Inflectional economy and politeness : morphology-internal and morphology-external factors in the loss of second person marking in Dutch

Aalberse, S.P.

Publication date
2009
Document Version
Final published version

Citation for published version (APA):
Aalberse, S. P. (2009). *Inflectional economy and politeness : morphology-internal and morphology-external factors in the loss of second person marking in Dutch*. LOT.

Chapter 5

Diachronic variation and the economy hypothesis

1   Introduction

In chapter 3, we concluded that the loss of a second person singular pronoun is cross-linguistically rare and not understandable from a socio-pragmatic angle. In chapter 4, we saw that feature structure provides the missing link in an explanation for the loss of the pronoun *du* and the suffix *–s*. We hypothesized that the pronoun *du* was not only replaced by the polite pronoun *gi* because of politeness reasons, but also because of inflectional economy. If we take inflectional economy into account, we can understand why English and Dutch have lost the second person singular pronoun whereas other languages have not. English and Dutch belong to the small group of languages where replacement of second person singular inflection by second person plural inflection yields a more economical paradigm. We will continue to follow Brown & Gilman (1960), who refer to informal address terms as T (derived from the second person singular *tu* in Latin). Formal address terms will be referred to as V (derived from the second person plural *vos* in Latin).

In this chapter, we will look at additional evidence for the economy hypothesis by comparing second person pronouns and inflection in thirteenth and sixteenth century Dutch texts. In section 2, we explain the selection of historical texts. In section 3, we formulate predictions based on the data. We present the results in section 4. Since the economy hypothesis is applicable to both Dutch and English, section 5 discusses studies on English which describe similar data searches. Section 6 concludes the chapter.

2   Data selection

All texts used for this chapter were gathered by Jacqueline Evers-Vermeul (2005) and Ninke Stukker (2005) from a number of different sources, such as CD-roms (e.g., the CD-roms *Middelnederlands* 'Middle Dutch' and *Klassieke literatuur* 'Classical literature') and the Internet (e.g., the project *Laurens Jansz. Coster*). For each period, we used a corpus of at least 2000 pages of computerized texts. We present an overview of all electronic sources in appendix A. We also specify the digital sources in each text in appendix B.

Computerized texts are useful because they make it possible to perform automatic searches. As a result, we are able to analyze a large amount of texts,

which leads to the analysis of many address forms. Computerized texts also allow other researchers to replicate the searches, thus facilitating verification of the data. Although the practical advantages of using computerized texts are obvious, there is a potential caveat worth mentioning. Namely, some texts not be are not properly represented in a digital format. It is possible that this will bias the sample we use. Gerritsen (1987: 18) and Evers-Vermeul (2005: 51) compare the exclusion of non-computerized texts with the refusal of people to take part in a survey of social differences. Ideally, all respondents would take part in a survey. Likewise, it would be ideal if all texts from a period would be preserved and digitalized. In reality, however, this is never the case. Some people (and in our case, some texts) are not included. Despite this caveat, we are confident that the pros of using computerized texts outweigh the cons, as it creates a much larger sample than would be possible in the alternative, i.e searching manually.

The texts in this chapter are taken from the thirteenth century and the sixteenth century. The oldest available Dutch data are from the thirteenth century. Vor der Hake (1908) shows that in the thirteenth century, V is already used in singular contexts. When the V-pronoun is first used, the preference for V is motivated only by socio-pragmatic factors. In the corpus of texts from the thirteenth century, we do not expect to find effects of economy strategies. The goal of this chapter is to contrast two periods: a period where the choice between T and V is motivated solely by socio-pragmatic factors and a period where inflectional economy also plays a role in the choice between T and V. Muller (1926: 82) dates the loss of *du* to around the sixteenth century. Vor der Hake (1908, 1915) marks 1618 as the year when the pronoun *du* died since this is the year when *du* into longer appeared in bible translations. If the loss of T is related to inflectional economy, we expect to find effects of inflectional economy on the choice between T and V in the sixteenth century, when use of the T-pronoun was waning in all contexts and in all domains.

For both the thirteenth and sixteenth century texts, we selected texts from all geographic regions. In the thirteenth century, the T-pronoun was not under inflectional pressure in any of the Dutch texts. In all regions, the choice between T and V was motivated only by socio-pragmatic factors. We contrast texts from the thirteenth century with texts from the sixteenth century because, in the sixteenth century, we expect effects of inflectional economy.

In chapter 4, we claimed that replacement of T by V yields a more economical paradigm when (i) V combines with first or third person singular inflection and (ii) if T combines with inflection that is not first or third person

singular inflection. We expect to find effects of inflectional economy only in those texts which obey both criteria.

In chapter 4, we also saw that in German, the T-pronoun is retained, despite the observation that the V-pronoun combines with more economical inflection. We related absence of T-loss in German to a more general observation that Standard German is less prone to deflection than other West-Germanic languages due to less dialect contact and language contact. It is possible that our data contain texts from a region where speakers use a more conservative dialect like Standard German. This might mean that texts from regions where inflectional economy does not play a role have been included in the sixteenth century corpus. If this is so, it will make it more difficult to find evidence in favour of the economy hypothesis.

Having selected the periods of the texts we will investigate, we now consider what kind of texts we should study. Research on English address forms and politeness theory has focused on drama as the main source, especially that of Shakespeare. Brown & Gilman (1989: 159) list three reasons for their focus on tragedies by Shakespeare: (i) dramatic texts provide the best information on colloquial speech of the period; (ii) the psychological soliloquies in the tragedies provide access to the inner life that is necessary for a proper test of politeness; and (iii) the tragedies represent the full range of society in a period that is of high relevance to politeness theory. In earlier research on address forms, Brown & Gilman (1960) base their theory on letters, plays, and legal proceedings. In the small number of studies investigating address forms on other types of texts, we find a particular interest in letters, trials and testimonies. These types of text are sometimes preferred because they include a large number of address forms (cf. Bentivoglio 2003, Betsch 2003, Walker 2003).

If one's goal is to investigate the influence of social factors such as class and gender on speakers' use of address forms or if one is interested in the pragmatic effects of adversion (cf. Aalberse 2005, Honegger 2003), it is only understandable to include texts that exhibit a large number of second person forms (for example letters, and trials). If we use texts with a few address forms, it is often difficult to interpret the actual meaning of an address form because we cannot contrast it with other uses of the same form within that same text. Comparison with address forms from other texts is more difficult because many additional variables enter the equation, for example, personal traits and style of the author and geographical region. In the present study, we are interested in formal factors that influence the choice between T and V. These formal effects should be the same for all texts in a certain time period, regardless of the author or region. Therefore, our sample can

include any genre of text, even if the total number of address forms within the text is small.

The next question we will consider is whether to include rhyming texts in our sample. Although we want as comprehensive sample as possible, this question is relevant because rhyming texts often include more archaic language then non-rhyming texts. Debaene (1977: 313) claims that in French prose, the rhymes are frequently directly copied from older versions of the text. Debaene (1977) claims that a similar strategy might have been used in the sixteenth century Dutch text *Droefliken strijt van Roncevale*, which is part of our corpus. There are more texts where the rhyming and the non-rhyming parts can be linked to different time periods. For example, in our sixteenth century corpus we find the texts *Een schoone historie vander borchgravinne van Vergi.* In this text, the rhyming schemes are copied from a text from 1524. The prose version dates back at least thirty years. (Resoort 1988: 156). Because of the reported differences between rhyming and non-rhyming texts we will distinguish these two types of text un our corpus search.

The relation between older texts and the texts in our corpus are not isolated cases. Many books published in the sixteenth century retell popular stories from earlier times. For example, the sixteenth century text *Historie van den wonderlicken Merlijn* shares content with the thirteenth century Dutch rhyming text *Boek van Merline* written by Van Maerlant which was written as a liberal adaptation of the French *Lestoire de Merlin*. We also find the story of Merlyn in older German, Spanish and Latin texts. The direct source for the Merlyn edition in the sixteenth century Dutch corpus is not a Middle Dutch text, but an English rhyming text. The English rhyming text is entitled *A lytel treatyse of ye byrth and prophecy of Marlyn* and was published first in 1499 (Resoort 2008).

Some translators and adaptors base their work on single sources, while others diversify their base. The prose text *Margarieta van Limborch* is so different from earlier Dutch rhyming versions that it is not clear whether the author ever had the opportunity to read a rhyming version (Resoort 1988: 44 fn2). In contrast, the text *Historie van den vier heemskinderen* relies directly on earlier texts. The original text was written substantially earlier than 1508. Debaene (1977: 313) gives an overview of reprints that appeared in the period between 1500 and 1540 and the only book on this reprint list which was included in our corpus is *Historie van den vier heemskinderen*. The text from 1508 is actually a reprint from a prose text from 1490, which is an adaptation of an earlier rhyming text. According to Debaene (1977: 305), the author of the prose version relies greatly on the older rhyming version.

The address term behaviour in *Historie van den vier heemskinderen* shows that the text is directly related to older texts. The number of T-forms is still high (75 T-

subjects as opposed to 19 T-subjects in all other sixteenth century prose texts together). Moreover, the grammatical behaviour of T in *Historie van den vier heemskinderen* is much more similar to thirteenth century texts than it is is to the sixteenth century texts. We do not include data from the text *Historie van den vier heemskinderen* in chapter 4, because the text is the only early reprint in our corpus, a reprint that relies directly on an even older source. Moreover, the text shows deviant behaviour in comparison to all other sixteenth century texts. Raw data from the text *Historie van den vier heemskinderen* are included in appendix D.

3   Predictions

Now that we looked at the selection criteria for the historical texts used in this chapter, let us move on to the predictions that follow from the economy hypothesis. In chapter 4, we looked at possible forms of independent empirical evidence for the role of inflectional economy in pronoun loss on the basis of pronoun loss in French and Brazilian Portuguese. In French and Brazilian Portuguese, the traditional first person plural pronouns *nous* and *nós* respectively were in competition with the indefinite pronoun *on* (French) and *a gente* (Brazilian Portuguese). The traditional first person plural pronoun combined with a unique suffix: *–ons* and *-mos*, whereas the indefinite pronoun combines with default third person singular inflection. Replacing the traditional first person plural pronoun by the indefinite pronoun provides the speaker with the opportunity to select more economical inflection while maintaining the pronoun suffix relation as provided in the input.

The question is: How can we test whether inflectional economy plays a role in the competition between the traditional pronoun and the new (i.e., indefinite) pronoun? In chapter 4, we looked at two pieces of evidence for the role of inflectional economy in pronoun loss, namely circumstantial evidence and independent formal evidence. In this chapter, we will focus on independent formal evidence for the role of inflectional economy in the loss of T. First, however, we will briefly review the evidence that was described in chapter 4.

The first piece of circumstantial evidence for the role of deflection in pronoun loss is the stylistic distribution of the change. In chapter 4, we saw that the new pronoun rises fastest in informal speech registers. Changes that begin in informal speech are associated with change from below and deflection is typically a form of change from below. The preference for the new pronoun in informal speech is thus in line with the hypothesis that pronoun loss is a form of deflection. The association of the V-pronoun with politeness interferes with the predicted

association of the new pronoun with informal settings. Politeness predicts use of V in formal settings (cf. Brown & Gilman 1960). What we see in English and in Dutch, however, is that *du* and *thou* are retained longest in archaic formal text types such as bible texts (Muller 1926a, Vor der Hake 1908, 1915, Wales 1983, 1996, 2004). Busse (2002: 95) shows that *thou* is used more frequently in archaic speech and in poetry. The decreased use of T in formal and archaic texts is in line with the economy hypothesis.

The second piece of circumstantial evidence that we discussed in chapter 4 was the correlation between immigration and loss of the traditional pronoun. In both French and Brazilian Portuguese, we saw that the rise of the new pronoun that combines with economical inflection coincides with massive migration. Migration implies language contact and language contact increases the chance of deflection. If pronoun loss is associated with deflection, the co-occurrence with migration is expected. As predicted, the timing of the loss of *du* in Dutch also coincides with an increase in migration: At the time *du* was lost, we also observe increased economic growth and urbanisation (Boyce Hendriks 1998, De Vries 1984, De Vries & Van der Woude 1995). Wales (1983, 1996) shows that the loss of *thou* in English also correlates with urbanization and immigration.

We will now move on to linguistic evidence for the role of inflectional economy for Dutch *du*. One piece of evidence has to do with the observation that the traditional pronoun is retained longest in combination with high frequency verbs. Since high frequency verbs can resist deflection longer, the preference for the traditional pronoun to combine with high frequency verbs is expected. Section 3.2 reports on the correlation between high frequency verbs and pronoun retention.

A second piece of evidence was presented in chapter 4 on French data. This had to do with the extension of grammatical domains where the new pronoun *on* is used. We saw that the pronoun *on* could be inserted in cleft sentences such as *c'est nous qu'on est les vainqueurs*. ('it is us who are the winners'). In this cleft sentence, the pronoun *nous* carries stress and would normally trigger first person plural inflection. The insertion of the pronoun *on* in this cleft sentence is noteworthy. In this sentence, *on* does not replace *nous* but it is inserted in a position that is usually empty. The insertion does not serve a pragmatic purpose. The only reason for the extra insertion of *on* is the fact that *on* triggers third person singular inflection.

The trend we just mentioned in French is not exactly the same as in Dutch. The difference between the French competition between the first person plural forms *on* and *nous* and the Dutch competition between *du* and *gi* is the grammatical status of the new pronoun. In Dutch, the pronoun *gi* can be used in the same grammatical

contexts where *du* is used. The French form *on*, is a clitic and is thus only in competition with *nous* in a limited context, namely only if *on* is as an unstressed subject. Since the pronouns *du* and *gi* have the same grammatical status, it is not possible that the use of *gi* increases as the result of pressure from inflectional economy because, there are no domains to extend to.

The observation that *du* and *gi* have a similar grammatical distribution does, however, yield another prediction. *Du* and *gi* are not only in competition in contexts where inflection is triggered (as was the case with French *nous* and *on*) but also in contexts where verbal inflection is not triggered. If inflectional economy is a motivation to use *gi* rather than *du*, we predict that the preference for *gi* is stronger in subject function than in non-subject function, since only subjects can trigger non-economical inflection. The data search investigating this prediction is presented in 3.1. In 3.2, we will look into the prediction that T is most likely to combine with high frequency verbs in the sixteenth century.

3.1 Subjects versus non-subjects

The first linguistic prediction that follows from the economy hypothesis is that non-subject *du* is retained longer than subject *du*, because avoiding uneconomical inflection is only directly relevant in subject forms. In order to test this hypothesis, we first have to make clear what we consider subjects and non-subjects. The rationale behind this first prediction is that, if there was pressure not to use the T pronoun on grounds of inflectional economy, we will see this effect most strongly in contexts where T actually triggers agreement, namely in subject position. The prediction is that T is used longer in functions where it does not combine with inflection. An alternative way of defining the same prediction is that, with time, there will be an increased preference for subject forms in V pronouns.

To test this hypothesis, we need to define all functions of T and need to assess whether a function triggers second person singular inflection. We list these functions in the table in (1) we will address each category below.

(1)   Functions of T

	Triggers 2s inflection
a. Subject du	+
b. Vocative	-
c. Part of plural reference	-
d. Comparison	-
e. Ellipsis	-
f. Object	-
g. Possessive marker	-

Walker (2005: 270-277) considers all nominative forms of the pronoun T as subjects. In contrast with Walker, we adhere to a more narrow definition whereby only T-forms which trigger second person singular inflection are considered subjects. In (2), -*tu* triggers second person singular inflection (*bis(t)* and *begeers(t)*) and is therefore a subject.

(2)   Wie   bistu   en   wat   begeerstu
         Who   art thou   and   what   desirest thou
         Who are you and what do you desire?
         (Vier Heemkinderen, 1508)

The examples in (3)-(6) that are printed in bold all show instances of *du* that Walker (2005) would classify as nominatives (and therefore as subjects) but which we classify as non-subjects.
        In (3), we present an example of vocative *du.* The vocative *du* does not trigger inflection. Thus, we consider it a non-subject.

(3)   O   siel   **du**   suyuer   bloem
         O   soul   thou   pure   flower
         'O soul, you pure flower'
         (Devoot ende profitelyck boecxken, 1539)

Item c in (1) is what we term 'part of plural reference.' An example of this type construction is presented in (4). In this particular case, *du* is in subject position. Since *du* is part of a coordinated subject in combination with *dine kinder* ('thine

children'), it triggers plural agreement. Because avoiding second person singular inflection is irrelevant in plural references, we consider these uses of *du* and *gi* as part of a coordinated subject as non-subjects.

(4)   **du**   ende   dine   kinder   sult   dod   sijn
       thou   and   thine   children   shalt   dead   be
       'you and your children will be dead'
       (Rijmbijbel, 1275-1300)

In (5), *du* is used as part of a comparison (item d in (1)) and does not trigger agreement. Because the lack of agreement, we consider instances such as (5) non-subjects.

(5)   dat   nieman en   weet   dan   **du**
       that   nobody NEG   knows   than   thou
       'that nobody knows it except you'
       (Nederrijns Moraalboek, 1270-1290)

In (6), we present an example of *ellipsis* (item e (1)). The pronoun *du* is used in an elliptic utterance that does not contain a finite verb. Since the finite verb is absent in combination with *du*, we have considered instances such as (6) non-subjects.

(6)   soutstu   mi   voeden, neen   **du**   niet
       should-thou   me   feed,   no   thou   not
       'would you feed me, no not you'
       (Esopet, 1215)

(7) and (8) are examples of non-nominative T which both Walker and ourselves consider non-subjects. In (7), we see the use of an object T *di* and in (8), two instances of possessive T.

(7)   iet   dat   **di**   quaet   doet
         something   that   thy   ill   does
         'something that harms you'
         (De boec van Catone, 1290-1300)

(8)   als **dijn** wijf   tonrecte claghet   ouer   **dine**   cnechte
         if thy wife   unjustly complains   over   thy   servants
         'if your wife unjustly complains about your servants'
         (De boec van Catone, 1290-1300)

All form and spelling variants of subjects and non-subjects that were used as search strings are listed in Appendix C.

A comparison of the number of subject forms of *du* and non-subject functions of *du* alone, does not tell us enough about our first prediction. As expected, we might find that the percentage of *du* subjects is higher in the sixteenth century than in the thirteenth century. This result is not necessarily the effect of an economy strategy. It could also be the case that our texts in the sixteenth century contain more non-subjects than subjects in comparison with the thirteenth century. If this were the case, the overrepresentation of subjects would not reflect inflectional economy, but instead a general preference for non-subjects in sixteenth century texts.

In order to control for this, we contrast the proportion of subject and non-subject functions in the V-pronoun with the proportion of subject and non-subject functions in the T-pronoun. The V-pronouns are classified in the same manner as the T-pronouns (see (1)). We do not distinguish V-forms with singular and plural reference. Since we are interested in the relationship between subject versus non-subject use in address forms, it is unnecessary to exclude other plural uses of V. We expect a similar subject versus non-subject ratio whether in singular or plural contexts. The only case where plural V is treated differently from non-plural V is in the case of coordinated subjects as in (4). As was the case for T-pronouns, uses of V pronouns as coordinated subjects are also considered non-subjects.

It is conceivable that a change in the ratio of subjects versus non-subjects in the T-pronoun could also be motivated by a pragmatic strategy. Dury (2005) argues that polite language correlates with an increased use of non-subject forms. Second person subjects are more face threatening than non-subjects. Therefore, subjects are avoided in polite contexts. If it is true that polite language triggers non-subject forms, this means that it is more difficult to prove a preference for non-subject forms

in the T-forms, since pragmatics motivate an overrepresentation of non-subjects in the V-forms.

3.2 High frequency verbs

The rate of deflection can differ from language to language, but it can also vary within a single language. Coveney (2000) shows that high frequency verbs are more resistant to deflection than low frequency verbs. We therefore expect less effect of the strategy to avoid uneconomical inflection in high frequency verbs than in low frequency verbs.

In order to determine which verbs were high frequency verbs, we calculated token frequencies for each verb used in each corpus. We consider the verbs with the highest 5 token frequencies as high frequency verbs, that is, the five most commonly used verbs.

Since high frequency verbs can resist inflectional pressure longer, we predict that, in the sixteenth century, there will still be a preference for the T-pronouns to combine with the most common 5 verbs. The fact that a high frequency verb frequently co-occurs with T does not necessarily have to be related to economy. The type of situation that triggers T-use might be precisely the type of situation that triggers high-frequency verbs. It is possible, for example, that informal settings push the selection of high frequency verbs. Knowing that modal verbs are among the most frequent, Busse (2002: 221) describes this possible pragmatic effect as follows: "intuitively it sounds convincing that modal verbs, which express modality or states of mind, go together well with the more affective of the two pronouns." Busse (2002) argues that T-pronouns are more likely to co-occur with modal verbs. Modal verbs are very frequent, which means that pragmatic factors led to an overrepresentation of high frequency verbs in combination with T.

If pragmatics support the same tendency as the economy hypothesis, then we expect that the preference for high frequency is stronger in the sixteenth century where both pragmatics and economy push towards the use of high frequency verbs as opposed to the thirteenth century where the preference for high frequency verbs is solely pragmatically motivated.

4   Results

In this section, we look at the results from the corpus search. In 4.1 we test whether the loss of the T pronoun in favour of the V pronoun is more pronounced in subjects than in non-subjects. In 4.2 we investigate whether the loss of the T pronoun is less

pronounced if the pronoun is used in combination with a high-frequency verb rather than with a non-high-frequency verb.

The odds ration is a well-established statistical measure to judge whether an event is more likely to occur in one situation than in another. For purposes here, the odds ratio will help determine if a T-pronoun is more likely than a V-pronoun to occur with high frequency verbs, and to determine if T is more likely than V to occur with subjects. Before we move to the actual data, let us first consider in more detail, what an odds ratio is. To understand odds ratios, we first need to know what odds are. Odds refer to the probability of occurrence for any given event. In a hypothetical corpus, we might find that one of every hundred words are swear words. The odds of finding a swear word is the probability of finding a swear word (10/100=0.1) divided by the probability of finding a non-swearword (90/100=0.9). Thus the odds of finding a swear word is 0.1/0.9=0.11. This is equivalent to dividing the number of swear words by the number of non-swear words (10/90=0.11). The odds can be converted to a percentage: 100 x odds/(odds+1). An odds of 1 corresponds to 50% ((100 x 1)/1+1=50), an odds of 2 to 66.7% ((100 x 2/2+1=66.7)) etc.

The odds ratio compares odds in two different situations in order to see whether one dichotomous (i.e. yes or no) feature is associated with another dichotomous feature. Say that we have a corpus of people with a religious background and a corpus of atheists. If we find that the odds of finding a swear word for the atheist corpus is 0.11 and it is also 0.11 for the people with a religious background, then the odds ratio of finding a swear word in the religious corpus versus the atheists corpus is 0.11/0.11 which is 1. An odds ratio of one means that there is no association, which, in our example, means that swearing is equally likely to occur in atheists as in people with a religious background.

The further away the odds ratio is from one, the stronger the association between the two features (religious background and tendency to swear). For example, if we find that the odds of finding a swear word is 0.11 for the group of people with a religious background, but 0.22 in the group of people with an atheist background, then the odds ratio of finding a swear word is the odds of finding a swear word in the religious group (0.11) divided by the odds of finding a swear word in the atheist group (0.22). If we divide 0.11 by 0.22 this yields 0.5. The odds of finding swear words in the religious group is thus half the odds of finding swear words in the atheist group.

Instead of saying that we are half as likely to find a swear word in the religious group as in the atheist group we can also say that we are twice as likely to find a swear word in the atheist group as in the religious group. Since the odds ratio

0.5 expresses the same strength of association as the odds ratio 2, the odds ratio plots are presented on a logarithmic scale, where 2 is equally distant from 1 as 0.5. The larger the distance is from 1, the stronger the association between the variables.

The odds ratio reflects the strength of an association. To indicate whether this value is statistically different from the value of 1 (i.e., no association), two approaches are available. One approach is to calculate the probability of finding this odds ratio purely by chance. This probability is calculated using the statistical chi-square test with Yates correction for continuity and is traditionally expressed as a p-value. A p-value of less than 0.05 is usually used to indicate that the odds ratio is statistically different from 1, An association is a near-certain if $p < 0.001$. The second approach is to calculate the 95% confidence interval (CI) for the odds ratio. If, for example, an odds ratio equals 4.7 and the 95% confidence interval is between 2.7 and 8.1, the value is unlikely to be lower than 2.7, so the results are incompatible with an odds ratio of 1, which tells us that there is a positive association. The strength of the association is probably close to an odds ratio of 4.7, but could be as low as 2.7 or as high as 8.1. If the value of 1 is outside the 95% confidence interval, then the chi-square calculation will show that $p < 0.05$. It should be noted that the correspondence between the calculation of the confidence interval is a close approximation of the calculation of the p-value based on the chi-square test (cf. Armitage & Berry 1994: 132-141, Bland & Altman 2000).

In some situations, the strength of the association (calculated as odds ratio) between two variables is influenced by a third variable. In statistical terms this effect is referred to as an "interaction". The interaction term is expressed as an odds ratio. The calculation of such an effect involves the calculation of the logistic regression equation, which requires the use of a statistical software package. For our analyses, we used SPSS (version 15).

4.1 Subjects versus non-subjects

We predicted that subjects would show more T-loss than non-subjects, because only subjects trigger agreement. Avoiding uneconomical inflection is one motivation to choose V rather than T. In this section, we will test the prediction that non-subjects show less loss of T than subjects. We will first consider the raw data and then move to the statistics to find out whether the observed differences are statistically significant.

If we compare the table in (9) (concerning the use of second person pronoun in the thirteenth century) with table (12) (concerning the use of second person pronouns in the sixteenth century), we observe two tendencies. First of all,

we can observe that the percentage of T-forms has decreased over time. Whereas T is attested 22% of the time in the thirteenth century, this number decreases to 3% in the sixteenth century. Secondly, we see that the proportion of subjects has decreased in the context of T. Whereas we find 50% subjects in the thirteenth century in the context of T, we find only 13% subjects in T in the sixteenth century. The percentage of subjects in V-pronouns increases slightly.

In (10) and (11), the data are divided for rhyme and non-rhyme for the thirteenth century and in (13) and (14), for the sixteenth century. The rhyming and the non-rhyming texts show the same two tendencies: a decrease in T over time and an overrepresentation of subjects in the context of T in the sixteenth century. A specification of the use of subjects and non-subjects per text can be found in appendix D.

(9)   Thirteenth century rhyming and non-rhyming combined

	Subject	Non-subject	Subj + Non-subj
Du	1155 (50%)	1172 (50%)	2327 (22%)
Gi	3177 (39%)	4884 (61%)	8061 (78%)
Du +Gi	4332 (42%)	6056 (58%)	10388

(10)   Thirteenth century rhyming corpus

	Subject	Non-subject	Subj + Non-subj
Du	863 (47%)	975 (53%)	1838 (19%)
Gi	3062 (40%)	4661 (60%)	7723 (81%)
Du +Gi	3925 (41%)	5636 (59%)	9561

(11)   Thirteenth century non-rhyming corpus

	Subject	Non-subject	Subj + Non-subj
Du	292 (60%)	197 (40%)	489 (60%)
Gi	115 (34%)	223 (66%)	338 (40%)
Du +Gi	407 (49 %)	420 (51%)	827

(12)   Sixteenth century rhyming and non-rhyming combined

	Subject	Non-subject	Subj + Non-subj
Du	55 (13%)	371 (87%)	426 (3%)
Gi	5643 (41%)	7976 (59%)	13619 (97%)
Du +Gi	5698 (40%)	8347 (60%)	14045

(13)   Sixteenth century rhyming corpus subjects versus non-subjects

	Subject	Non-subject	Subj + Non-subj
Du	36 (14%)	255 (86%)	291 (4%)
Gi	2614 (39%)	3955 (61%)	6569 (96%)
Du +Gi	2650 (39%)	4210 (61%)	6860

(14)   Sixteenth century non-rhyming corpus with subjects versus non-subjects

	Subject	Non-subject	Subj + Non-subj
Du	19 (14%)	116 (86%)	135 (2%)
Gi	3029 (43%)	4021 (57%)	7050 (98%)
Du +Gi	3048 (42%)	4137 (58%)	7185

Now that we have considered the raw data, the question is whether our results are statistically significant. We will look first at the decrease of T from the thirteenth century to the sixteenth century. It is a well known fact that T-use has decreased over this time period in Dutch. By considering this well known change first, we can become more familiar with odds ratios.

In our complete dataset, we found 24,433 second person pronouns, of which 2,753 were T-pronouns. The odds of finding a T-pronoun in the complete corpus are 0.127. The odds can be converted to a percentage: 100 x odds/(odds+1). An odds of 0.127 thus corresponds to (100 x 0.127)/1.127 which yields 11.3%. We thus have approximately 11% chance of finding T in the complete corpus.

The odds of finding T in the thirteenth century (odds 0.29) are much higher than in the sixteenth century (odds 0.03). Expressed as an odds ratio, this loss is 0.03 (odds of finding T in the sixteenth century) divided by 0.29 (odds of finding T in the thirteenth century) which yields 0.108 (95% confidence interval: 0.097 - 0.121). The

odds ratio of 0.108 indicates a strong negative association between T and the sixteenth century ($p < 0.0001$).
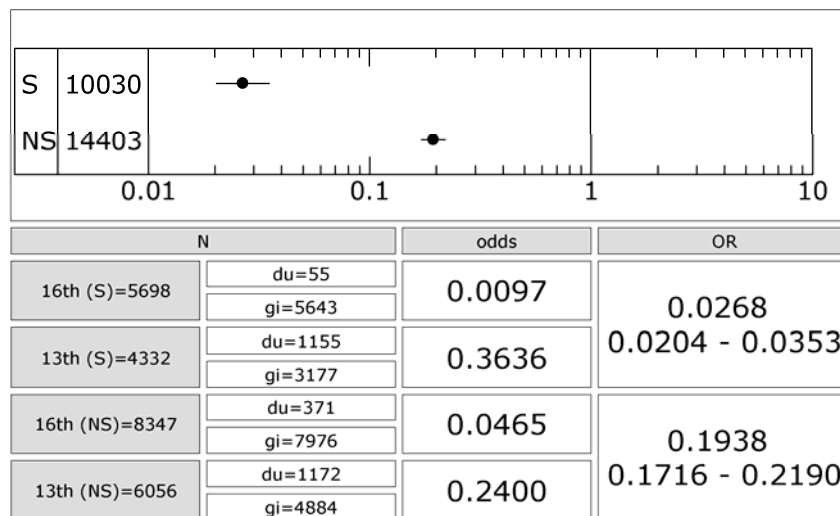
In the top panel in figure 1, the odds ratio, here abbreviated as 'OR' of finding *du* (rather than *gi*) in the sixteenth century (rather than in the thirteenth century) is shown as an odds ratio (the dot) and its 95% confidence interval (the small line through the dot). In the second panel, the actual data are shown. The OR is 0.1084. Because this value is smaller than 1, we can conclude that the odds of finding *du* in the sixteenth century is substantially smaller than the odds of finding du in the thirteenth century.

		0.05	0.1	0.2	0.5	1	2
total	24433						

N		odds	OR
16th=14045	du=426	0.0313	0.1084
	gi=13619		0.0974 - 0.1206
13th=10388	du=2327	0.2887	
	gi=8061		
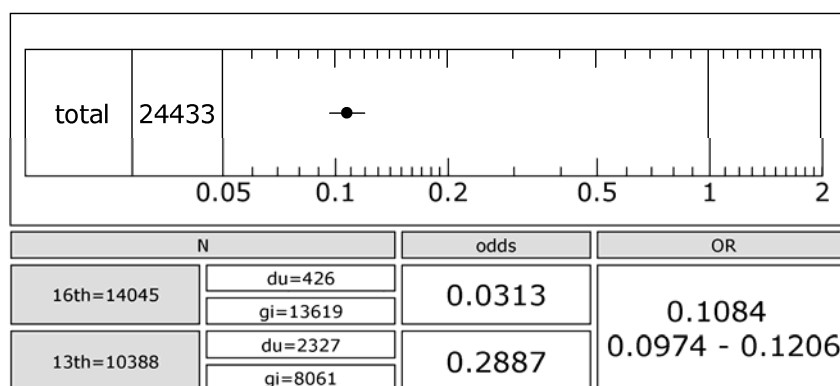
Figure 1

Just like figure 1, figure 2 shows the OR of finding *du* (rather than *gi*) in the sixteenth century (rather than in the thirteenth century). In figure 2, however, the corpus is split into two groups: subjects and non-subjects. Because only subjects trigger agreement, we predicted that the use of T decreases more in the context of subjects than it does in the context of non-subjects. This means that, in the sixteenth century, we expect to find a stronger association with T-loss with subjects than with non-subjects. In figure 2, we observe that the OR of finding *du* (rather than *gi*) in the sixteenth century (rather than the thirteenth century) is indeed lower in the context of subjects than in non-subjects. The OR for subjects is 0.0268 ($p < 0.0001$) and the OR for non-subjects is 0.1938 ($p < 0.0001$). Both OR's indicate a negative association between T and the sixteenth century, but, as predicted, this negative association is stronger in subjects than in non-subjects.

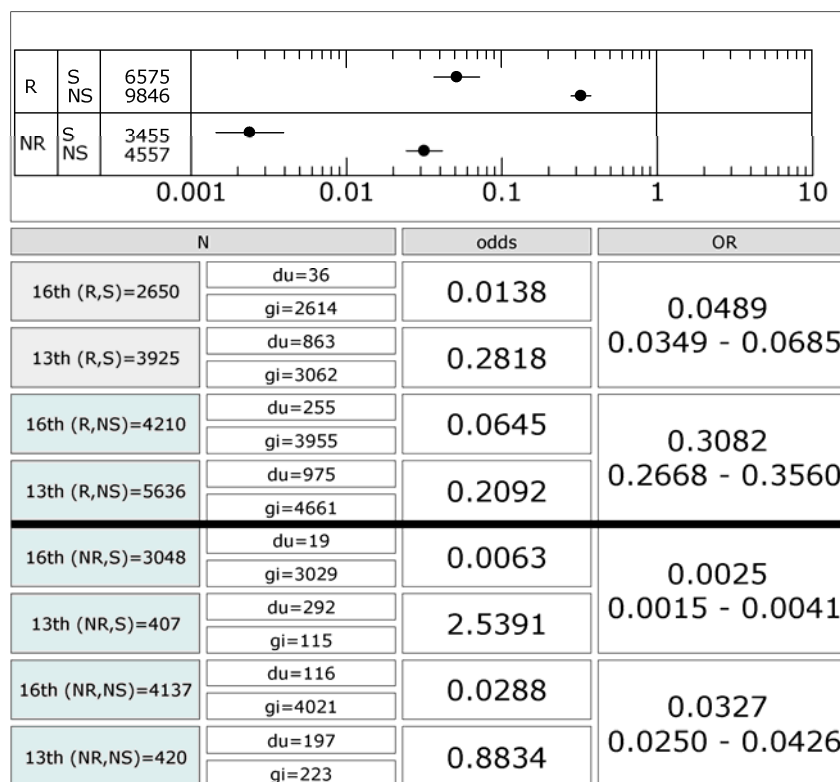|	N		odds	OR
16th (S)=5698	du=55		0.0097	0.0268
	gi=5643			0.0204 - 0.0353
13th (S)=4332	du=1155		0.3636	
	gi=3177			
16th (NS)=8347	du=371		0.0465	0.1938
	gi=7976			0.1716 - 0.2190
13th (NS)=6056	du=1172		0.2400	
	gi=4884			

Figure 2

In figure 2, we observe that the OR of finding *du* in the sixteenth century is lower in subjects than in non-subjects. Since the confidence intervals of the two OR's in subjects and non-subjects do not overlap, we can assume that the difference between these two OR's is statistically significant. With logistic regression, we can make this assumption more precise. We can calculate the difference between the OR of finding *du* in the sixteenth century in subjects versus non-subjects and we can calculate the statistical significance of the difference between these two OR's.

We can understand logistic regression as follows. The difference between the OR of finding *du* in the sixteenth century in subjects versus non-subjects can be expressed as a ratio (0.1938/0.0268). If there is no association between the variables 'non-subject' and 'sixteenth century' the association between T-loss and the sixteenth century would be equally strong in subjects and non-subjects. Dividing two identical numbers yields 1. An outcome of 1 would indicate that there is no association between the two variables 'sixteenth century' and 'non-subject'. The actual difference we calculate for the two OR's is 7.2 (0.1938/0.0268). This means that the OR of T is 7.2 times larger in non-subjects than in subjects. In other words, the strength of the association between T and the sixteenth century is 7.2 stronger in non-subjects than in subjects.

So far, we have considered only three variables, namely '*du*/*gi*', 'thirteenth/sixteenth' and 'subject/non-subject. When a fourth variable ('rhyme') is included, the calculation becomes more complex, but the principle remains the same: We want to know whether the difference between the OR of finding *du* in the sixteenth century in subjects and non-subjects is statistically significant.

In figure 3 the data are split into a subgroup for rhyming texts and non-rhyming texts. From figure 3, it is evident the OR is lower in subjects than it is in non-subjects in the rhyming and the non-rhyming corpora. The OR's of finding *du* in the sixteenth century in subjects and non-subjects are not identical for the rhyming and the non-rhyming corpus because there are interaction effects between the variables 'rhyme' and 'thirteenth century' and interaction effects between the variables 'rhyme' and 'non-subjects'. What the logistic regression analysis does is calculate the probability of finding T after controlling for these other associated effects. Logistic regression with all four variables yields an adjusted OR of 8.06 (CI 5.88-11.05, $p < 0.0001$), which means that the association between T and the sixteenth century in non-subjects is 8.06 times stronger than in non-subjects after controlling for interaction effects. Its high value indicates a strong, highly significant, conservative property of non-subjects with regards to the loss of the T pronoun.

		N		odds	OR
R	S	6575			
	NS	9846			
NR	S	3455			
	NS	4557			

0.001　　　0.01　　　0.1　　　1　　　10

N		odds	OR
16th (R,S)=2650	du=36	0.0138	0.0489
	gi=2614		0.0349 - 0.0685
13th (R,S)=3925	du=863	0.2818	
	gi=3062		
16th (R,NS)=4210	du=255	0.0645	0.3082
	gi=3955		0.2668 - 0.3560
13th (R,NS)=5636	du=975	0.2092	
	gi=4661		
16th (NR,S)=3048	du=19	0.0063	0.0025
	gi=3029		0.0015 - 0.0041
13th (NR,S)=407	du=292	2.5391	
	gi=115		
16th (NR,NS)=4137	du=116	0.0288	0.0327
	gi=4021		0.0250 - 0.0426
13th (NR,NS)=420	du=197	0.8834	
	gi=223		

Figure 3

In the rhyme subcorpus we found a more pronounced drop in the use of the T pronoun than in the non-rhyme subcorpus. We therefore had some concern that the T-conserving property of the non-subject form might not apply in general, but might be restricted to rhyme subcorpus. However, both subcorpora showed the T-conservative property of non-subjects. In both subcorpora the interaction term '16 x NS' shows a positive value distant from one, namely 6.31 (CI 4.37-9.10, p < 0.0001) in the non-rhyming corpus and 13.22 (CI 7.5-23.3, p < 0.0001) in the rhyming corpus. This indicates that in both corpora there is a statistically significant protective effect against T-loss for non-subjects.

Our data thus confirm the prediction that follows from the economy hypothesis: T is lost more in the context of subjects than in the context of non-subjects. This tendency holds for the rhyming corpus and the non-rhyming corpus,

and the combined corpus. The question is whether the economy hypothesis is the only explanation for our results. Could other mechanisms explain the observation that T is overrepresented in the subject forms in the thirteenth century and underrepresented in the sixteenth century?

Let us consider an alternative explanation for our observations. Dury (2005) claims that non-subject pronouns are politer then subject pronouns. Honegger (2003) claims that politeness can be perceived as the result of the combination of multiple strategies. We observed that the T-pronoun gained a more negative association in the sixteenth century. It could be the case that some T forms are more face threatening than other forms. For example, subject-T might be considered more face threatening than possessive-T. If speakers attempt to maintain a minimum level of politeness, they might want to avoid the very impolite T-subject combination.

This alternative explanation is conceivable, but not supported by the data. Many non-subject forms in the use of T are vocative and not indirect objects. Whereas indirect objects might be considered less face threatening (because they occur less frequently in non-sentence initial position) there is nothing about vocatives that makes them less impolite or less face threatening than subjects. On the contrary, the offensive intention of vocatives is often strengthened by combining vocative *du* with offensive epithets like *vuyl* ('dirty'), *valsch* ('mean') and *quade* ('evil').

There are two reasons suggesting against the notion that vocative T combines with subject V forms as an attempt to maintain a minimal level of politeness: First the fact that vocatives are no less face threatening than subjects and second, because vocatives are used in very offensive contexts. There is no minimum level of politeness to be maintained and the preferred sentence position of vocatives and subjects is equally face threatening. We can better understand the observation that we find T-vocatives and V-subjects from the perspective of inflectional economy. Vocatives do not trigger inflection and therefore, T can be used. Subjects do trigger inflection and therefore, the pronoun that triggers the most economical inflection, (V) is preferred.

In chapter 4, we already hypothesized that T was retained longest in those contexts where T adds information, and that T is likely overrepresented in emotional contexts. This hypothesis is supported by the data. All vocatives that were used in emotional contexts or are used in religious contexts. Of the 20 vocatives located in the sixteenth century non-rhyming corpus, 18 vocatives are used to express negative emotions. There are two instances where vocatives were used in religious contexts. In the sixteenth century subcorpus of rhyming texts, we find two instances of vocative *du*. One of these instances is religious (*o siel du suyuer bloem*

'o soul thou pure flower') in *Devoot ende Profitelyck Boecxken* ('Devout and Benifcial Little Book') and the other is motivated by intimate positive emotions, namely *du schoonste wijf, ghi verblijt dat herte mijn* ('thou most beautiful woman you make my heart happy') from *Antwerps Liedboek* ('Songbook from Antwerp').

It is interesting to note that in 8 out of the 20 instances, vocative *du* combines with subject *gi*. An example is given in (15).

(15) **Du** arghe theve **ghi** liechter aen
Thou horrible bitch you lie there of
'You horrible bitch you are lying about it'
(Historie van Margarieta van Lymborch, 1516)

There is one example of vocative *du* that combines with subject *du*, found in *Merlijn.* This sentence is given in (16).

(16) **du** vuyl wicht gaet van
thou dirty girl go from

ons **du** bist een kint onwettelijc
us thou art a child unlawful

'you dirty girl, go away from us you are an unlawful child'
(Historie van den wonderlicken Merlijn,1530-1550)

In the remainder of the 11 cases of vocative *du*, a second person finite verb is absent. An example is given in (17).

(17) Die hertoginne seyde **du** valsche verrader
The duchess said thou mean traitor

dat is ghelogen ende si sloech den grave
that is lied and she hit the count

'The duchess said: you mean traitor those are lies and she hit the count.'
(Historie van Margarieta van Lymborch, 1516)

Given the frequently observed negative epithets like *vuyl* ('dirty') and *valsch* ('mean') used in combination with vocative *du,* it does not seem that speakers are attempting to avoid being too polite. This suggests against Honegger's proposal. Instances like (15) (where vocative *du* is combined with subject *gi*), follow naturally from the economy hypothesis: Vocative *du* is used to express the offensive nature of the utterance. Subject *gi* is not used for pragmatic reasons, but instead, only as a means to avoid the inflectional marker that would otherwise combine with subject *du*. It is important to note that the reverse case (where vocative *gi* occurs in a sentence with second person singular inflection) is never attested. We do have one example of a *gi*-vocative with a *du*-subject in the text *Merlijn* as shown in (18).

(18)	**ghi**	heer	rechter	**du**	sijt	niet	goet
	you	lord	judge	thou	are	not	good

ghenoech	mijn	moeder	te	verdoemen
enough	my	mother	to	condemn

 'You lord judge you are not good enough to condemn my mother'
(Historie van den wonderlicken Merlijn, 1530-1550)

The section from *Merlijn* is noteworthy because the verb seems to agree with the vocative rather than with the subject pronoun. (Notice how the plural form *sijt* ('are') is chosen rather than the expected singular *bist* ('art').) There is only one other example in the corpus where *du* combines with plural inflection rather than singular inflection. This example occurs in *Antwerps Liedboek* ('Songbook from Antwerp') where we find T combining with *hebt* ('have') rather than *hebst* ('hast'). The observation that we find instances of vocative T in combination with second person plural inflection and never the reverse is in line with the hypothesis that inflectional economy plays a role in the under-representation of non-subjects. Since the uses of vocative T are offensive, we cannot understand the absence of subject T in combination with vocative T as a pragmatic strategy to maintain a minimal level of politeness. The utterances combine with offensive epithets and politeness seems irrelevant in these contexts.

A second conceivable alternative explanation for the overrepresentation of non-subjects in the sixteenth century is that T in non-subject positions is preferred on stylistic grounds. According to this explanation, there are more stylistic benefits in the use of non-subjects than of subjects. First of all the non-subject forms of first and second person are minimal pairs. The first person object form *mi* ('me') rhymes

with second person singular *di* ('thy'). Similarly, the first person object form *mijn* ('mine') and the second person singular *thine* ('thine') are also minimal pairs. In contrast to the non-subjects forms, the subject pronouns of first and second person singular *ic* ('I') and *du* ('thou') are not minimal pairs. A second argument for the stylistic preference for non-subject T is that, in contrast to subjects, we often find non-subjects in sentence final position. The sentence final position is associated with rhyming pairs in sixteenth century Dutch poetry. The use of words like *fijn* ('fine') and *wijn* ('wine') in sentence final position can trigger possessive *dijn* ('thine') in the following sentence.

There are indeed examples of the use of non-subject T motivated on stylistic grounds. The excerpts in (19) through (21) illustrate examples of these uses. In (19), *dijn vader* ('thine father') forms a chiasmus with *sone mijn* ('son mine'). The choice for *dijn* ('thine') rather than *uw* ('yours') strengthens the relation between the two items *dijn vader* and *sone mijn.* In (20), *dijn* ('thine') is chosen to rhyme with *sijn* ('to be'). In (21), the object form *di* ('thee') combines well with the object form *mi* ('me'). The set combination *Ick beuele mi di* ('I command myself to you') is the final sentence of the refrain of the song *God gruet v suuer bloem* ('God greets you pure flower'). This song is included in two songbooks in the rhyming corpus, namely *Devoot ende Profitelyck Boecxken* ('Devout and Benifcial Little Book') and *Suverlijc Boecxken* ('Immaculate Little Book'). In total, *Ick beuele mi di* occurs 14 times in the corpus from the sixteenth century. *Dus blijve ik dijn* ('thus I remain yours') also occurs four times in the corpus.

(19)	Was Hillebrant	**dijn**	**vader**,	so bistu		die	**sone mijn**
	Was Hillebrant	thine	father	so art thou		that	son mine

'If Hillebrant was your father you are my son. '
(Antwerps Liedboek, 1544)

(20)	Goet	en	getrouwe		sal	ik	u	**sijn**
	Good	and	faithful		shall	I	you	be

.... Wanneer	ic	ligghe	in	den	arm	van	**dijn**
.... When	I	lie	in	the	arm	of thee	

'I shall be good and faithful to you… when I lie in your arm(s)'
(Antwerps Liedboek, 1544)

(21)   Ick   beuele   mi   di
        I   command   me   thee
        'I command myself to you'
        (Devoot ende Profitelyck Boecxken, 1539 and Suverlijc Boecxken, 1508)

Of the total 367 non-subject T-forms, 90 can be attributed to stylistic consideration. In the rhyming corpus (where we attested 36 subjects and 251 non-subjects), 46 occurances of non-subject T can be attributed to stylistic consideration. 41 can be attributed to chiasmus and parallelism. In the rhyming corpus, 35% of the overrepresentation of non-subjects in T might be attributed to stylistic considerations. The remaining 174 items do not reflect authors' stylistic consideration. In the non-rhyming corpus, we found 19 subjects and 116 non-subjects. Only 3 of the total 116 non-subjects in the non-rhyming corpus can be understood on the basis of stylistic grounds. If we disregard these examples motivated on stylistic grounds, T non-subjects are still strongly overrepresented. Style can thus explain a part of the overrepresentation of non-subjects in the rhyming corpus, but surely not all.

To conclude, we have seen that T use has decreased over time. T-loss is stronger in subjects than in non-subjects (OR 8.06, CI 5.88-11.05, p < 0.0001). The observation that subjects are more affected by T-loss than non-subjects is observed both in the rhyming and in the non-rhyming corpus. This outcome is exactly the outcome that follows from the economy hypothesis, namely that T-subjects need to be avoided in the sixteenth century because in the sixteenth century, second person singular inflection is under pressure. Since subjects trigger inflection, it follows that subjects are avoided the most.

The hypothesis that the use of *gi* is not always pragmatically motivated is further supported by examples such as (15), where there seems to be no motivation for the choice for *gi* in the context. The only reason we do find vocative *du* and subject *gi* in sentences like (15) is inflection: Vocatives do not trigger inflection and thus the T-pronoun is allowed in the vocative. Subjects do trigger agreement and therefore subject *gi* is used.

Our primary data confirm the hypothesis that subject T was lost before non-subject T. There is additional evidence from secondary sources that the inflection that combined with the pronoun T played a role in its loss. Zwaan (1939: 130) cites the grammarian De Hubert who includes the non-nominative forms of the pronoun *du*, namely *dijn* ('thine') and *dij* ('thee') in his 1624 grammar of Dutch. He also excludes subject *du* because of *de hardigheid des gevolgs* ('the roughness of what

follows'). With 'des gevolgs' De Hubert refers to the inflectional marker that combines with *du*.

There is no pragmatic reason why the use of a second person singular inflectional marker would be considered especially unpleasant when compared to a second person singular pronoun. We are able to understand the inflectional marker's strong negative connotations from the perspective of the economy hypothesis. Since second person singular inflection is under pressure, this form is least used. The T-subject and agreement marker remain in use only when there is strong pragmatic need. The use of second person singular inflection can be triggered by anger. In this context, the inflectional marker will also gain negative associations.

Several authors have commented on the problematic status of second person singular agreement. Among them are the grammarian De Hubert, and advisors for the translation of the *Statenbijbel* ('Dutch Authorized Version [of the bible, S.A.]'). In 1618, thirty-seven advisors from different European countries gathered to decide about important translation issues for the new version of the bible. Accounts of the gatherings are based on minutes of the gatherings, themselves, as well as on diaries, reports, and letters. These texts indicate that participants discussed what the correct address term for God should be. Most non-Dutch advisors preferred the pronoun T because it mirrored the use in Latin and because a real singular pronoun T was a more proper term of address than the originally plural pronoun V. Some Dutch advisors, however, noted that it was inappropriate the use of T as an address form to God. Participants cited inflection as one argument against the use of plural V. One argument they presented in favour of plural V had to do with inflection. Specifically, they claimed that inflection in combination with the T-pronoun was no longer in use.[1]

The prediction that inflection in combination with *du* was under pressure is therefore confirmed by our primary data. We find extra support for this prediction in the remarks of translators and a grammarian from the early seventeenth century who report that inflection in combination with *du* is problematic. An overview of the data in each text concerning the subject versus non-subject ratio in T-pronouns and in V-pronouns can be found in appendix D.

---

[1] Exoleverunt etiam compositiones, quae solent addi voci du: du hebst, du solst.' (Sibelius 1618 in Kaajan 1914: 118 fn.2) ('The compositions that are added to the word *du hebst, du solst* have gone out of use.')

4.2 High frequency verbs

We expected an overrepresentation of high frequency verbs in the T-pronoun in the thirteenth century. We predicted an even stronger overrepresentation of high frequency verbs in combination with T-forms in the sixteenth century, because, in the sixteenth century, the preference for high frequency verbs in the T-pronoun is pushed both by pragmatic factors and by economy: high frequency verbs resist deflection longer. An overview of all finite verbs in combination with T and V can be found in appendix E.

In (22), we present the raw data. The question of whether T combines with high frequency verbs more than V is only relevant for subjects. We have seen in the previous section that T-loss is strongest in the group of non-subjects. Since we are looking only at a subset of the data (subjects) which is underrepresented in the sixteenth century, I decided to collapse the rhyming and the non-rhyming corpus into one single corpus.

The raw data support both of our predictions. In the thirteenth century, we already find that T combines more frequently with high frequency verbs (61%) than V (55%). This difference is more pronounced in the sixteenth century, where we find that 79% of the T-pronouns combine with high frequency verbs as opposed to 59% of the V-pronouns.

(22)   Percentage of high frequency verbs in combination with *du* and *gi*

	High	Non-high	Total verbs
13 DU	706 (61%)	449(39%)	1155
16 DU	44 (79%)	12 (21%)	56
13 GI	1755 (55%)	1422 (45%)	3177
16 GI	3316 (59%)	2326 (41%)	5642

The raw data are thus in line with our predictions. We will now look into whether these data are statistically significant. When we stratified for combination of the pronoun with high (N=5806) or low (N=4224) frequency verbs, we found that, as predicted, high-frequency verbs partially resisted the loss of the T pronoun (see figure 4).
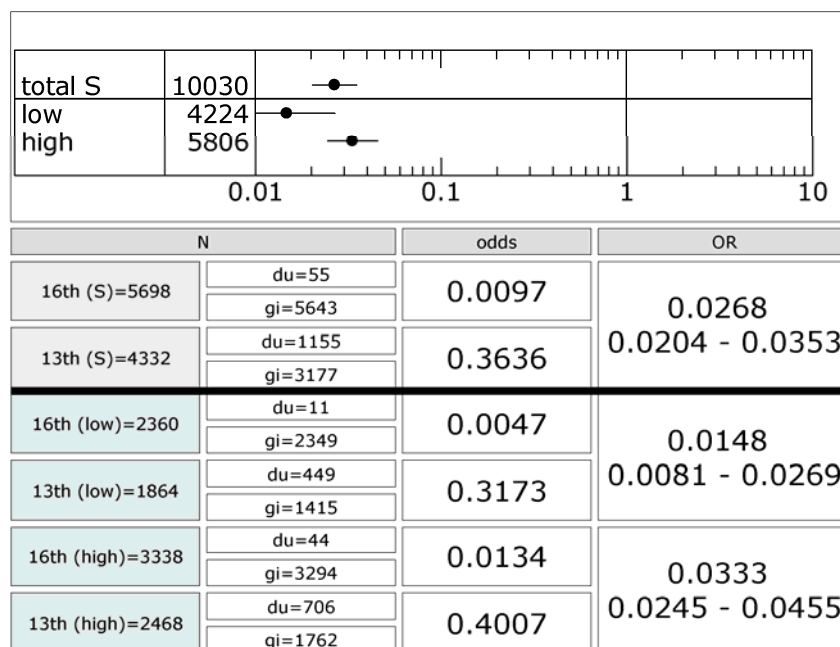
		odds	OR
total S	10030		
low	4224		
high	5806		

N		odds	OR
16th (S)=5698	du=55	0.0097	0.0268
	gi=5643		0.0204 - 0.0353
13th (S)=4332	du=1155	0.3636	
	gi=3177		
16th (low)=2360	du=11	0.0047	0.0148
	gi=2349		0.0081 - 0.0269
13th (low)=1864	du=449	0.3173	
	gi=1415		
16th (high)=3338	du=44	0.0134	0.0333
	gi=3294		0.0245 - 0.0455
13th (high)=2468	du=706	0.4007	
	gi=1762		

Figure 4

In order to establish the statistical significance of the difference between the OR's (between high frequency verbs and low frequency verbs), we performed a logistic regression analysis. The odds ratio of the interaction term (sixteenth century x high frequency) was positive: 2.26 (1.15-4.45), with a p-value < 0.02. 2.26 indicates the probability of finding *du* with a high frequency verb rather than a low frequency verb in the sixteenth century. Its positive value indicates the conservative property of T in high frequency verbs as opposed to the loss of the T pronoun. We found this result whether or not the variable: rhyme, was included (rhyme excluded: OR 2.27 CI 1.15 - 4.47 p <0.02) (rhyme included: OR=2.26, CI 1.15 - 4.45, p<0.02) This indicates that the decision to coalesce the subcorpora (rhyme and non-rhyme) is statistically justified. Our second prediction: that T combines longer with high frequency verbs than with low frequency verbs, is also confirmed by the data.

5   **The economy hypothesis and English**

We have looked at Dutch data. We predict that the economy hypothesis is applicable for both Dutch and English. Thus, we would expect similar results on English data. In this section, I discuss studies on the competition between the forms *thou* and *ye/you* in English. We will use a similar methodology and analysis as we used for the Dutch data.

5.1 Subjects versus non-subjects

The proportion of subjects versus non-subjects in T-forms and V-forms in English is discussed by Walker (2005, 2007) and by Busse (2002). As discussed in section 3, one difference between our study and their studies is the definition of the category 'subject'. In the work by Busse (2002) and Walker (2005, 2007), vocatives are considered subjects. Only subjects trigger agreement. From the perspective of the economy hypothesis, categorizing vocatives as subjects contaminates the data. Despite this inconsistency, the results of the Dutch data and the English data are still similar. Busse (2002: 260) reports on earlier work by Bock (Bock 1938: 65 appendix 2), where subjects and objects are compared in English drama texts dating from 1497 to 1775. Bock's data show that after 1633, the percentage of T-forms is higher in objects than in subjects. For example, the text *The Gamester* (1633) includes 22% T-forms in the subjects and 47% T-forms in objects. In all texts published after 1633, there is a higher percentage of T-objects than T-subjects. The preference for T-objects is clear in all texts of 1633, except for one text: *Way of the World*, where 13% T-subjects are attested and only 10% T-objects.

Walker (2005, 2007) looked at three different corpora: drama texts, trials and depositions. In the drama corpus, she finds that, from 1640 onwards, there is a slight overrepresentation of T in objects (Walker 2005: 271). The same is true for the trials (Walker 2005: 271). In the depositions, however, there are more T-subjects than T-objects, even in the texts from the period between 1720-1760 (Walker 2005: 272). At first sight, these results seem to provide evidence against the economy hypothesis. But if we look more closely at the regional origin of the texts from the trial corpus, it becomes clear that the deviant behaviour of the deposition corpus actually provides further support for the economy hypothesis.

Nearly all of the depositions that Walker studied are from Northern England. In (23), we repeat the verbal paradigm of Northern English, which was originally presented in chapter 4. From (23), we can see that, in Northern English, the T-pronoun combines with the same inflection as the V-pronoun.

(23)   Northern English since 1300 (cf. Lass 1992)

	Singular	Plural
1	-(e)	-es
2	-es	-es
3	-es	-es

Replacing T by V does not yield a more economical paradigm in the English from Northern England. If inflectional economy does not play a role in the choice between T and V, the observed preference for subjects would be expected. In a sense, the Northern English texts are similar to the thirteenth century Dutch texts, where there was not yet any inflectional pressure. Without pressure from the inflectional system, we also observed a slight preference for subjects in the T-pronouns in the thirteenth century Dutch data. As long as there is no pressure from inflection, pragmatics supports a slight preference for T-subjects.

We can conclude that the English data show a pattern similar to the Dutch data. Non-subjects are overrepresented in T-forms in later stages. The overrepresentation of non-subjects in T is predicted by the economy hypothesis. T-subjects trigger second person singular agreement. Avoiding this agreement is only directly relevant in subjects. In an ideal comparison between English and Dutch, an identical categorization of subjects and non-subjects would be used. Additionally, we would only include texts from regions where T was lost and where the replacement of T by V yields a more economical paradigm.

5.2 High frequency verbs

Walker (2005, 2007) and Busse (2002) followed Barber (1981) and Mulholland (1967) who looked into the possibility that T combines with closed class verbs and V combines with lexical verbs. As discussed in 3.1, Busse (2002) and Walker (2005,2007) argue that the tendency for T to combine with closed class verbs (in particular, with modal verbs) is related to pragmatics. Modal verbs express states of mind and states of mind are personal. Therefore, we expect the use of modal verbs

to co-occur with the more personal pronoun T. Neither Walker (2005,2007) nor Busse (2002) found this to be true.

In the Dutch data, we did observe a preference for high frequency verbs. This was not the case for English. We speculate that this difference is related to the operationalization of the prediction. Neither Walker (2005,2007) nor Busse (2002) was interested in the role of inflection in the choice for T or V. The effect of co-ocurrence with certain verbs they predicted was pragmatically oriented and this different motivation led to a somewhat different operationalization of apparently similar predictions. Both Walker (2005,2007) and Busse (2002) opposed closed class verbs to lexical verbs. As we just saw, some closed class verbs occur more frequently than others. We expect that T will only combine with high frequency verbs. If we oppose closed class verbs to lexical verbs, we mix high frequency and non-high frequency verbs in one class, and are less likely to find an effect of co-occurrence with T.

In the depositions Walker (2005: 260-267) also looked at the use of the three most frequently attested finite verbs in combination with second person pronouns. These most frequent verbs are *be*, *shall* and *will*. Walker reported a strong tendency for T to combine with *be*. Walker's (2005: 262) explanation for this tendency differs from the economy hypothesis. Instead of relating the preference for the verb *be* to the tendency for high frequency verbs to resist deflection, she relates the tendency to pragmatics. Namely, Walker argues that the T-pronoun is suitable for defamation, where the verb *be* is frequently used. An example of this use of the verb *to be* is *thou art a whore witche* (Walker 2005: 261). Although Walker's theory fits the English data, it cannot explain the prevalence of the combination of T and the verb *be* in the Dutch data. In the Dutch data, the use of the T-pronoun in combination with the verb *be* can be related to defamation cases in only three out of 24 instances.

At first sight, the Dutch data described in this chapter seem comparable to the English data. Upon further investigation, however, this claim is not entirely true. The different angles from which the data are perceived yield different operationalizations of co-occuring patterns between T and certain verbs. For the Dutch texts, we looked at the effect of frequency. In the English texts, we compared closed class verbs with lexical verbs. The group of closed class verbs does not sufficiently overlap with high frequency verbs in all cases and thus, it becomes impossible to compare the data. A second conclusion is that verbs which frequently occur with second person in the English texts also occur frequently with second person in Dutch texts. In both languages, there is a strong preference for T to

combine with the verb *be*, and, interestingly the T-pronoun combines the longest with this verb.

6   Summary

In this chapter, we looked at additional evidence for the economy hypothesis in Dutch historical data. We compared texts from the thirteenth century with texts from the sixteenth century. In the thirteenth century, the choice between T and V is based solely on pragmatics. In the sixteenth century corpus, we predicted an effect of pragmatic factors in the choice between T and V, in addition to an effect of inflectional economy.

We formulated two predictions to investigate the role of inflectional economy in the choice between T and V. First, we predicted that non-subject T would be retained longer than subject T. This prediction was confirmed. The second prediction was that, if a T-subject was used in the sixteenth century, it would usually combine with high frequency verbs, since high frequency verbs resist deflection the longest. This prediction was also borne out.

We also looked into special combinations, for instance, where the vocative T combines with subject V as in: *Du arghe theve ghi liechter aen* ('thou horrible bitch you are lying about it'). This clearly offensive utterance shows no pragmatic motivation for the use of a V-pronoun. The economy hypothesis, however, can easily explain why we find a vocative T-form and a subject V-form. The vocative T-form is selected on pragmatic grounds: vocative T expresses contempt. Since vocatives do not trigger inflection, the T-form can be used. The choice for the subject is affected by inflectional agreement. And, while vocatives do not trigger inflection, subjects do. Inflectional economy pushes the choice for the V-pronoun in subjects, because the V-pronoun combines with more economical inflection than the T-pronoun.

The data presented in this chapter in support of the economy hypothesis was based on historical data. In chapter 6, we will focus on synchronic data based on dialectal variation in Dutch.