adhere to the standard case system described in section 2.3.

Considering now the percentages for the rest of the oblique search results (second and third person singular as well as first, second, and third person plural). Here, we see that a similar tendency is prevailing: The Present Day Danish searches *dig og, og dig, ham og, og ham, hende og, og hende, og os, og dem* and the Present Day English searches *him and, and him, her and, and her, and them* all have rather high percentages of results with unexpected case. Not all the above-mentioned searches have as high percentages as the first person singular oblique, which has ≈ 10-20% results with unexpected case. Despite this, it is clear that the oblique search results behave differently from the nominative search results, which, most of them, have percentages between 0-5% (see figure 14).

In Early Modern Danish and Early Modern English, this tendency cannot be observed in the same way as in the present day variants indicating that the phenomenon of using the oblique case in subject position has developed after the end of the Early Modern period. This and other questions regarding oblique pronouns used in subject position will be discussed in subsection 5.2.1.

4.4 Nominative pronoun form in complement position

Let us now consider the percentages for expected and unexpected case for the nominative searches (see the left column of figures 13 and 14). Compared to the oblique searches, which have relatively high percentages for unexpected case (see the above section), the present day nominative searches have very low percentages for unexpected case. In fact, for most of the nominative searches, both Present Day Danish and Present Day English, it is only 0-5% of the results that have the unexpected case. Thus, none of the nominative searches stand out in the same way as the oblique searches. Despite this, it is worthwhile to look at the searches containing the first person singular nominative pronoun *I/jeg* – that is *I and, and I, jeg og, og jeg*. Except for the *I and* search, these searches all have percentages ranging from 1.5 - 1.7%, which is worth noting when compared to most of the other nominative percentage rates, which are very close to zero. In subsection 5.2.2, I will return to why the percentages for unexpected case for the first person singular nominative are of interest for the current project.

4.5 Odd distribution of first person singular nominative and oblique in Early Modern Danish

One final observation worth mentioning when considering the calculations of percentages for frequency and expected and unexpected case is the seemingly random distribution of case forms in Early Modern Danish, which is evident in both figure 11 and figure 13. The Early Modern Danish column in figure 11, which contains the percentages calculated from the comparisons in figure 10, depicts a rather random and vague picture of the use of the pronoun case forms. First of all, many of the comparisons have a 100.0% - 0.0% or 0.0% - 100.0% distribution. In most instances, this simply means that there have been a deficient number of results in the corpora searches involved in those comparisons. Second of all, there are seven comparisons which I was unable to carry out because there were no results at all in the corpora searches involved in those comparisons (e.g., there were no occurrences of *vi* as part of CoDPs in the entire Holberg corpus meaning that the *first vs second conjunct vi* comparison is unable to be carried out). Such comparisons are marked with 0.0% - 0.0% in figure 11.

Essentially, the lack of occurrences of CoDPs containing pronouns in the Early Modern Danish Holberg corpus leaves us with highly vague percentages which are not suitable to draw any conclusions upon. The lack of data arguably also explains why the Early Modern Danish percentages do not conform to the percentages from the other three corpora investigations. As was found in section 4.2, all corpora except for the Early Modern Danish corpora conformed to the same pattern – namely that the first person singular nominative *I/jeg* was significantly more popular in the second conjunct, whereas the other first person singular nominative pronouns had the opposite distribution being more popular in the first conjunct. The percentages for the Early Modern Danish investigations do not support this tendency – in fact, they show the exact opposite picture, with *jeg* being more common in the first conjunct, not the second. This is not the only place where the Early Modern Danish percentages seem to disturb the picture. In the second section of comparisons in figure 11, the other three corpora have a distribution of ≈ 35.0% - 65.0%, meaning that out of all first person singular pronouns occurring in the first conjunct of a DP, 35% were the nominative form *I/jeg* whereas 65% were the oblique form *me/mig.* Conversely, Early Modern Danish has a 51.4% - 48.6% distribution.

Early Modern Danish also seem to differ from the other corpora investigations with respect to figure 13, which displays the percentages for expected and unexpected case. The two searches *du og* and *han og* stands out from the rest of the searches having respectively 7.1% and 12.5% results with unexpected case. None of the other three corpora investigations has similarly high numbers for the *thou and/du og* and *he and/han og* searches indicating that something strange is happening in the Early Modern Danish corpora investigation.

The reason for all the strange numbers and percentages for Early Modern Danish compared to the three other corpora is, unfortunately, simply lack of data. Because there are so few occurrences of CoDPs containing pronouns in the Holberg corpus (see appendix 1 for Early Modern Danish compared to appendix 3 for Present Day Danish), we are left with a highly vague data set upon which we should be hesitant to draw any real conclusions. Because of this lack of data, we must remain critical of all conclusions which involve data from the Holberg corpus.

Chapter 5: Discussion

5.1 Corpus composition

In section 4.1, we saw that the percentages which determine the ratio of the nominative form to the oblique form in the first person singular (*I vs me/jeg vs mig* in figures 11 and 12) were not consistent in all four corpora as would be expected. More precisely, the two corpora from the Early Modern period (Holberg corpus and Shakespeare corpus) showed a parallel distribution (≈ 57.0% nominative - 43.0% oblique), while the present day corpora (KorpusDK and the BNC) showed a different but also parallel distribution (≈ 73.0% nominative - 27.0% oblique). For convenience, the numbers are repeated below in figure 17:

Early Modern Danish	Early Modern English	Present Day Danish	Present Day English
Jeg vs *mig*: 58.7% - 41.3%	*I* vs *me*: 55.3% - 44.7%	*Jeg* vs *mig*: 75,0% - 25,0%	*I* vs *me*: 70.9% - 29.1%
Du vs *dig*: 59.3 - 40.7 %	*You* (*thou*) vs *you* (*thee*): 46.7% - 53.3%	*Du* vs *dig*: 49,3% - 50,7%	*You* vs *you*: -
Han vs *ham*: 66.7% - 33.3%	*He* vs *him*: 41.7% - 58.3%	*Han* vs *ham*: 67,8% - 32,2%	*He* vs *him*: 69.3% - 30.7%
Hun vs *hende*: 20.0% - 80.0%	*She* vs *her*: 45.7% - 54.3%	*Hun* vs *hende*: 64,2% - 35,8%	*She* vs *her*: 62.4% - 37.6%
Vi vs *os*: 0.0 % - 100.0%	*We* vs *us*: 28.0% - 72.0%	*Vi* vs *os*: 42,3% - 57,7%	*We* vs *us*: 37.6% - 62.4%
I vs *jer*: 44.4% - 55.6%	*You* vs *you*: -	*I* vs *jer*: 24,0% - 76,0%	*You* vs *you*: -
De vs *dem*: 66.7% - 33.3%	*They* vs *them*: 25.0% - 75.0%	*De* vs *dem*: 32,3% - 66,7%	*They* vs *them*: 37.2% - 62.8%

Figure 17: Extract from figure 11 and figure 12. The ratio of nominative to oblique in CoDPs.

Because the two time periods show such different frequencies with respect to the two case forms, it seems reasonable to believe that something has happened between the two periods that have made the nominative form *I/jeg* increase in popularity compared to the oblique form *me/mig*.

Although it seems as if the first person singular nominative has risen drastically in popularity between the two time periods, we should be careful with jumping to such conclusions. As was thoroughly described in subsection 3.5.2, the above numbers were found by comparing the total number of occurrences of the first person singular nominative to the total number of occurrences of the first person singular oblique, a procedure repeated for all four corpora. Although this type of comparison does tell us something about the ratio of nominative to oblique in the given corpus, it does not arm us with enough information to draw any conclusions about what caused the apparent increase in popularity of the nominative form. To fully understand the increase in nominative occurrences, we need to look at the numbers from another angle.

The first step is dividing the number of first person singular nominative occurrences (*I/jeg*) in the given corpus – the absolute frequency – by the total number of words in that corpus. By doing so, we get the relative frequency of the first person singular nominative form. This procedure is repeated for both case forms, nominative and oblique, alone (*I/jeg*) and in

coordination (*I and* + *and I*/*jeg og* + *og jeg*) in all four corpora giving the percentages found in figures 18 and 19[2] below.

	Holberg Corpus 708,558 words	Shakespeare Corpus 958,268 words	KorpusDK 56,000,000 words	BNC 100,000,000 words
'jeg'	17,241 = 0.24%		405,298 = 0.07%	
'I'		23,923 = 0.24%		858,211 = 0.09%
'mig'	5667 = 0.08%		90,272 = 0.02%	
'me'		8,287 = 0.09%		127,872 = 0.01%

Figure 18: Percentages for how big a part of the given corpora the first person singular nominative and the first person singular oblique constitute.

	Holberg Corpus 708,558 words	Shakespeare Corpus 958,268 words	KorpusDK 56,000,000 words	BNC 100,000,000 words
'jeg og' + 'og jeg'	54 = 0.0008%		3,575 = 0.0006%	
'I and' + 'and I'		94 = 0.001%		1,700 = 0.0002%
'mig og' + 'og mig'	38 = 0.0005%		1,194 = 0.0002%	
'me and' + 'and me'		76 = 0.0008%		697 = 0.00007%

Figure 19: Percentages for how big a part of the given corpora the first person singular nominative and the first person singular oblique constitute when part of a coordination.

From the numbers in figure 18 and figure 19, the following ratios of nominative to oblique can be calculated:

Early Modern Danish Holberg Corpus	**Early Modern English** Shakespeare Corpus	**Present Day Danish** KorpusDK	**Present Day English** BNC
First person singular pronouns alone			
Jeg vs *mig*: 75% - 25%	*I* vs *me*: 74% - 26%	*Jeg* vs *mig*: 82% - 18%	*I* vs *me*: 87% - 13%
First person singular pronouns in coordination			
Jeg vs *mig*: 59% - 41%	*I* vs *me*: 55% - 45%	*Jeg* vs *mig*: 75% - 25%	*I* vs *me*: 71% - 29%

Figure 20: The ratio of nominative to oblique in first person singular pronouns alone and in coordination.

[2] The numbers in figure 19 are calculated the same way as in figure 18, even though we are dealing with coordination structures containing two words. Hence, the relative frequency is calculated by dividing the absolute frequency by the total number of words in the corpus. It is possible that, because we are dealing with coordination structures, this is a faulty method of calculating the relative frequency. When dealing with CoDPs, each instance consists of two words (e.g., *I and*/*jeg og*), not one as with the simple pronouns. Arguably, this problem could be solved by dividing the absolute frequency by half the number of words in the given corpora, although this method has not been thoroughly investigated. The above reflections do not constitute a significant issue for any of the following conclusions since the proportions will be the same no matter the method.

The percentages for the first person singular pronouns in coordination found in the bottom part of figure 20 are identical to the original numbers found in figure 17. However, by comparing these numbers to the same numbers but for the first person singular pronouns alone (where the pronoun is not part of CoDPs), we gain essential knowledge about the development in case distribution.

As inferred from figure 20, the first person singular nominative alone (where it is not part of CoDPs) increases by 7 percentage points between Early Modern Danish and Present Day Danish. Likewise, between Early Modern English and Present Day English, the first person singular nominative alone (where it is not part of CoDPs) increases by 13 percentage points. Let us now compare these numbers to the numbers for the same pronouns but in coordination: Between Early Modern Danish and Present Day Danish, the first person singular nominative in coordination increases by 16 percentage points. Likewise, between Early Modern English and Present Day English, the first person singular nominative in coordination increases by 16 percentage points.

These four increases in the nominative form are relatively similar. The first person singular nominative in coordination has risen slightly more than the first person singular nominative, which are not part of a coordination, but essentially the distributions of nominative to oblique in all four corpora, both alone and in coordination, are parallel. On the basis of this parallelism, we can rule out the coordination structure as the reason for the increase of the nominative form and the decrease of the oblique form. It could have been that something had happened with respect to the coordination structure, which made the nominative form more popular, but from the numbers found above, this does not seem to be the case. Hence, the findings of figures 18, 19, and 20 have enabled us to eliminate the coordination structure as the reason for the increase in the nominative between the Early Modern period and present day. That said, we are still left more or less clueless as to why the data shows this particular increase.

It is very unlikely that the increase of the first person singular nominative has something to do with the speakers; the speakers of English and Danish most likely use the first person singular nominative and oblique just as much in 2022 as they did in the 17th and 18th centuries. This neither constitutes us with an explanation for the increase of the nominative form.

One thing that could, however, account for the increase is the corpora compositions. The four corpora are composed very differently – or more precisely, the two Early Modern corpora have a significantly different composition than the present day corpora. As mentioned in chapter 3, the Holberg corpus consists of the Danish playwright Ludvig Holberg's complete works; that is plays which imitate everyday dialogue, prose, letters, and the so-called epistles which

are also a type of letters (*Holberg Corpus* 2022). Likewise, the Shakespeare corpus consists of the British playwright William Shakespeare's complete works; that is plays which imitates dialogue and a number of sonnets[3].

The two present day corpora, KorpusDK and the BNC, differ significantly from the Holberg and Shakespeare corpora with respect to composition. In KorpusDK, "[t]he texts have been composed in order to give a broad and varied picture of the wealth of appearances which language can take (...) Contrary to the examples which you can find on the internet by means of a search engine, the texts have been carefully selected to show a broad section of both text types and language users" (Det Danske Sprog- og Litteraturselskab, n.d.). Hence, KorpusDK consists of texts from national news media, magazines, publishing companies, schools, associations, companies, websites, and private individuals (Det Danske Sprog- og Litteraturselskab, n.d.). Likewise, the BNC is a "word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century, both spoken and written" (Burnard 2015). The written part of the BNC makes up 90% of the corpus and includes "extracts from regional and national newspapers, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, school and university essays, among many other kinds of text" (Burnard 2015). The spoken part makes up 10% of the corpus and "consists of orthographic transcriptions of unscripted informal conversations (recorded by volunteers selected from different age, region and social classes in a demographically balanced way) and spoken language collected in different contexts, ranging from formal business or government meetings to radio shows and phone-ins" (Burnard 2015).

Returning to figures 18 and 19, we see the difference in the corpora compositions clearly. As seen from the percentages, both the nominative form and the oblique form (alone and in coordination) constitute a larger part of the Early Modern corpora than the present day corpora. For example, the first person singular nominative *jeg* makes up 0.07% of KorpusDK. In comparison, *jeg* makes up 0.24% of the Holberg corpus. The same goes for the English corpora, where *I* makes up 0.09% of the BNC, but 0.24% of the Shakespeare corpus. This pattern is consistent for all first person singular forms (see figures 18 and 19).

The variation in the relative frequencies of the first person singular form can be ascribed to differences in the corpora compositions. As was described above, the present day corpora

[3] The sonnets only constitute a very small fraction of the Shakespeare corpus, which is fortunate since they are written in the rhyme scheme, iambic pentameter, and we thus cannot rely fully on their syntactic composition to be representative of the language situation in the first half of the 18th century.

are designed to represent a wide cross-section of English and Danish. Contrary to the present day corpora, the two playwright corpora primarily consist of plays that imitate dialogue and letters – two very narrow genres, that are not fully representative of language at the time. For letters, first person narration is a characteristic; hence it is natural that *I/jeg* occurs frequently. In plays, the first person singular nominative is also standard – especially in dialogues and monologues, which comprise a significant part of both Shakespeare's and Holberg's plays. Furthermore, both letters and plays often revolve around relationships between people and the conflicts arising from these relations. Hence, sentences where the nominative form *I/jeg* and the oblique form *me/mig* occur are standard in both Shakespeare's and Holberg's works.

The first person singular also occurs regularly in the present day corpora although it constitutes a significantly smaller part of the corpora than in the Early Modern corpora. The discrepancy can be ascribed to the broader and more varied corpora composition of KorpusDK and the BNC. KorpusDK and the BNC consist, among other sources, of news media, magazines, and websites – genres in which *I/jeg* and *me/mig* are used less often than in prose and plays. Hence it is natural that the first person singular forms occur less often in the present day corpora than the Early Modern corpora, which are composed only of genres with a high frequency of the aforementioned forms.

It is not unlikely that these distinct differences in corpora composition have given rise to what could be a statistical illusion – namely, the apparent increase in the nominative form *I/jeg* and the decrease of the oblique form *me/mig*. Because of the large discrepancies found in the composition of the corpora, we should be careful when comparing results across the two time periods. Hence, we cannot use the numbers for the ratio of nominative to oblique to draw any conclusions about the diachronic change that the English and Danish case systems have undergone.

The different corpora compositions constitute a possible source of error for the investigation of the behaviour of English and Danish pronouns and should be considered throughout the following sections.

5.2 Unexpected case

5.2.1 Unexpected oblique

As we saw in section 4.3, the searches *me and, and me, mig og, og mig* trigger a large number of results which have the unexpected case according to the standard case theory within generative linguistics (see section 2.3 for an account of this theory). In fact, as much as 18.9% of all Present Day English CoDPs in which *me* is the first conjunct (*me and*) have *me* appearing in

unexpected contexts. The other three search strings, *and me, mig og* and *og mig,* also trigger strangely high numbers of occurrences where the oblique is used unexpectedly – that is in subject contexts. In sentences (36) to (39), a few examples from the data sets (see appendices 3 and 4) are given of the oblique form *me/mig* appearing in subject contexts in Present Day English and Present Day Danish:

(36) Mig (obl.) og Jeppe var begyndt at komme sammen igen.
"Me (obl.) and Jeppe had started seeing each other again" (*KorpusDK*, n.d.).

(37) Julie og mig (obl.) ville jo have masser af tid.
"Julie and me (obl.) would have plenty of time" (*KorpusDK*, n.d.).

(38) Me (obl.) and my boyfriend play cards, and watch telly, and that sort of thing (*British National Corpus*, n.d.).

(39) My sister and me (obl.) are very close (*British National Corpus*, n.d.).

As was established in section 2.5, the unexpected use of case that we see in examples (36)-(39) is perfectly acceptable in both Present Day English and Present Day Danish CoDPs. Although not all Danish speakers would produce clauses such as (36) and (37), they would all certainly understand the meaning. The same applies to English speakers and the clauses in (38) and (39).

Even though CoDPs containing *me/mig* are regularly used in unexpected contexts in both English and Danish, clauses such as the above have often been subject of linguistic prescriptivism, being denounced as highly incorrect and even intolerable by some speakers (Straaijer 2020). Hence, in grammar books, constructions with oblique case in subject position are often considered to be "very informal colloquial language or children's language, which are not to be regarded as normal or as usable outside such circles"[4] (A. Hansen 1965, 102). Many speakers of both Present Day English and Present Day Danish follow this prescriptive rule, which directs that only *I/jeg* can be used for subjects, never *me/mig.* However, as found from the numbers in chapter 4 and examples (36)-(39), there are also speakers who disregard the prescriptive rule. In fact, 10% - 20% of all instances of the oblique form *me/mig* occur in unexpected contexts,

[4] Original text in Danish: "I sideordnet forbindelse med et substantiv er der tendens til oblik form af pronominet: Far og dig har vel været unge sammen (...) Dette er lavt talesprog eller barnesprog og altså ikke at regne for normalt eller blot anvendeligt uden for disse cirkler" (A. Hansen 1965, 102).

indicating that a relatively large part of Present Day English and Present Day Danish speakers use the oblique case in non-standard positions.

Considering these high numbers, it seems insufficient to deem this phenomenon as something that only occurs in very informal colloquial language or children's language. In line with this perception, Heegård, Jensen, and Schack (2019, 449) argue that the phenomenon of using oblique in subject contexts has "changed character from being primarily associated with low status to gradually being used by all language users in all social strata."[5] Together with the high numbers found in chapter 4, this could indicate that prescriptivism is in decline and that the phenomenon of using oblique case in subjects is increasing in acceptability, however, at this point, these are purely unbacked assumptions.

As has been established above, the use of oblique pronoun forms in subject position is a common phenomenon in both Present Day Danish and Present Day English CoDPs. This non-standard, yet common, use of case gives rise to an essential and important question about the assignment of case: How is it syntactically possible for oblique to be used in subject position? Within the Government and Binding framework, we assume oblique case to be assigned by a V° or a P° to the NP/DP which it governs, while nominative case is assigned by I° to the NP/DP which it governs. This process is demonstrated in figure 21 below:

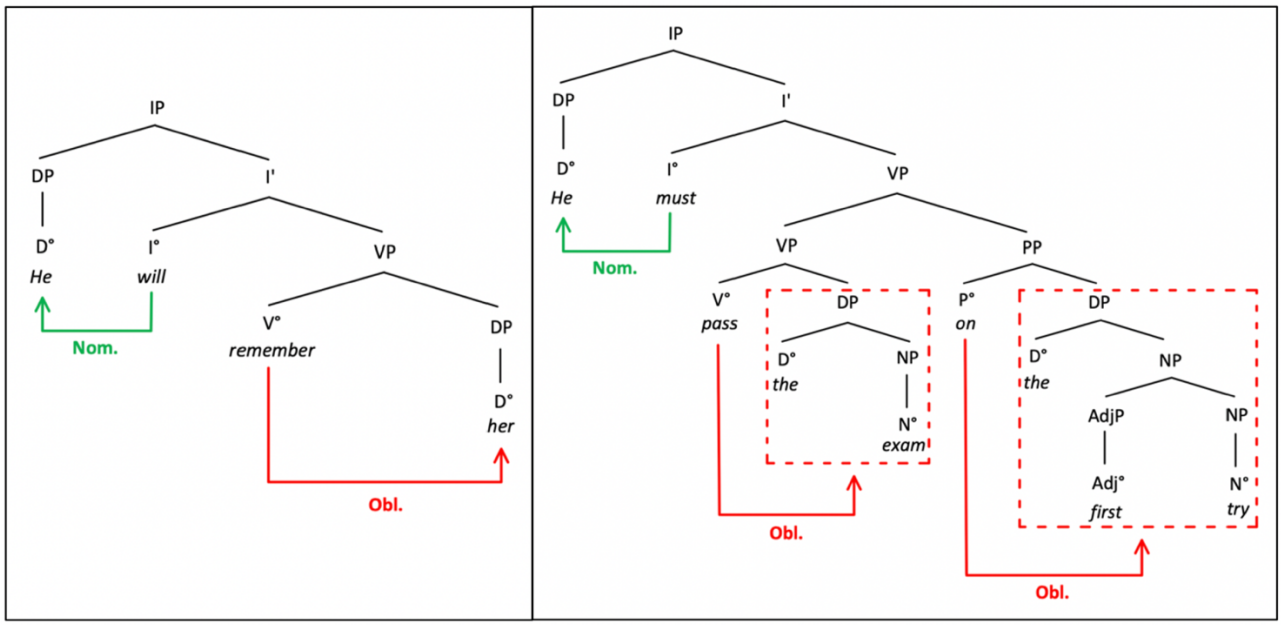


Figure 21: Assigment of nominative and oblique case according to the standard generative case conventions.

[5] Original text in Danish: "Med hensyn til udbredelse og sociolingvistisk valør har brugen undergået forandringer op gennem 1900-tallet, hvor den har skiftet karakter fra overvejende at være associeret med lavstatus til efterhånden at blive brugt af sprogbrugere i alle sociale lag" (Heegård, Jensen, and Schack 2019, 449).

Figure 21 demonstrates the standard, generative assignment of nominative and oblique case in Present Day English and Present Day Danish. However, as seen in (36)-(39), this assignment of case is regularly violated. See (37), which is repeated below as (40):

(40) Julie og mig (obl.) ville jo have masser af tid.
"Julie and me (obl.) would have plenty of time" (*KorpusDK*, n.d.).

Julie og mig is the subject of the clause and is positioned in IP-Spec. Hence, the DP which makes up *Julie og mig* should be assigned nominative case by *ville* (placed in I°). How come the second conjunct of the CoDPs then has oblique case and from where does it get it?

Several scholars have attempted to answer these questions resulting in a number of different theories linking case to anything from stress to information structure (see Jørgensen 2002; E. Hansen 1972; Heltoft 1990, 127-135; 1992; E. Hansen and Heltoft 2011, 29-30, 181; Jensen 2019, 72-74; Heegård, Jensen, and Schack 2019, 448-449). One theory which is placed within the generative framework (and the one which I intend to follow in this article) is Schütze's (2001) theory of default case. By introducing the notion of default case, Schütze (2001) provides a parsimonious and general[6] account for a number of unexpected phenomena which occur in the distribution of pronoun case. One of the phenomena which can be accounted for with reference to default case is the unexpected use of oblique in coordination structures (CoDPs) which function as subjects.

Essentially, the theory takes its starting point in the assumption that some DPs, "for a variety of reasons, have not had their case inflection determined by the syntax and are therefore sent to the morphological spell-out procedures without any indication of which case form they should be realized with" (Schütze 2001, 205). As required by the Case Filter (Chomsky 1981, 49) a DP without Case is ungrammatical. Hence, the pronouns not affected by any of the case mechanisms and thus not assigned case by the syntax, needs to obtain case from elsewhere for the sentence to be grammatical. Here, default case enters the picture. Schütze (2001, 206) defines default case as those case forms "that are used to spell out nominal expressions (e.g., DPs) that are not associated with any case feature assigned or otherwise determined by syntactic mechanisms". Default case is never assigned by anything to anything; rather, it is a "mechanism by which case marking can arise", which is supplied after spell-out in the instances where "there is no case assigner for the DP in question" (Schütze 2001, 210).

[6] Default case is assumed to be a feature of Universal Grammar (UG). Hence, "all languages have default case available" (Schütze 2001, 208).

It is important to mention that default case can never "save a DP from violating the Case Filter" (Schütze 2001, 206). Default case plays no role in licensing the DP in the first place "because structural licencing [referred to as abstract Case in this project] is a syntactic requirement" (Schütze 2001, 206). By recognising that licencing (abstract Case) and morphological case are separate systems, we circumscribe the distribution of default case so that default case is only supplied to those DPs that are already licensed (Schütze 2001, 208). In this way "default case can never save an otherwise invalid syntactic structure" (Schütze 2001, 208). Hence, no matter the default case of English, the Case Filter violations in (41)-(43) are still ungrammatical, meaning that default case cannot simply be inserted everywhere to save invalid sentences:

(41) *Him (obl.)/*He (nom.) to leave would be rude.

(42) *It is important him (obl.)/he (nom.) to be on time.

(43) *My desire him (obl.)/he (nom.) to succeed led me astray (Schütze 2001, 209).

Oblique (referred to as accusative in Schütze (2001)) is the default case form of both English and Danish (Schütze 2001, 210, 227). Hence, in both languages, the oblique forms are used whenever there is doubt about where case should come from and, thus, what the correct case should be. Sentences (44)-(48) are examples of English clauses where the pronoun DP is sent to spell-out without any indication of which case form it should be realised with and hence, is supplied with default case.

(44) My sister and me (obl.) are very close (*British National Corpus*, n.d.).

(45) Me (obl.), I'd rather try something different (*British National Corpus*, n.d).

(46) Who wants to try this game? Me (obl.)! (Schütze 2001, 211).

(47) We can't eat caviar and him (obl.) (eat) beans (Schütze 2001, 212).

(48) The real me (obl.) is finally emerging (Schütze 2001, 215).

The examples above show five different syntactic environments in which default case supplies a pronoun DP with oblique. Nothing licenses accusative case on *me* in any of the five clauses, indicating that oblique must be the morphological default form of English.

Considering the strong parallelisms between the English and Danish case systems, it comes as no surprise that the Danish default case is also oblique. Many of the syntactic environments in which default case occurs in English also have default case in Danish (see sentences (49)-(53) below).

(49) Julie og mig (obl.) ville jo have masser af tid.
"Julie and me (obl.) would have plenty of time" (*KorpusDK*, n.d.).

(50) Kom herhen! Mig (obl.)?
"Come here! Me (obl.)?" (E. Hansen 1972, 102).

(51) Ham (obl.), der står derovre, er min nabo.
"Him (obl.), who stands over there, is my neighbour" (Parrott 2009b, 288).

(52) Længe leve mig (obl.)!
"Long live me (obl.)!" (E. Hansen 1972, 102).

(53) Ham (obl.), han er en skat.
"Him (obl.), he is a darling" (Jensen 2019, 77).

What is of interest for the present project is the default-case environment which contains CoDPs consisting of at least one personal pronoun, that is, sentences (44) and (49). For reasons unknown, the DPs, which function as subjects in (44) and (49), are not assigned nominative case by I° as is expected according to the standard case conventions. Opposing these standard assumptions of case distribution, it is instead the oblique pronoun form which is used.

The theory of default case offers an explanation for this unexpected distribution of case. As mentioned above, the DPs in (44) and (49) (and in similar clauses) have no case assigner to assign them nominative case. Why I° does not assign nominative case into certain CoDPs remains a mystery. It could be that the pronoun positioned within the CoDPs is simply buried too deeply in the syntactic structure and thus that the connection to I° has been cut off in some

way[7]. It is beyond the scope of this paper to delve into this, but see Johannessen (1998, 109), Hartmann (2015, 489), E. Hansen and Heltoft (2011, 440), and Jensen (2019, 75), who all deal with the topic.

Instead of remaining caseless, which would make the clauses ungrammatical cf. the Case Filter, these subject CoDPs are assigned default oblique case after spell-out. The default oblique case "is never assigned by anything to anything" (Schütze 2001, 207). Hence, the subject CoDPs are not assigned oblique by anything; this would be impossibly for locality reasons[8], as demonstrated in figure 22 below:

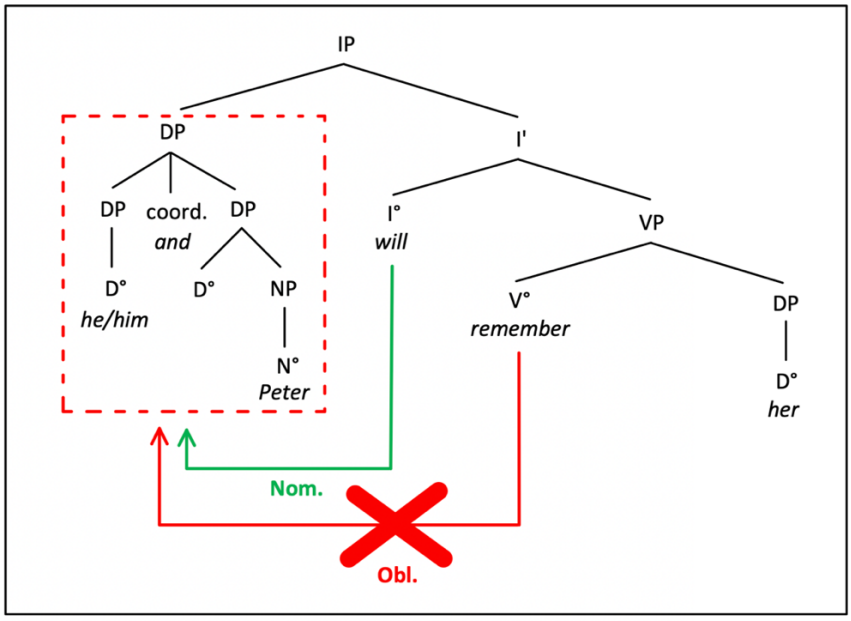
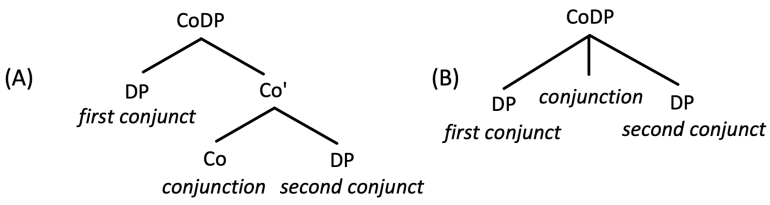


Figure 22: Impossible assignment of oblique case to the subject DP.

As seen above, default oblique case is assigned to a DP by neither I° nor V°. Rather, default case "is used to spell out a terminal node of the syntactic tree" (Schütze 2001, 207). In this way, "the presence of default case is, by design, invisible to the syntax" (Schütze 2001, 207) since the supplementation of default case happens after the morphological spell-out.

By assuming default case to be a language general feature of Universal Grammar (UG), we can account for why clauses such as (44) and (49) are grammatical and acceptable even

[7] This requires that we view coordination structures as asymmetrical as in (A) (Johannessen 1998, 109; Hartmann 2015, 489), not as symmetrical as in (B). For reasons of simplification, (B) will be used in all figures throughout this thesis. However, the present project still assumes the deep structure of CoDPs to look like (A).



[8] The closest potential case assigner for the subject will always be I° (which assigns nominative), never V°.

though their subjects (which consists of CoDPs) have oblique case – and we can do so without any ad hoc rules[9]. Hence, from a generative perspective, the default-case theory appears to be the most suitable theory for accounting for the unexpected use of oblique case in subjects.

The phenomenon of using the oblique form in subjects is not limited to the first person singular *me/mig*. As is clear from the results in chapter 4 (see figure 23 below), most of the oblique forms have a noticeably high percentage of unexpected occurrences:

Percentages for expected and unexpected case - Present Day Danish

	Unexpected case	Expected case undeterminable	Expected case		Unexpected case	Expected case undeterminable	Expected case
jeg og	1.5%	3.3%	95.2%	mig og	8.7%	4.1%	87.2%
og jeg	1.6%	5.3%	93.1%	og mig	8.6%	7.2%	84.2%
du og	2.1%	7.7%	90.2%	dig og	2.3%	4.8%	92.9%
og du	5.6%	5.6%	88.8%	og dig	13.8%	1.7%	84.5%
han og	0.2%	1.2%	98.6%	ham og	1.6%	3.1%	95.3%
og han	0.0%	4.9%	95.1%	og ham	15.2%	6.5%	78.3%
hun og	0.4%	2.3%	97.3%	hende og	3.2%	2.3%	94.5%
og hun	1.2%	6.2%	92.6%	og hende	7.7%	13.5%	78.8%
vi og	0.0%	0.7%	99.3%	os og	0.0%	5.5%	94.5%
og vi	4.8%	1.6%	93.6%	og os	2.0%	2.0%	96.0%
i og	0.0%	37.5%	62.5%	jer og	0.0%	0.0%	100.0%
og i	0.0%	0.0%	100.0%	og jer	0.0%	0.0%	100.0%
de og	0.9%	0.0%	99.1%	dem og	1.0%	1.0%	98.0%
og de	0.0%	11.1%	88.9%	og dem	29.4%	2.9%	67.7%

Percentages for expected and unexpected case - Present Day English

	Unexpected case	Expected case undeterminable	Expected case		Unexpected case	Expected case undeterminable	Expected case
i and	0.5%	2.6%	96.9%	me and	18.9%	2.6%	78.5%
and i	1.7%	2.9%	95.4%	and me	13.9%	10.7%	75.4%
you and	----------	----------	----------	you and	----------	----------	----------
and you	----------	----------	----------	and you	----------	----------	----------
he and	0.2%	0.6%	99.2%	him and	1.9%	0.3%	97.8%
and he	0.0%	0.0%	100.0%	and him	27.5%	2.0%	70.5%
she and	1.3%	0.3%	98.4%	her and	2.0%	0.0%	98.0%
and she	0.0%	12.5%	87.5%	and her	9.7%	0.0%	90.3%
we and	0.0%	2.5%	97.5%	us and	0.0%	0.0%	100.0%
and we	0.0%	0.0%	100.0%	and us	0.0%	5.4%	94.6%
you and	----------	----------	----------	you and	----------	----------	----------
and you	----------	----------	----------	and you	----------	----------	----------
they and	0.0%	1.4%	98.6%	them and	0.0%	0.0%	100.0%
and they	0.0%	0.0%	100.0%	and them	2.6%	0.0%	97.4%

Figure 23: Figure 14 repeated. Present Day Danish and Present Day English percentages for expected and unexpected case. The oblique forms are marked with red to show the disproportionally large amount of occurrences with unexpected case.

[9] Other analyses that account for the unexpected oblique have been proposed – among these, the suggestion that the subject DP is assigned oblique by a source within the CoDPs, e.g., that the conjunction is a case assigner (Schütze 2001, 214). Several other explanations exist as well, including a theory that assumes English to have no morphological case at all. See subsection 5.2.4 for this alternative explanation of unexpected oblique.

Although the numbers in figure 23 indicate that the phenomenon of unexpected oblique in subjects is most common with the first person singular *me/mig*, examples of unexpected oblique are also common in second and third person singular and first, second, and third person plural. The Present Day Danish searches *dig og, og dig, ham og, og ham, hende og, og hende, og os, og dem* and the Present Day English searches *him and, and him, her and, and her, and them* all have rather high percentages of unexpected case making sentences such as (54) and (55) occur regularly in KorpusDK and the BNC.

(54) Hende (obl.) og Thyra aftalte, at de ville komme tilbage.
"Her (obl.) and Thyra agreed that they would come back" (*KorpusDK*, n.d.).

(55) His sister and him (obl.) never spoke anything else but the Gaelic (*British National Corpus*, n.d.).

Schütze's theory can be generalised to cover all numbers, persons, and genders, meaning that the unexpected oblique in (54) and (55) can be accounted for by the default-case theory just as in the first person singular examples above. Essentially it is the same procedure that takes place: A clause with CoDPs in subject position is sent to spell-out without any indications of which case form the CoDPs should be realised with. After morphological spell-out the case-less CoDPs are supplied with the default case to satisfy the Case Filter and ensure that the clause is grammatical.

Apart from explaining why oblique pronouns – both singular and plural – can be used in subject environments, Schütze's theory on default case can account for the phenomenon of mixed case within CoDPs. As found in section 2.5, it is not only possible for English and Danish to have unexpected oblique in subject environments but also for CoDPs to have mixed case (Parrott 2009a, 166-169, 181-182):

(56) Him (obl.) and I (nom.) got on very well together (*British National Corpus*, n.d.).

(57) I dag er det en sød hemmelighed mellem hende (obl.) og jeg (nom.).
"Today it is a sweet secret between her (obl.) and I (nom.)" (*KorpusDK*, n.d.).

Parrott (2009b, 284) categorises languages in which unexpected oblique and mismatched case within CoDPs are salient to native speakers as vestigial-case languages. These languages are

languages such as English and Danish, which have lost most of their morphological case and in which distinctive case forms are limited to a subset of the personal pronouns. Interestingly, these languages all have oblique as their default case, whereas transparent-case languages such as German, Latin, etc. have nominative as their default case. In the transparent-case languages, unexpected oblique and case mismatches inside CoDPs "appear to be completely unattested and unacceptable" (Parrott 2009b, 277).

In his theory about default case, Schütze (2001) accounts for why vestigial-case languages such as English and Danish allow unexpected and mixed case marking in coordination structures and why transparent-case languages such as German and Latin do not. He assumes that "languages may have or lack case spreading (concord) within DP – that is, the case feature of the D head [D°] may or may not be copied or spread to other constituents of the DP" (Schütze 2001, 217). Transparent-case languages have this kind of spreading, whereas vestigial-case languages such as English and Danish do not. Essentially, this entails that in English and Danish, some pronouns that are buried deeply within a DP will not be assigned case by the case assigner. Hence, "the pronoun itself can lack a case feature even though it is contained within a DP that has a case feature" (Schütze 2001, 217). Cf. the Case Filter, the pronoun must receive case in order for the sentence to be grammatical, and thus the pronoun is supplied with default oblique case resulting in CoDPs with mismatched case as in (56) and (57).

Lack of case spreading can result in mismatched case, or it can result in unexpected oblique in subjects, as was described earlier in the present section. With unexpected oblique in subject position, it is the same procedure that takes place, but instead of case appearing on one of the conjuncts and not the other, case appears on neither of the two conjuncts. These conjuncts are then later, after spell-out, supplied with default case to satisfy the Case Filter. It is beyond the reach of this article to examine why in some instances case is assigned to both conjuncts of a coordination, in some instances only to one of the conjuncts, and in some instances to neither of the conjuncts.

The high number of occurrences of unexpected oblique in Present Day English and Present Day Danish indicates that it is not the standard case system alone that determines what morphological case is used where. Another important syntactic mechanism, namely default case, also plays a role in the distribution of case. It is the alternation between these two mechanisms that causes the unexpected and unpredictable use of oblique form in subject position in Present Day English and Present Day Danish.

5.2.2 Unexpected nominative

Although frequently used in both English and Danish, sentences in which an oblique pronoun is used as subject are often deemed as childlike and 'incorrect'. If a child asks *can Peter and me have an ice cream?/må Peter og mig få en is?*, they are corrected with the explanation that *me/mig* is wrong and that it should be *I/jeg* instead. Despite this correction being completely legitimate (if one follows the standard case patterns), "such corrections can lead to the belief that *og jeg* [and *and I*] is the only correct way of forming a coordination which contains a first person singular pronoun"[10] (Jensen, Kragh, and Strudsholm 2018, 83). If this happens, speakers will avoid using the oblique form *me/mig* altogether – replacing it with *I/jeg* – even in places that would, according to standard case assignment, require the oblique form (objects and complements of prepositions). In this way, sentences such as (58)-(61) have become common in everyday speech and writing in both English and Danish.

(58) Da vi kom, omfavnede han både min kone og jeg (nom.).
"When we arrived, he embraced both my wife and I (nom.)" (*KorpusDK*, n.d.).

(59) Hun skulle vise min familie og jeg (nom.) rundt på skolen.
"She would show my family and I (nom.) around at the school." (*KorpusDK*, n.d.).

(60) It has given my wife and I (nom.) a lot of comfort (*British National Corpus*, n.d.).

(61) It gave my walking partner and I (nom.) quite a scare (*British National Corpus*, n.d.).

"The use of a particular (form of a) word, phrase, etc. by a speaker or writer who is under the impression that this conforms to 'correct' (more prestigious) usage" is known as hypercorrection (Aarts 2014). Insisting on using *I/jeg* instead of *me/mig* even in complement position is one of the most common hypercorrections, and the phenomenon can be spotted regularly in both English and Danish – not only in everyday speech and writing but also in the worlds of politicians, magazine editors, in erudite prose and newspapers, and on television (Angermeyer

[10] Original text in Danish: "Disse rettelser resulterer i at *og jeg* af nogle kan opfattes som den eneste rigtige måde at danne paratagmer på hvori der indgår et pronomen i første person singularis" (Jensen, Kragh, and Strudsholm 2018, 83).

and Singler 2003, 171). Even among royals such as Queen Elizabeth II, this hypercorrect nominative can be found:

(62) It is a wonderful moment for my husband and I (nom.) after nearly six months away to be met and escorted by ships of the Home Fleet (Queen Elizabeth II, in 1954, quoted in Angermeyer and Singler 2003, 174).

Contrary to the unexpected oblique, unexpected use of nominative is restricted to one syntactic environment – that is coordination structures (CoDPs) in complement position. Additionally, unexpected nominative never occurs with the plural pronoun forms, only with the singular forms, indicating that the hypercorrection is limited to the latter (see (63) below).

(63) *It is a deal between we and they.
It is a deal between us and them (Schütze 2001, 220).

Unexpected nominative is restricted to the singular pronoun forms because speakers are taught "to think of *me* as a naughty word" (Bernstein 1965, 322). The denunciation of the oblique form occasionally leads to unexpected use of the other nominative singular pronoun forms (*du*[11], *he/han*, *she/hun*). However, it never affects the plural pronoun forms as seen from (63) above. Because the hypercorrection does not generalise to plurals, Schütze (2001, 220) does not consider the unexpected nominative to be part of grammar in the same way as unexpected oblique is thought to be a consequence of grammatical rules. Instead, Schütze follows Sobin (1997) in assuming the unexpected nominative to be a 'virus' that can easily override the default oblique case in those syntactic environments where case specifications have not been given prior to spell-out (Schütze 2001, 220). Hence, the use of unexpected nominative has to do with the absence of a syntactic case specification, just as the use of unexpected oblique. Those instances where case specifications have not been given prior to spell-out, and where default oblique case has been supplied, are more easily overridden than 'regular' oblique case assigned from either V° or a P°. Hence, prescriptive rules (what Sobin (1997) calls viruses) – in this case in the form of nominative – can easily replace the default case resulting in the nominative case form being used unexpectedly (Schütze 2001, 220).

[11] *You* is left out of any considerations of Present Day English case due to the syncretism of second person singular nominative and oblique and second person plural nominative and oblique.

Although the phenomenon of using nominative forms in complement position is considered one of the most common hypercorrections in both English and Danish, it is, very unexpectedly, not reflected in the results found in chapter 4. The investigation of expected and unexpected case (see figure 13) shows that approximately 1.5% of all Present Day Danish occurrences of the first person singular nominative *jeg* as part of a coordination occur in unexpected environments. In Present Day English, approximately 1.1% of all occurrences of *I* as part of a coordination are used unexpectedly. Compared to the numbers for unexpected use of the first person singular oblique (where there were between 10-20% unexpected occurrences), the numbers for unexpected use of *I/jeg* seem insignificant.

When we consider the numbers for all nominative pronoun forms (see figure 24 below), we do see the expected distribution of unexpected case – namely, that unexpected use of the nominative forms is seen exclusively with the singular forms[12]. The pattern, however, is vaguer than anticipated. We would have expected to see the numbers for the singular forms – and especially the first person singular nominative *I/jeg* – resemble the numbers for unexpected oblique. However, instead of having between 10% and 20% unexpected occurrences, *I/jeg* only has ≈ 1.5% and ≈ 1.1% respectively.

[12] The Present Day Danish *og vi* is unexpectedly high (4.8%) but is, in this case, merely considered a random outlier.

Percentages for expected and unexpected case - Present Day Danish

	Unexpected case	Expected case undeterminable	Expected case		Unexpected case	Expected case undeterminable	Expected case
jeg og	1.5%	3.3%	95.2%	mig og	8.7%	4.1%	87.2%
og jeg	1.6%	5.3%	93.1%	og mig	8.6%	7.2%	84.2%
du og	2.1%	7.7%	90.2%	dig og	2.3%	4.8%	92.9%
og du	5.6%	5.6%	88.8%	og dig	13.8%	1.7%	84.5%
han og	0.2%	1.2%	98.6%	ham og	1.6%	3.1%	95.3%
og han	0.0%	4.9%	95.1%	og ham	15.2%	6.5%	78.3%
hun og	0.4%	2.3%	97.3%	hende og	3.2%	2.3%	94.5%
og hun	1.2%	6.2%	92.6%	og hende	7.7%	13.5%	78.8%
vi og	0.0%	0.7%	99.3%	os og	0.0%	5.5%	94.5%
og vi	4.8%	1.6%	93.6%	og os	2.0%	2.0%	96.0%
i og	0.0%	37.5%	62.5%	jer og	0.0%	0.0%	100.0%
og i	0.0%	0.0%	100.0%	og jer	0.0%	0.0%	100.0%
de og	0.9%	0.0%	99.1%	dem og	1.0%	1.0%	98.0%
og de	0.0%	11.1%	88.9%	og dem	29.4%	2.9%	67.7%

Percentages for expected and unexpected case - Present Day English

	Unexpected case	Expected case undeterminable	Expected case		Unexpected case	Expected case undeterminable	Expected case
i and	0.5%	2.6%	96.9%	me and	18.9%	2.6%	78.5%
and i	1.7%	2.9%	95.4%	and me	13.9%	10.7%	75.4%
you and	----------	----------	----------	you and	----------	----------	----------
and you	----------	----------	----------	and you	----------	----------	----------
he and	0.2%	0.6%	99.2%	him and	1.9%	0.3%	97.8%
and he	0.0%	0.0%	100.0%	and him	27.5%	2.0%	70.5%
she and	1.3%	0.3%	98.4%	her and	2.0%	0.0%	98.0%
and she	0.0%	12.5%	87.5%	and her	9.7%	0.0%	90.3%
we and	0.0%	2.5%	97.5%	us and	0.0%	0.0%	100.0%
and we	0.0%	0.0%	100.0%	and us	0.0%	5.4%	94.6%
you and	----------	----------	----------	you and	----------	----------	----------
and you	----------	----------	----------	and you	----------	----------	----------
they and	0.0%	1.4%	98.6%	them and	0.0%	0.0%	100.0%
and they	0.0%	0.0%	100.0%	and them	2.6%	0.0%	97.4%

Figure 24: Figure 14 repeated. Present Day Danish and Present Day English percentages for expected and unexpected case. The nominative forms are marked with red to show the low amount of unexpected occurrences.

Strong evidence for this type of hypercorrection cannot be found in the data from the present investigation of case. This, however, does not mean that the phenomenon does not exist in English and Danish. As has become apparent from examples (58)-(62), unexpected nominative can be found in various different genres and sources – both spoken and written – in both languages. If we believe the results from chapter 4 to give a representative picture of English and Danish, maybe we should acknowledge that this type of hypercorrection is less common than first anticipated and significantly less common than the phenomenon of unexpected oblique.

5.2.3 A historical look at unexpected case

As has become apparent by now, case is frequently used in unexpected ways in Present Day English and Present Day Danish, oblique regularly being used in subject position and

nominative regularly being used in complement position. However, using the case system in this untraditional manner has not always been as common as it is today. If we compare figure 13 to figure 14, we see that the numbers for unexpected case for Early Modern English and Early Modern Danish are significantly lower than the corresponding present day numbers. In fact, for many of the pronouns, there are no occurrences of unexpected case at all in the Early Modern corpora.

It is plausible that the lack of unexpected occurrences of case in the Early Modern corpora is connected to the transformation of English and Danish from transparent-case languages to vestigial-case languages. As described elaborately in section 2.4, the two languages went from being transparent-case languages to vestigial-case languages as their morphological case systems were dismantled. (Parrott 2009b, 284). The development from being a highly inflecting language to being a case-impoverished one entailed a number of new syntactic properties including the ability to use oblique case in unexpected environments and to have case mismatches inside CoDPs (see subsection 5.2.1). Hence, in Present Day English and Danish, unexpected oblique and case mismatches are socially salient, whereas in transparent-case languages – as English and Danish once were – these phenomena are "completely unattested and unacceptable" (Parrott 2009b, 277). For this reason, it is expected that we will see fewer occurrences of unexpected case as we go further back in the history of English and Danish.

It is hard to say when the rearrangement of the case system took place in English and Danish. However, scholars generally agree that the case inflections were lost gradually starting in the Late Old English/Danish period before being finally established by the end of the Early Modern English/Danish period (Quinn 2005, 14; Rissanen, Kytö, and Heikkonen 1997, 2; Skautrup 1968, 266). During this process, as the traditional case inflections were lost and new properties were entering the languages, the use of case seemed unsystematic and, at times, random (Jensen 2003, 222). The in-between stage in which various competing case systems existed simultaneously gave rise to a very complex case situation in the Early Modern period. As found in figure 13, case is rarely used unexpectedly in Holberg's and Shakespeare's works. In a similar corpora investigation covering the years 1707 to 1913, McFadden (2017, 21) conclude that "the vast majority of examples show consistent use of SF [subject form, i.e., nominative] and OF [object form, i.e., oblique] in the places where we would expect them" and thus that the standard case system is consistently being adhered to.

Hence, from these data, it can be concluded that Early Modern English and Early Modern Danish speakers adhered to the standard case system where nominative = subjects of finite

clauses and oblique = all other the majority of times and to a far greater extent than in Present Day English and Danish.

With that said, there are some indications that it was indeed possible to use case in untraditional ways even in the Early Modern period – see example (64) from the Shakespeare corpus and (65) from the Holberg corpus.

(64) All debts are cleared between you and I (nom.) (*Shakespeare Corpus* 2022, "The Merchant of Venice").

(65) Du har stillet dig afsindig an, for at sætte Vox-Næse paa din Far, og siden løbet bort med en Landstryger, hvorved mig (obl.) og mit heele Huus er sat saadan Klik paa, som aldrig kand afstryges.
"You have positioned yourself badly by deceiving your father and afterwards running away with a vagabond, whereby me (obl.) and my entire house have been besmirched in a way which can never be changed (*Holberg Corpus* 2022, "Kilde-Reysen").

(64) is an example of the hypercorrect unexpected nominative, whereas (65) is an example of the unexpected oblique in subject position. Although these two sentences are some of the only examples of unexpected nominative and unexpected oblique in the Early Modern corpora compiled for the present study, there is ample evidence of the nominative's and oblique's unexpected use in other sources from the Early Modern Period (Angermeyer and Singler 2003, 173). McFadden (2017, 22), for example, gives a few examples of what he calls "PDE-style" in Early Modern English texts:

(66) … and their estates were to go immediately over to their next of kin being protestants, and them (obl.) and their families left to starve (McFadden 2017, 22).

(67) Yet, I hope it won't break the friendship so long subsisting between you and I (nom.) (McFadden 2017, 22).

Examples of unexpected use of case can even be found even earlier than the Early Modern Period. In Jensen's (2003, 223) investigation of Early Middle Danish in the Scanic Law in B 69, she found that as much as 18 out of 74 subjects (that is 24.3%) have unexpected oblique

case. Likewise, 9 out of 152 objects (that is 5.9%) have unexpected nominative (Jensen 2003, 223). The numbers for unexpected case in the Scanic Law are disproportionately high compared to other early Danish and early English texts. For the time being, the Scanic Law will remain a mystery, but it does, however, demonstrate that the unexpected and, at times, seemingly random distribution of nominative and oblique have been present in language for several centuries.

The fact that both unexpected nominative and unexpected oblique can be attested in earlier variants of English and Danish indicates that the syntactic properties that allow for these phenomena are not new but have been developing for a long time. If we leave the Scanic Law out of account, the phenomena seem to have accelerated in popularity during the past centuries going from rare to common. This acceleration has probably taken place simultaneously with the rearrangement of the syntactic properties that allow for sentences such as (64)-(67). In other words: Something indicates that the ability to use case unexpectedly emerged as the English and Danish case systems were dismantled and the remains were reorganised. From here on, the phenomena have only been rising in popularity, as seen by comparing figure 13 to figure 14. Arguably, the phenomena are still increasing in popularity, but these are purely speculations, which this article is unable to delve into.

5.2.4 Morphological case or allomorphs?

As has become evident from the above sections, it simply cannot be that speakers of English and Danish "are wishy-washy when it comes to case" (Schütze 2001, 220). Case variation in English and Danish only occurs in certain environments and certain syntactic constructions, indicating that we are not dealing with random speaker variation but with systematised variation occurring due to certain syntactic properties. That we are dealing with something else than speakers being wishy-washy with their choice of case, can be seen from the lack of vacillation in simple, declarative sentences:

(68) *Me (obl.) am tired (Schütze 2001, 220).

(69) *John likes she (nom.) (Schütze 2001, 220).

(70) *Mig (obl.) er træt.

(71) *John elsker hun (nom.).

In this paper, I follow Schütze (2001) by explaining the oblique case variation as being the result of the default case, which is supplied after spell-out in the instances where "there is no case assigner for the DP in question" (Schütze 2001, 210). The nominative case variation likewise has to do with the absence of a syntactic case specification. The default case, which is supplied in those syntactic environments where case specifications have not been given prior to spell-out, is more easily overridden than 'regular' oblique case assigned from either V° or a P°. Hence, prescriptive rules or viruses in the form of nominative case can easily override the default case resulting in the nominative case form being used in complement position.

As mentioned briefly at the beginning of the current chapter, several scholars have attempted to explain the phenomenon of unexpected case, resulting in a number of different theories linking case to anything from stress to information structure (see Jørgensen 2002; E. Hansen 1972; Heltoft 1990, 127-135; 1992; E. Hansen and Heltoft 2011, 29-30, 181; Jensen 2019, 72-74; Heegård, Jensen, and Schack 2019, 448-449). It is beyond the scope of this project to delve into all these theories, but one explanation for the case variation worth mentioning is the one propounded by Emonds (1985), who suggests that English, in fact, does not have morphological case. Emonds (1985, 297) argues that "morphological "case" of English pronouns is not case of the sort that depends on grammatical relations or abstract case", and hence that "the remnants of case found on English pronouns should not be generated by the mechanisms for morphological case" (Emonds 1985, 220). Instead, Emonds asserts that it is "some other language-particular (hence local) rule [that] must assign nominative case to subject pronouns" (Emonds 1985, 239). This rule can be given as follows: Nominative pronouns reflect the structural property of surface sisterhood with I°, while oblique pronouns appear elsewhere (Emonds 1985, 297). Furthermore, Emonds argue that the nominative pronoun forms are licit only when directly dominated by an S node. Hence, the choice between *I* and *me* in sentences such as (72) is determined solely by their position in the tree (Hudson 2008, 378).

(72) Me (obl.) and Mary went together (Hudson 2008, 378).

When *me* is used unexpectedly, it is because there has been introduced an extra node between the pronoun and S, Emonds (1976, 197) explains.

Emonds (1976; 1985) theory applies only to personal pronouns and only to pronouns in one structural position. "This is very different from the orthodox analyses in which the choice between *I* and *me* is a uniquely visible manifestation of the much more widespread distinction between abstract Case distinctions which apply to all NPs [DPs] in all positions" (Hudson 2008,

378). Whereas the standard generative explanation is that the distribution of nominative and oblique depends on the general theory of abstract Case, Emonds' theory suggests that the nominative and oblique pronoun forms are acquired as allomorphs, which occur only in certain morphological contexts on the basis of certain rules (see also Hudson (2008) and McFadden (2017, 31) for similar analyses).

As accounted for above, this approach to the distribution of pronoun forms requires a number of idiosyncratic ad hoc rules (Emonds (1976; 1985) proposes several other rules as part of his theory, however, these will not be accounted for in the present article). Contrary to Emonds' language-particular theory, the theory of default case, which was elaborately described in subsection 5.2.1, provides us with a parsimonious and general account for a number of unexpected phenomena which occur in the distribution of pronoun case forms. For this reason, it is assumed in this article that Present Day English (and Present Day Danish) does have morphological case (although only a few remnants), and hence that the pronoun forms are distributed by the mechanisms of morphological case. When this mechanism fails to assign case to certain DPs, default case enters the picture and, in the case of English and Danish, supplies the DP in question with default oblique case.

5.3 Sequence in CoDPs

In coordination structures, case is inextricably intertwined with the sequence of the conjuncts. As demonstrated in section 2.6, certain pronoun forms are more acceptable in certain sequences than others. Some sequences, most speakers would avoid using, for example, *X and he* as exemplified in (73) below:

(73) You and <u>he</u> (nom.) have a few problems to sort out (*British National Corpus*, n.d.).

In the BNC, there are no more than 41 examples of CoDPs with the third person singular nominative *he* in the second conjunct, whereas the opposite sequence *he and X* can be found 1,491 times (see appendix 4). The same pattern recurs in KorpusDK, where *X og han* is only found 97 times, while *han og X* occurs 1,443 times (see appendix 3).

Some sequences are also regarded as more acceptable in certain environments. Above, in subsection 5.2.3, we found that many speakers "readily use constructions like *to she and I* and *between he and his friends*" (Angermeyer and Singler 2003, 176), although the use of nominative in such constructions is untraditional and hypercorrect. Although the nominative form is

regularly used in complement positions, not all sequences and combinations are acceptable in this position. *X and I*/*X og jeg*, for example, is commonly used in complement CoDPs, whereas other combinations, "all – or virtually all – speakers would reject: for example, **told I and Kim*" (Angermeyer and Singler 2003, 175). Hence, there seem to be some implicit rules which decide what combinations and sequences are acceptable in coordination structures in certain positions and what are not. The present section will try to identify these special properties in order to account for the behaviour of personal pronouns within CoDPs.

Angermeyer and Singler (2003, 176-178) present three competing patterns for sequence within CoDPs: Vernacular order, Standard order, and Polite order. The three patterns are summarised below in figure 25:

	Subject	Object
Vernacular	me and X	me and X
Standard	X and I	X and me
Polite	X and I	X and I

Figure 25: Three patterns for sequence within CoDPs. (Adapted from Angermeyer and Singler 2003, 178).

What Angermeyer and Singler (2003) call Vernacular order is when the oblique form is used in both complement position (referred to as object position in Angermeyer and Singler (2003)) and in subject position. The last-mentioned is what I described in subsection 5.2.1 as unexpected oblique. In the Vernacular order, the first person singular pronoun is positioned as the first conjunct, which breaks with "the rule of politeness which stipulates that first person pronouns should occur at the end of the coordinate construction" (Quirk et al. 1985, 335). What Angermeyer and Singler (2003) call Polite order is oppositely, when the nominative form is used in both subject position and in complement position. When used in complement position, it is, as I described in subsection 5.2.2, a matter of unexpected oblique (also known as hypercorrection), which occurs "by children's overgeneralizing of adult correction of their use of the Vernacular in subject position" (Angermeyer and Singler 2003, 177). In the Polite order, the rule of politeness is being conformed to. The third pattern for sequence in CoDPs is the Standard pattern, which obeys the standard generative case theory and the rule of politeness.

Both the Standard pattern and the Polite pattern adhere to the rule of politeness by positioning the first person singular pronoun as the last conjunct. This rule constitutes one of the

primary pillars in the formation of coordination structures in both Present Day English and Danish. This can be seen from the great majority of first person singular pronouns which occur in the second conjunct of the CoDPs in both languages:

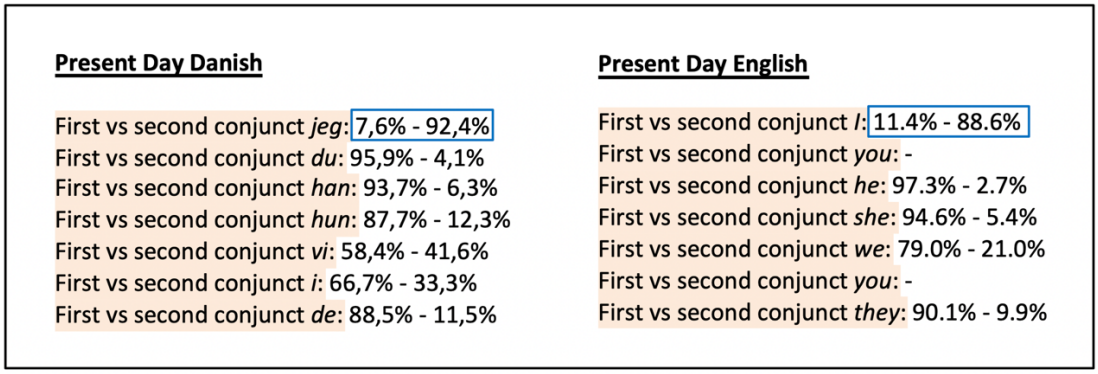


Figure 26: Extract from figure 11 and figure 12. The distribution of the first person singular nominative in the two conjuncts of the CoDPs.

The fact that ≈ 90% of all instances of *I/jeg* occur in the second conjunct of the CoDPs in Present Day English and Danish demonstrates the impact of the politeness rule. According to Jensen, Kragh, and Strudsholm (2018, 82), the superiority of the politeness rule originates in a more general behavioural rule proclaiming always to put yourself last – whether it is at the dinner table, when going through a doorway, or in the queue at the supermarket. This behavioural rule has then "carried over to the linguistic field, where it has become the norm to always mention others before yourself"[13] (Jensen, Kragh, and Strudsholm 2018, 82).

The politeness norm is so strong that for many native English and Danish speakers, it is questionable to have the first person singular pronoun in the first conjunct as in (74) and (75) below, even though the standard generative case conventions (the Standard order) allow it:

(74) Jeg (nom.) og min kone har lært at sætte meget pris på København.
"I (nom.) and my wife have learned to appreciate Copenhagen" (*KorpusDK*, n.d.).

(75) I (nom.) and my sisters visited her (*British National Corpus*, n.d.).

In fact, sentences (74) and (75) sound so 'wrong' to many native speakers of English and Danish (myself included) that the Vernacular order, where the nominative *I/jeg* is replaced by *me/mig* as in (76) and (77) below is more acceptable.

[13] Original text in Danish: "Denne adfærdsregel har afsmittende effekt på det sproglige, hvor det er helt almindeligt at håndhæve at det mest høflige er at nævne andre før sig selv" (Jensen, Kragh, and Strudsholm 2018, 82).

(76) Mig (obl.) og min kone har lært at sætte meget pris på København.
"Me (obl.) and my wife have learned to appreciate Copenhagen."

(77) Me (obl.) and my sisters visited her.

This is even though both the standard generative case theory and the rule of politeness are violated here. Thus, it seems as if the rule of politeness is so strict with the first person singular nominative, that speakers would rather use the Vernacular order and disobey both the standard generative case theory and the rule of politeness (and instead use the oblique *me/mig*), than have *I/jeg* in the first conjunct.

Syntactically, it is also favourable to have the first person singular nominative in the second conjunct – especially when using *I/jeg* in unexpected environments. In sentences such as (78), where the first person singular nominative pronoun is used in complement position, we instinctively want the pronoun as far away from the case assigner as possible.

(78) It has given my wife and I (nom.) a lot of comfort (*British National Corpus*, n.d.).

In (78), the case assigner for the CoDPs *my wife and I* is the V° *given*. Hence, we would expect the CoDPs to have oblique case and the pronoun within the CoDPs to be *me*, not *I*. As given above in subsection 5.2.2, this hypercorrection occurs as a 'virus' that can easily override the default oblique case (Sobin 1997). The *I* which arises from this virus, however, is still, in some sense, affected by the oblique case assigner in V°. Thus, for the clause to be as acceptable as possible, the nominative pronoun *I* should be positioned as far away from V° as possible; that is, in the second conjunct.

The tendency to have the first person singular nominative in the last conjunct of the CoDPs has arguably arisen as a combination of the politeness rule and the syntactic preference to have *I/jeg* as far away as possible from the original case assigner, V°. As was found in figure 26, the tendency, however, is only prevailing for the first person singular pronouns *I/jeg* and *me/mig* – arguably because the norm originates in the politeness of putting others before oneself. As seen from figure 26 above, neither the second and third person singular forms nor the first, second, and third person plural forms show the same distribution with respect to sequence. Oppositely, they show a clear tendency to have the highest frequency in first conjunct – not in the second conjunct. Consider, for example, the English third person singular *he*, which occurs

in the first conjunct 97.3% of the time and only occurs in the second conjunct 2.7% of the time. All English and Danish nominative pronouns – except for the first person singular nominative – follow this pattern. Some pronouns, such as *he,* rarely occur in the second conjunct. For other pronouns, such as the Danish first person plural nominative *vi*, the pattern is less distinct; here, 58.4% of the instances occur in the first conjunct, while 41.6% of the instances occur in the second conjunct. Nevertheless, the pattern is clear: In both English and Danish, the first person singular nominative is most attracted to the second conjunct, whereas the remaining pronouns are most attracted to the first conjunct of the CoDPs.

If we compare the present day numbers to the Early Modern numbers shown in figure 27 below, we see an interesting connection with respect to the development of the politeness norm. Just as in Present Day English and Present Day Danish, the numbers for Early Modern English[14] demonstrate a tendency to have the first person singular pronoun appearing in the second conjunct. Furthermore, as in the present day variants of the languages, the remaining pronouns display the reverse tendency by being far more common in the first conjunct than in the second conjunct (see the red and blue markings in figure 27).

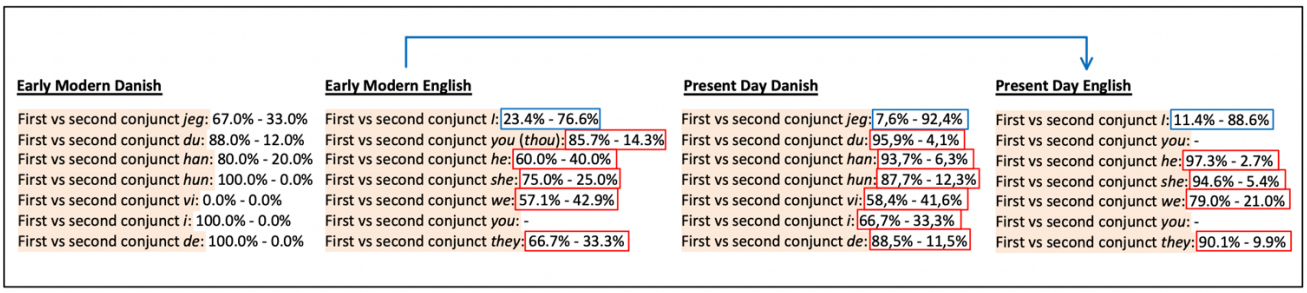


Figure 27: Extract from figure 11 and figure 12. The distribution of the first person singular nominative in the two conjuncts of the CoDPs.

In Present Day English, 88.6% of the instances of *I* occur in the second conjunct of the CoDPs. Although the majority of the instances of *I* also occur in the second conjunct in Early Modern English, the number is slightly lower: Here, 76.6% of the instances of *I* occur in the second conjunct. Hence, the percentage of instances of *I* appearing in the second conjunct has increased by 12 percentage points from Early Modern English to Present Day English. This development illustrates that the politeness norm, which tells us to put ourselves last, has increased in popularity during the past 400 years.

[14] Because the numbers diverge greatly from the other three corpora investigations, Early Modern Danish has been left out of this exact discussion. Due to the small corpora size and few instances of CoDPs, these numbers are believed to be misleading, which is why they have been excluded.

The remaining pronouns have also undergone a change but in the opposite direction. Hence, in the development from Early Modern English to Present Day English, they have become more attracted to the first conjunct instead of the second conjunct. Take, for example, the third person singular pronoun *he*. In Present Day English, *he* occurs in the first conjunct 97.3% of the time. In Early Modern English, *he* only occurs in the first conjunct in 60.0% of the instances. Thus, from Early Modern English to Present Day English, there seems to have been a development in which all pronouns apart from the first person singular *I* became more attracted to the first conjunct. Simultaneously, the first person singular *I* underwent the opposite development, becoming more attracted to the second conjunct. This development has led us to the current point in the English language, where the placement of *I* in the second conjunct rather than the first conjunct has become nearly categorical (Grano 2006, 17).

In a presentation held at the University of Lund in 2017, McFadden presented numbers similar to the numbers proposed above, which showed that the first person singular nominative form *I* was disproportionately common as the second conjunct in the years 1707-1913, whereas the remaining pronouns had the reverse distribution being more attracted to the first conjunct (see figure 28) (McFadden 2017, 26).

	1st conjunct	2nd conjunct	% 1st conjunct
I	32	110	23%
she	20	1	95%
he	81	20	80%
we	10	3	77%
they	19	7	73%
me	54	29	65%

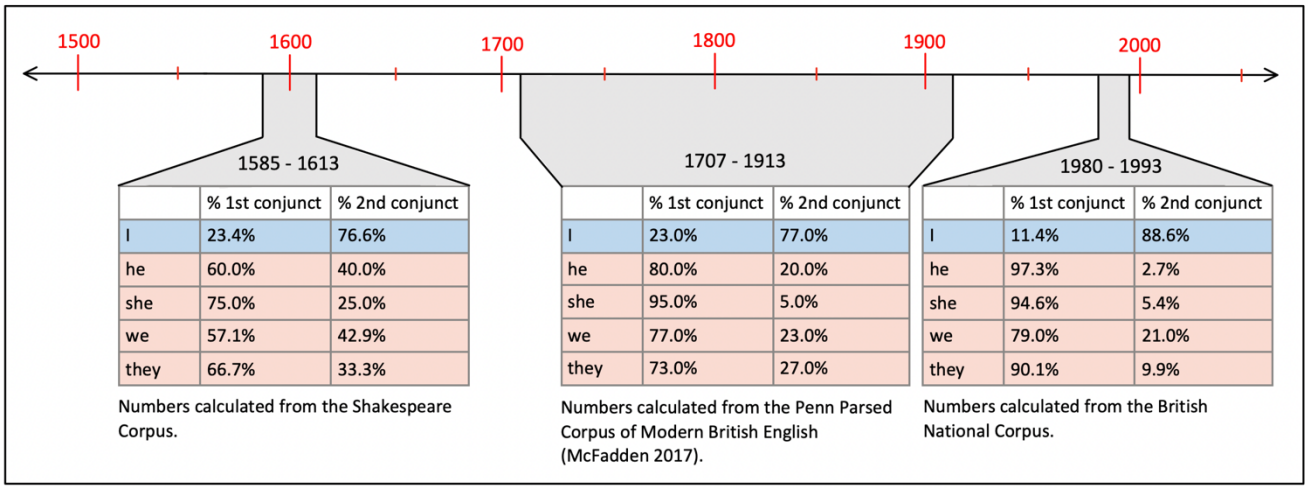
Figure 28: The distribution of the pronouns in CoDPs in the Penn Parsed Corpus of Modern British English, which covers 1707 to 1913 (Adapted from McFadden 2017, 26).

As seen in figure 28, McFadden's (2017, 26) numbers resemble the numbers extracted from the Shakespeare corpus and the BNC. McFadden's numbers are calculated[15] based on data from the Penn Parsed Corpus of Modern British English, which covers the period from 1707 to 1913. Thus, these numbers cover a period that comes after the period which my own numbers from the Shakespeare corpus cover; the numbers extracted from the Shakespeare corpus can quite precisely be determined to be from the time when Shakespeare was active – that is, from 1585

[15] McFadden's (2017, 26) numbers are expectedly calculated in the same way as my numbers and can thus be directly compared.

to 1613. Covering the period from 1707 to 1913, McFadden's numbers come before my own Present Day English numbers, extracted from the BNC, covering the years 1980 to 1993 (Burnard 2015). Thus, timewise, McFadden's data fit perfectly between my two datasets.

Taking McFadden's data from the Penn Parsed Corpus of Modern British English into consideration, we now have numbers from three points in time with roughly 100 years in between them. If we arrange the data from figure 27 and figure 28 into a timeline, we get the following overview of the development of the distribution of pronouns in CoDPs in the English language:



	% 1st conjunct	% 2nd conjunct
I	23.4%	76.6%
he	60.0%	40.0%
she	75.0%	25.0%
we	57.1%	42.9%
they	66.7%	33.3%

Numbers calculated from the Shakespeare Corpus.

	% 1st conjunct	% 2nd conjunct
I	23.0%	77.0%
he	80.0%	20.0%
she	95.0%	5.0%
we	77.0%	23.0%
they	73.0%	27.0%

Numbers calculated from the Penn Parsed Corpus of Modern British English (McFadden 2017).

	% 1st conjunct	% 2nd conjunct
I	11.4%	88.6%
he	97.3%	2.7%
she	94.6%	5.4%
we	79.0%	21.0%
they	90.1%	9.9%

Numbers calculated from the British National Corpus.

Figure 29: Distribution of pronouns in the first and second conjunct of CoDPs in Shakespeare Corpus, Penn Parsed Corpus of Modern British English (McFadden 2017), and the British National Corpus.

From the timeline in figure 29, we see that the percentage of *I* occurring in the second conjunct has risen from 76.6% in 1585-1613 to 77.0% in 1707-1913 and finally to 88.6% in 1980-1993. Hence, over the course of almost 400 years, the tendency to have *I* in the last conjunct has been increasing drastically. From the timeline above, it seems as if most of this development has taken place between the two last datasets, hence between 1913 and 1980.

In subsection 5.2.3, we found that – although there were a few occurrences of unexpected nominative in earlier texts – the phenomenon of using nominative pronoun forms in non-subject environments seemed to have accelerated during the past few centuries. Interestingly, this period coincides with the period in which the politeness rule seems to have increased – that is, from 1913 and forth. Together, these two phenomena constitute the Polite form (Angermeyer and Singler 2003, 178), and hence it can be concluded from the above investigations that, during the last few centuries (or perhaps only during the last century), there has been a rise in the Polite form, where the *X and I* sequence has gained popularity in both subject position and complement position. This finding goes hand in hand with Hock and Joseph's (1996) claim that having a nominative pronoun in initial position in object CoDPs was strictly unacceptable in Early

Modern English. Hence, as the hypercorrect usage of nominative in complement CoDPs became more popular, so did the politeness rule, which required the first person singular nominative pronoun to occur in the last conjunct.

Returning to the timeline in figure 29, we see the reverse tendency in the remaining nominative pronouns. The third person singular nominative and the first and third person plural nominative have gradually, over the course of the last 400 years, become more attracted to the first conjunct. For these pronouns, the development cannot be dated to have taken place during the 20th century. Instead, it seems as if the development has happened gradually during the past five centuries. Consider, for example, the third person singular nominative *he*, which in 1585-1613 occurred in the first conjunct 60.0% of the time. A couple of hundred years later (from 1707-1913), this number had increased to 80.0%. In the most recent calculations, which cover the years 1980-1993, 97.3% of all instances of *he* occur in the first conjunct.

We cannot, with any certainty, conclude when the third person singular nominative and the first and third person plural nominative pronouns started becoming attracted to the first conjunct. We cannot either, with certainty, say when the first person singular nominative *I* started appearing more frequently in the second conjunct than in the first, but from the above findings, we can estimate this change to have accelerated during the past few centuries (see figure 29). However, one thing that can be concluded with certainty is that an interesting and conflicting development has taken place regarding pronoun forms and their sequence within CoDPs.

Conclusively, we can say that the distribution of first person singular pronouns in Danish and English follows the following principles:

- 1sg (almost) always occur as the second conjunct of the CoDPs due to the rule of politeness (see figure 26). This tendency has been increasing during the past 400 years and is presumably still increasing (see figure 29).
- Apart from being used in its expected position (complement position), the 1sg oblique is often used in unexpected environments (that is, in subject position) (see figure 23).
- The 1sg nominative can also be used in unexpected environments (that is, in complement position), but this does not frequently happen (see figure 24). When it does, it is a case of hypercorrection.

So far, the discussion about sequence within CoDPs has revolved primarily around the first person singular pronouns (i.e., *I/jeg* and *me/mig*). Within the pertinent descriptive and formal

literature, most of the attention is indeed devoted to first person singular pronouns (Angermeyer and Singler 2003, 176); however, the remaining pronoun forms also deserve brief attention.

In contrast with the first person singular, "explicit ordering constraints for 3sg pronouns are virtually non-existent" (Grano 2006, 19). However, with that said, a sub-pattern to the Polite pattern described on page 63 does, in fact, permit the third person singular nominative pronouns *he/han* and *she/hun* to occur as part of a coordination structure in complement position, but only in initial position of the CoDPs (Angermeyer and Singler 2003, 178). Hence, the *3sg og/and X* pattern can be found in complement position in both Korpus DK and the BNC:

(79) A separation between she (nom.) and her husband came to seem inevitable (*British National Corpus*, n.d.).

(80) Der opstod tumult imellem han (nom.) og kannikerne.
"Commotion arose between he (nom.) and the canons" (*KorpusDK*, n.d.).

Earlier, it was stated that unexpected nominative (also known as hypercorrection) was restricted to the singular pronoun forms, the majority of occurrences being with the first person singular nominative *I/jeg*. Examples (79) and (80) show that the hypercorrect phenomenon is able to spread to other singular pronoun forms such as the third person singular. It should, however, be mentioned that (79) and (80) are extremely rare constructions, which are only found a few times in KorpusDK and the BNC. Hence, not all speakers who allow for the 'regular' hypercorrection (*X og/and 1sg* pattern used in complement position) allow for constructions such as (79) and (80).

Conclusively it can be stated that the English and Danish third person singular pronoun forms usually behave consistently when part of a coordination structure; that is, they follow the below observations:

- 3sg (almost) always occur as the first conjunct of the CoDPs (see figure 26). This tendency has increased over the past 400 years and is presumably still increasing (see figure 29).
- The 3sg nominative forms always occur in expected environments (that is, in subject position), except for the rare examples in (79) and (80).
- The 3sg oblique forms can be used both in expected and unexpected environments but are, of course, most used in expected environments (that is, in complement position) (see figure 23).

Very little literature has been published on the behaviour of second person singular pronouns. For English, the reason is apparent: Due to syncretisms between the second person singular nominative and oblique and the second person plural nominative and oblique, no distinctions are made between these four forms. Danish still distinguishes between the four forms (see figure 5), but at the time of writing, no published literature brings up the topic of second person singular pronouns in CoDPs. However, based on the investigations carried out for this article, following two observations are made about the Danish *du* and *dig*:

- 2sg (almost) always occur as the first conjunct of the CoDPs (see figure 26 and figure 12). This tendency has been increasing during the past 400 years and is presumably still increasing (see figure 29).
- When the 2sg forms, both nominative and oblique, do occur in second conjunct of the CoDPs, they are likely to also occur in unexpected environments.

The plural pronoun forms are rarely used in CoDPs in neither English nor Danish (see appendices 3 and 4). When coordination constructions with plural pronouns do occur, they usually have the expected case according to standard generative case theory. Unexpected nominative never occurs in CoDPs with plural pronouns, since this phenomenon is restricted to the singular forms – hence, the unacceptability of (63). Unexpected oblique can occur with plural pronouns, as with singular pronouns, since the default case theory can be generalised to cover all numbers, persons, and genders. However, unexpected oblique is rarely seen with plural pronouns. In theory, the default case theory also allows for mixed case to occur in CoDPs with plural pronouns, however, no examples of this are found in the data sets. Due to this lack of occurrences, no conclusions will be drawn about the plural pronoun forms in the present article.

5.4 Possible pitfalls of the study

The present study does have its drawbacks. Most of them are inevitable when working within comparative and historical linguistics but will still be highlighted here for future purposes. Other drawbacks could possibly have been avoided if the time had not been limited.

Arguably, the biggest pitfall of the current project is the lack of sources from older variants of English and Danish. In the initial stages of the project, it was believed that the diachronic development of case from Early Modern English and Danish to Present Day English and Danish would be the focal point of the study. It was even the intention to go further back in time to investigate the dismantling of the case systems in Old- or Middle English and Danish. However,

in both languages, very few reliable sources exist from these time periods. In Danish, there is even an empirical gap lasting from 1000-1300, in which there exists no sources at all from (Jensen 2002, 162). With no sources from Old- or Early Middle Danish, it would be impossible to carry out a comparative study of English and Danish from this period in time. Hence, it was decided to begin the investigation later than first expected, namely in Early Modern English and Danish.

Even here, however, challenges concerning the sources arose. It was decided to use a corpus composed of Shakespeare's complete works and a corpus composed of Holberg's complete works to investigate the two languages in the Early Modern period. However, as discussed in section 5.1 on corpus composition, neither of the playwright corpora are fully representative of language at the time. Not only are the corpora are limited in size – the Shakespeare corpus consisting of 958,268 words and the Holberg corpus consisting of 708,558 words (in contrast, KorpusDK consists of 56 million words and the BNC consists of 100 million words) – but they also have a slightly different composition than the present day corpora. The BNC and KorpusDK, on the other hand, are fully representative of Present Day English and Present Day Danish. This difference in corpora compositions does inevitably constitute a possible source of error for the comparisons of Early Modern English and Danish and Present Day English and Danish.

It is "undoubtedly necessary to consider many data before one can draw firm conclusions on which to base an analysis" (Kemenade 1987, 4). The current project has not drawn any firm conclusions about the Early Modern variants of the two languages for this exact reason. To do so, more extensive corpora of both Early Modern Danish and Early Modern English is required. This is one of the major drawbacks of the current project – a drawback which could partially have been eliminated if time had not been a factor.

Another pitfall, which is a general problem in working on historical syntax, is the absence of information on the ungrammaticality of sentences (Kemenade 1987, 2). "Work on syntactic theory aims to capture the internal grammar of native speakers" (Kemenade 1987, 2), however, we have no native speakers of earlier variants of English and Danish at our disposal. We only have texts, which, at best, reflect "the performance of the scribe who wrote (or copied) the text" (Kemenade 1987, 2). With no native speakers to perform grammaticality judgements, we cannot know if the Early Modern English and Danish sentences in (64)-(67) are even grammatical.

Connected to this problem is the problem of counterexamples. As Kemenade (1987, 3) points out: "There will always be counterexamples to most of the observations about language facts." When dealing with a corpus of written texts in a dead language, "it is often difficult to

assign a status to counterexamples. Are they 'real'? Are they 'slips'? Are there stylistic considerations involved that may well be beyond our reach?" (Kemenade 1987, 3). Such questions will always arise when dealing with historical syntax, and the current project is no exception.

A final, more technical problem for the current study was the lack of software which could distinguish different syntactic structures from each other to identify only the ones which were relevant for my project (that is CoDPs containing at least one personal pronoun). The lack of such software complicated the corpora investigations significantly since all sorting needed to be done manually. With KorpusDK and the BNC, this sorting was especially time-consuming due to the great size of the corpora and thus a large number of search results. Sorting the data manually also entails the possibility of errors. Thus, to lower the error margin, we would need a software to perform this sorting automatically.

Most of the abovementioned drawbacks are related to the investigation of case and sequence in CoDPs in early variants of English and Danish. Because of these problems, I have been hesitant to draw any major conclusions based on the data from the Early Modern English and Early Modern Danish corpora. In this way, the drawbacks mentioned above are not as critical for my study as they could have been. As most drawbacks and issues which arose in the process have been circumvented, the current project is sufficiently reliable to allow conclusions to be drawn with some confidence.

Chapter 6: Conclusion

The present thesis has sought to uncover the behaviour of English and Danish pronouns within CoDPs with respect to case and sequence. Through the study of four corpora from two periods in time, the project aimed to examine the special properties of personal pronouns, which give rise to unexpected use of case in CoDPs in English and Danish.

Placed within a generative framework, the current study assumes the remnants of morphological case in the English and Danish pronominal systems to be distributed based on abstract Case (Chomsky 1981, 170). Assuming that morphological case agrees with the abstract Case (Crystal 2008, 67), we expect the following distribution of case in Present Day English and Present Day Danish: Subjects of finite clauses are assigned nominative case, and everything else (most often that is objects (of verbs) and complements of prepositions) is assigned oblique case (e.g., Crystal 2008, 6, 129, 210, 328; Quinn 2005, 26; McCreight 1988, 2; Chomsky 1981, 170; 1993; Pollard and Sag 1994; Burzio 2000).

However, this account of case does not consider the numerous examples of case being used in unexpected ways in English and Danish. In CoDPs in both English and Danish, oblique pronoun forms are regularly used in subject position and occasionally, nominative pronoun forms are also used unexpectedly, that is, in complement position. Furthermore, in both languages, case can occur mixed within CoDPs, e.g., *Him* (obl.) *and I* (nom.) *got on very well together.* Contributing further to the complexity of the behaviour of pronouns within CoDPs is the inextricable intertwining of case and the sequence of conjuncts making *Him and I* significantly more acceptable than *I and him.*

From the investigations of four corpora – two corpora from the Early Modern period, Holberg corpus and Shakespeare corpus, as well as two present day corpora, KorpusDK and the BNC – it was found that the use of oblique in unexpected syntactic environments is a widespread phenomenon in both English and Danish. In fact, ≈ 10-20% of all occurrences of the first person singular oblique *me/mig* in CoDPs occur in unexpected positions, that is, in subject position. This deviation indicates that speakers of Present Day English and Present Day Danish do not adhere to the standard conventions of case in generative linguistics, where oblique is assumed to be assigned by a V° or a P° to the DP which it governs. Instead, this unexpected distribution of oblique case can be explained by Schütze's (2001) theory of default case, which proposes that those DPs that, for a variety of reasons, have not had their case inflection determined by the syntax are supplied with default case after spell-out to satisfy the Case Filter and ensure that the clause is grammatical (Schütze 2001, 205; Chomsky 1981, 49). The default case of English and Danish is oblique, meaning that in those instances where a subject DP is not

assigned nominative case by I° as is expected according to the standard case conventions, it will receive oblique case (Schütze 2001, 210, 227). Default case can be supplied to both conjuncts of the coordination, only one of the conjuncts, and in most instances to neither of the conjuncts resulting in the expected case distribution. In the instances where only one of the conjuncts receives default case, we are left with the phenomenon of mixed case, which is socially salient in both English and Danish.

The high number of occurrences of unexpected oblique in Present Day English and Present Day Danish indicates that it is not the standard case system alone that determines what morphological case is used where. Another important syntactic mechanism, namely default case, also plays a role in the distribution of case. It is the alternation between these two mechanisms that causes the unexpected and unpredictable use of oblique form in subject position in Present Day English and Present Day Danish.

The opposite phenomenon, unexpected nominative, also known as hypercorrection, is also apparent in the data, although much less prominent than unexpected oblique. Unexpected nominative has to do with the absence of a syntactic case specification, just as unexpected oblique. Those instances where case specifications have not been given prior to spell-out and where default oblique case has been supplied are more easily overridden than 'regular' oblique. Hence, prescriptive rules and viruses – in this case in the form of nominative – can easily replace the default case resulting in the nominative case form being used unexpectedly (Schütze 2001, 220). Although seen regularly in English and Danish speech and writing, unexpected nominative appears less common than anticipated. In fact, it is only between 0.5% and 1.7% of all occurrences of *I/jeg* in CoDPs that are used unexpectedly. However, with that said, figures 13 and 14 do indicate that there has been a rise in both unexpected nominative and unexpected oblique from the Early Modern period to the present day.

The corpora investigations of pronouns in CoDPs in English and Danish furthermore demonstrated tendencies that help uncover the implicit rules which decide what sequences are acceptable in coordination structures and what are not. One of the primary tendencies in the formation of coordination structures is the almost categorical tendency to have *I/jeg* positioned in the second conjunct, while the remaining pronouns in the vast majority of cases are placed in the first conjunct. This so-called politeness norm, "which stipulates that first person pronouns should occur at the end of the coordinate construction" (Quirk et al. 1985, 335) has only increased in popularity during the past 400 years. By comparing the Present Day English numbers (see figure 27) to the same numbers but for Early Modern English (see also figure 27) as well as similar numbers extracted from McFadden (2017, 26) (see figure 28), we find that the

percentage of *I* occurring in the second conjunct has risen from 76.6% in 1585-1613 to 77.0% in 1707-1913 and finally to 88.6% in 1980-1993. Together with the occasional use of unexpected nominative, the politeness norm constitutes what Angermeyer and Singler (2003, 178) call the Polite form. The Polite form, i.e., the *X and I* sequence, seemed to have gained popularity in both subject position and complement position. However, at the same time, and even more evident from the corpora investigations, unexpected oblique has increased drastically in popularity during the past 400 years.

As has become evident, it simply cannot be that speakers of English and Danish "are wishy-washy when it comes to case" (Schütze 2001, 220). Case variation in English and Danish only occurs in certain environments and certain syntactic constructions, indicating that we are not dealing with random speaker variation but with systematised variation occurring due to certain syntactic properties. The present thesis has accounted for this systematised variation in case and sequence in English and Danish coordination structures by ascribing the variation primarily to default case and politeness norms. For future purposes, improving the methodology according to the limitations described in section 5.4 could provide more reliable results allowing conclusions to be drawn with even more confidence. Especially it would benefit future studies to have a broader set of sources from older variants of English and Danish as well as more advanced software to distinguish only the relevant syntactic structures in the given corpora. With the necessary tools, it will be feasible and beneficial to conduct further and more exhaustible research within the field of case and sequence in coordination structures.

Appendices

Appendix 1: Early Modern Danish data set.

Appendix 2: Early Modern English data set.

Appendix 3: Present Day Danish data set.

Appendix 4: Present Day English data set.

Reference List

Aarts, Bas. 2014. “Hypercorrection.” In *The Oxford Dictionary of English Grammar*. 2nd ed. Oxford: Oxford University Press.

Allen, Cynthia. 1995. *Case marking and reanalysis*. Oxford: Clarendon Press.

Angermeyer, Philipp S., and John Victor Singler. 2003. “The case for politeness: Pronoun variation in co-ordinate NPs in object position in English.” *Language Variation and Change* 15: 171-209. https://doi.org/10.10170S0954394503152027.

Anthony, Laurence. 2022. AntConc (Version 3.5.9). Computer Software. Tokyo: Waseda University. Available from https://laurenceanthony.net/software/antconc.

Bergs, Alexander, and Laurel J. Brinton. 2012. *English Historical Linguistics. Volume 1: English Historical Linguistics*. Berlin: De Gruyter.

Bernstein, Theodore M. 1965. *The careful writer: A modern guide to English usage*. New York: Atheneum.

Bertelsen, Henrik. 1939. *Dansk Sproghistorie til Skolebrug*. København: Nordisk Forlag.

Brezina, Vaclav, and William Platt. 2022. #LancsBox X 1.1.0. Computer software. Lancaster: Lancaster University. Available from http://corpora.lancs.ac.uk/lancsbox/downloadx.php.

British National Corpus (BNC 1994 XML Edition). n.d. British National Corpus Consortium, Oxford Text Archive. Available from http://hdl.handle.net/20.500.12024/2554.

Brøndum-Nielsen, Johs. 1935. *Gammeldansk Grammatik i sproghistorisk Fremstilling III. Substantivernes Deklination.* København: J.H. Schulz.

Burnard, Lou. 2015. “What is the BNC?” British National Corpus. Accessed July 4, 2022. http://www.natcorp.ox.ac.uk/corpus/index.xml?ID=intro.

Burzio, Luigi. 2000. “Anatomy of a Generalization”. In *Arguments and Case: Explaining Burzio’s Generalization,* edited by Eric Reuland, 195-240. Amsterdam: John Benjamins. https://doi.org/10.1075/la.34.10bur.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110884166.

———. 1993. “A Minimalist Program for Linguistic Theory”. In *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, edited by Ken Hale and Samuel Jay Keyser, 1-52. Cambridge, Massachusetts: MIT Press. https://doi.org/10.7551/mitpress/10174.003.0005.

Crystal, David. 2008. *A Dictionary of Linguistics and Phonetics*. Hoboken: John Wiley & Sons, Incorporated. https://doi.org/10.1002/9781444302776.

Det Danske Sprog- og Litteraturselskab. n.d. “Facts about KorpusDK”. KorpusDK. Accessed

July 4, 2022. https://ordnet.dk/korpusdk_en/facts?set_language=en.

Donaldson, Bruce. 2017. *Dutch: A Comprehensive Grammar*. London: Routledge. https://doi.org/10.4324/9781315620787.

Dyrlund, Niels Kristian Folmer. 1895. "Anmældelse med sproghistoriske indskud." Review of *Dansk Sproglære med sproghistoriske Tillæg: Haandbog for Lærere og viderekomne*, by Kristian Mikkelsen. *Arkiv för nordisk filologi 11*, 1895. Accessed August 27, 2022. https://journals.lub.lu.se/anf/issue/view/1754.

Emonds, Joseph. 1976. *A transformational approach to English syntax*. New York: Academic Press.

———. 1985. *A unified theory of syntactic categories*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110808513.

———. 2010. "Case Theory Revisited: Nominative and Accusative Super Case." In *Karlík a továrna na lingvistiku,* edited by Aleš Bičan, Jan Klaška, Petra Macurová, and Jana Zmrsliková, 98-124. Brno: Masaryk University Press.

Fritz, Matthias. 2003. "Proto-Indo-European Syntax." In *Indo-European Linguistics*, edited by Michael Meier-Brügger, 238-276. Berlin: De Gruyter. https://doi.org/10.1515/9783110895148.238.

Grano, Thomas. 2006. ""Me and her" meets "he and I": Case, person, and linear ordering in English coordinated pronouns." Honors Thesis. Stanford University.

Hansen, Aage. 1965. *Vort Vanskelige Sprog.* 2nd ed. Copenhagen: Grafisk Forlag.

Hansen, Erik. 1972. "Dr. Jekyll og Mr. Hyde i Dansk Grammatik. En filologisk skitse. Papir I, bunke 4:1-13." In *Glæden ved grammatik. Udvalgte artikler og afhandlinger*, edited by Henrik Galberg Jacobsen and Henrik Jørgensen. København: Hans Reitzels Forlag.

Hansen, Erik, and Lars Heltoft. 2011. *Grammatik over det Danske Sprog*. Odense: Syddansk Universitetsforlag.

Hartmann, Katharina. 2015. "Coordination." In *Syntax - Theory and Analysis*, edited by Tibor Kiss and Artemis Alexiadou, 478-514. Berlin: De Gruyter Mouton. Accessed August 24, 2022. https://www.degruyter.com/serial/hsksyn%20ta-b/html#volumes.

Heegård, Jan, Eva Skafte Jensen, and Jørgen Schack. 2019. "Tak til alle de der gør en indsats for vores sikkerhed: Bagom dronningens kasusvalg." *MUDS - Møderne om Udforskningen af Dansk Sprog* 17: 443-456.

Heltoft, Lars. 1990. "Talesprogets tale." *MUDS - Møderne om Udforskningen af Dansk Sprog* 3: 126-143.

———. 1992. "Topologiens plads i en sprogteori." *Sprogvidenskabelige arbejdspapirer fra*

Københavns Universitet 2: 67-98.

Heltoft, Lars, and Marita Akhøj Nielsen. 2019. "Fra gammeldansk til nutidsdansk." In *Dansk Sproghistorie III. Bøjning og bygning*, edited by Ebba Hjorth, Henrik Galberg Jacobsen, Bent Jørgensen, Birgitte Jacobsen, Merete Korvenius Jørgensen and Laurids Kristian Fahl, 129-225. Aarhus: Aarhus Universitetsforlag.

Hock, Hans Henrich, and Brian D. Joseph. 1996. *Language history, language change, and language relationship: An introduction to historical and comparative linguistics.* Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110214307.

Holberg Corpus. 2022. Compiled and edited by Freja Bang Lauridsen.

Hudson, Richard. 1995. "Does English really have case?" *Journal of Linguistics* 31 (2): 375-392. https://doi.org/10.1017/S0022226700015644.

Jensen, Eva Skafte. 2002. "Kasus i gammeldansk." *Ny forskning i Grammatik* 9: 161-184. https://doi.org/10.7146/nfg.v10i9.23902.

———. 2003. "Case in Middle Danish: A double content system." In *Historical Linguistics 2001*, edited by Barry J. Blake and Kate Burridge, 221-236. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/cilt.237.15jen.

———. 2017. "Det er mig: En forskningshistorisk redegørelse for kasusvalget i ældre danske prædikativer." *Danske Studier 2017*: 5-26.

———. 2018. ""It is me" - the replacement of the nominative by the oblique form in Danish subject complements." *Acta Linguistica Hafniensia* 50 (1): 52-72. https://doi.org/10.1080/03740463.2018.1436414.

———. 2019. "Peter og mig, vi gik i biografen: Paratagmer i ekstraposition i dansk." *Ny forskning i Grammatik* 26: 71-88. https://doi.org/10.7146/nfg.v0i26.115994.

Jensen, Eva Skafte, Kirsten Jeppesen Kragh, and Erling Strudsholm. 2018. "Kasusdistribution i personlige pronominer i danske, franske og italienske paratagmer." *Ny forskning i grammatisk* 25: 70-88. https://doi.org/10.7146/nfg.v26i25.109303.

Johannessen, Janne Bondi. 1998. *Coordination*. New York: Oxford University Press.

Jørgensen, Henrik. 2000. *Studien zur Morphologie und Syntax der festlandskandinavischen Personalpronomina – mit besonderer Berücksichtigung des Dänischen*. Aarhus: Aarhus University Press.

Kemenade, Ans van. 1987. *Syntactic Case and Morphological Case in the History of English*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110882308.

KorpusDK. n.d. Det Danske Sprog- og Litteraturselskab. https://ordnet.dk/korpusdk.

Lass, Roger. 1992. "Phonology and Morphology." In *The Cambridge History of the English*

Language Volume II, edited by Norman Blake, 23-155. Cambridge: Cambridge University Press. https://doi.org/10.1017/CHOL9780521264754.003.

Levin, Israel. 1844. "Om Ordforklaringerne til Holbergs Comedier." *Fædrelandet*, February 16, 1844. Accessed August 27, 2022. https://go.ly/Cybh9.

McCreight, Katherine L. 1988. "Multiple Case Assignments." PhD dissertation. Massachusetts Institute of Technology.

McFadden, Thomas. 2017. "The diachrony of impoverished case in English: Workshop on Case-impoverished Germanic." Presentation, Lund, Sweden, October 10, 2017. Accessed August 22, 2022. https://goo.gl/RSRpf3.

———. 2020. "Case in Germanic." In *The Cambridge Handbook of Germanic Linguistics,* edited by Michael T. Putnam and B. Richard Page, 282-312. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108378291.014.

Nielsen, Michael Lerche, and Marie Stoklund. 2019. "Runetiden." In *Dansk Sproghistorie III: Bøjning og bygning*, edited by Ebba Hjorth, Bent Jørgensen, Henrik Galberg Jacobsen, Birgitte Jacobsen, Merete Korvenius Jørgensen, and Laurids Kristian Fahl, 11-40. Aarhus: Aarhus Universitetsforlag.

Parrott, Jeffrey Keith. 2009a. "Case variation in coordination: Danish vs. Faroese." *Nordlyd* 36 (2): 165-185. https://doi.org/10.7557/12.231.

———. 2009b. "Danish Vestigal Case and the Acquisition of Vocabulary in Distributed Morphology." *Biolinguistics* 3 (2-3): 270-304. https://doi.org/10.5964/bioling.8711.

———. 2013. "Introduction: Case variation and change in the Nordic languages." *Nordic Journal of Linguistics* 35 (3): 213-218. https://doi.org/10.1017/S0332586513000061.

Petersen, Niels Matthias. 1858. *Bidrag til den danske Literaturs Historie 4. Holbergs tidsalder 1700-1750*. København: Brødrene Berling.

Pintzuk, Susan. 2002. "Morphological case and word order in Old English." *Language Sciences* 24 (3-4): 381-395. https://doi.org/10.1016/S0388-0001(01)00039-0.

Pollard, Carl, and Ivan Andrew Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Quinn, Heidi. 2005. *The Distribution of Pronoun Case Forms in English.* Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/la.82.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Radford, Andrew. 2004. *Minimalist Syntax: Exploring the Structure of English*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511811319.

Ringe, Don. 2017. “Proto-Germanic.” In *From Proto-Indo-European to Proto-Germanic*, 241-330. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198792581.001.0001.

Rissanen, Matti, Merja Kytö, and Kirsi Heikkonen. 1997. *Grammaticalization at Work: Studies of Long-term Developments in English.* Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110810745.

Rosenbach, Anette. 2002. *Genitive Variation in English: Conceptual Factors in Synchronic and Diachronic Studies*. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110899818.

Sapir, Edward. 2014. “Types of Linguistic Structure.” In *Language: An Introduction to the Study of Speech*, 127-156. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139629430.007.

Schütze, Carson T. 2001. “On the Nature of Default Case.” *Syntax* 4 (3): 205-238. https://doi.org/10.1111/1467-9612.00044.

Shakespeare Corpus. 2022. Compiled and edited by Freja Bang Lauridsen.

Skautrup, Peter. 1968. *Det Danske Sprogs Historie: Første Bind*. Copenhagen: Gyldendal. First published 1944.

Straaijer, Robin. 2020. “Linguistic Prescriptivism”. Oxford Bibliographies. Accessed July 5, 2022. https://doi.org/10.1093/OBO/9780199772810-0208.

Sobin, Nicholas. 1997. “Agreement, Default Rules, and Grammatical Viruses.” *Linguistic Inquiry* 28 (2): 318-343.

Thráinsson, Höskuldur. 2007. *The Syntax of Icelandic.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511619441.

Valin, Jr, Robert D. van. 2001. *An Introduction to Syntax*. Cambridge: Cambridge University Press.