1. Introduction

1.1 Central research question: mechanisms of language change

This study takes a detailed look at language change and considers how and why it takes place. The research has been supported by analysis of the decline in the use of unstressed, full vowels in the Frisian language between 1300 and 1550. For example, the word-final /a/ in Old Frisian /sit-ta/ ‘to sit’ > Modern Frisian /sit-ta/. The reasons for choosing this subject are as follows:

- Written Frisian texts from ±1300 until 1550 are particularly suited to this kind of research. They are technically easily accessible in a digital database, they are dated, can be localised and cover a long period of time, making them ideally suited for monitoring language change;
- Diachronic linguistics illustrate how languages develop and evolve, while synchronic approaches underline the communicative validity and systematicness of language at any moment in time. The understanding of the phenomenon ‘language’ may benefit from insight into the mechanisms of language change that respect both the intrinsic dynamics and the synchronic validity.

The arguments behind this study are given in four stages:

- Firstly, the corpus of historical texts is analysed for their suitability for historical phonological research. Most scholars of Frisian are not convinced that spelling different vowels in unstressed syllables (<a>, <e>, <u> and <i>) reflects the different phonemes in the language of the time (§ 3.1). It will be shown that the mediaeval Frisian texts used constitute a reliable and accurate source for this type of research (§ 1.3);
- The second stage is the actual retrieval of linguistic facts from the historical sources. This provides considerable detailed information about Frisian phonology over the period ± 1300 - 1550, with a focus on vowels in unstressed syllables. Within the constraints as outlined at the first stage, this presentation of facts will be theory-free as far as possible (chapter two);
- The third stage is a phonological interpretation of the data. Chapter three presents a traditional phonological analysis of the data. The phonological analysis indicates that some additional phonetic feature may have been prevalent in 15th century Frisian. The character of this feature is discussed separately in chapter four. Arguments are provided for the assumption that 15th century Frisian had two contrasting tone contours, similar to those in modern Norwegian and Swedish;
- In chapter five, two dynamic models of language change are presented, illustrating that language change can be the result of deterministic
processes and as a result can be highly predictable. The detailed reconstruction in chapter two is needed to formulate, calibrate and test the models in chapter five.

The introduction section 1.2 contains:

- A brief introduction into the Frisian language: 1.2.1;
- An explication about the language sources used: 1.2.2;
- A description of language period labels applied: 1.2.3;
- An outline of the main linguistic changes in Frisian during the studied period, as far as these are relevant for the understanding of this subject: 1.2.4;
- Delineation of the main theoretical problems of language change in diachronic and synchronic approaches: 1.2.5.

From § 1.3.2 onwards, almost every section is followed by a section summary, to structure the content and facilitate browsing through the study.
1.2 Introducing the subject

1.2.1 The Frisian Language

Frisian is a West Germanic language, traditionally spoken on the southern littoral of the North Sea in what are now the Netherlands and Germany (map 1.1). Frisian is linguistically divided into three main dialect groups: West, East and North Frisian. Versions from the different groups are mutually barely intelligible. The West Frisian dialects are generally mutually intelligible, so West Frisian constitutes a single-speech community. Within the small region of the North Frisian dialects, mutual understanding is often problematic. There is an old and deep contrast between the North Frisian dialects from the islands and the mainland as a result of different stages of immigration (8th century / 10th century; cf. Sjölin 1969, 3).

Nowadays, East Frisian is only spoken in the municipality of Saterland. Sources ranging from the 17th to the early 20th century are available in Frisian as it was once spoken in the Harlingerland, on the island of Wangerooge and in Land Wursten. The last three dialects share several linguistic features (cf. detailed discussion below). They are sometimes referred to as Weser Frisian. One accessible and reliable source of information about the Frisian language, its history and sociological position, is the Handbook of Frisian Studies (Munské 2001a).

The Frisian language area was more extensive in the past than it is nowadays. The

Map 1.1: Frisian language area around 1200 and present day. The map includes references to regions and places mentioned in the text.
largest Frisian speaking community is currently found in the Dutch province of Fryslân. A recent survey revealed that 54% of the population of Fryslân speak Frisian as their 'mother tongue,' or about 330,000 people (Provincje Fryslân 2007, 9). The number of native speakers of Frisian in Germany is probably no more than 7,000 (cf. Walker 2001, 267; Fort 2001, 410). These days all 'mother tongue' speakers of Frisian are at least bilingual, speaking Frisian and Dutch or Frisian and High German. Frisian is an officially recognised minority language, both in the Netherlands and Germany. Modern West Frisian has a written standard form and has gained some position in the educational system and the media.

In the West Germanic dialectal continuum, ranging from English to High German, Frisian is historically the nearest relative of English. Both languages belong to the so-called group of North Sea Germanic languages. Also, Old Saxon and north-western varieties of Dutch belong to or belonged to this group. North Sea Germanic consists of a rather vague collection of shared linguistic relics and innovations (cf. Nielsen 2001, 512ff). Typical examples of North-Sea Germanic phenomena are:

- Nasal drop before fricatives. For example: Proto-Pan Germanic */gans/ > Proto-North Sea Germanic */gons/ > */gøs/ > Modern English goose, Modern West Frisian goes /gu.øs/, Low Saxon Goos /gos/, contrasting with Dutch gans /yəns/, High German Gans /gans/;
- Palatalisation of Proto-Germanic /k/ before front vowels. For example: Latin căsens > Proto-Pan Germanic */kasjus/ > Old English */Old Frisian /ke:s/ > Modern English cheese, Modern West Frisian tsiis /tsi:s/ contrasting with Low Saxon Kees /ke:s/, Dutch kaas /ka:s/, High German Käse /keːs/.

These historical similarities should not detract from the fact that no variant of Modern Frisian is intelligible alongside any form of Modern English, while, with some effort, Modern West Frisian, for example, is fairly understandable for speakers of Modern Standard Dutch.

As North Sea Germanic languages, Old English and Old Frisian share some features with the North Germanic languages. For example, the Old Nordic word gâs 'goose' with the aforementioned nasal drop before the fricative. A unique Frisian-North Germanic parallel is the accent shift in the products of the Proto-Germanic /eu/. For example: Old Frisian biðda 'to offer', Modern Frisian / West-Terschelling dialect bjœade [bjœ.ədə], Modern Icelandic bjöða, cf. Old English bêdan, Old High
German *biotan* and Modern High German *bieten* [biːtən].\(^1\) Often, these North Sea Germanic parallels date back more than a 1000 years. Separate to this is the impact the Danish language had on the neighbouring North Frisian dialects, since the colonisation of that region by Frisians in the 8th century. The close language contact between North Frisian and Danish lasted until the decline of Danish in Southern Jutland during the 19th century.

This study deals with part of the history of West Frisian, so when reference is made to modern Frisian forms, this will generally be from the standard spelling of Modern West Frisian. Where relevant, examples from other dialects are given. There are three clearly differentiable dialects of Modern West Frisian:\(^2\)

- The dialect of the small town of *Hindeloopen* in the south-west is a rather archaic dialect, more or less ‘petrified’ in the 16th century;
- The dialect of *Terschelling* resembles late 18th century north-western dialects from the mainland;
- The dialect of *Schiermonnikoog* combines many archaic traces from the 16th century with some intensive phonological innovations. The language of the mid-16th century *Proverbs*, written by Reyner Bogerman (De Boer 1900), can be traced back to the region *Dongeradeel* in the far north-east of the mainland. This language form is an almost direct precursor to the modern *Schiermonnikoog* dialect (cf. Spenter 1968, 14).

Typologically, Modern West Frisian resembles Modern Dutch in many aspects. Both are not only geographically but also typologically positioned between English and High German. The consonant system of Modern West Frisian resembles the Dutch system. The vowel system of Modern West Frisian is rich in diphthongs compared to other West Germanic standard languages (Bussmann 2004). Modern West Frisian is a mostly analytical language, with some reflexes of the former synthetic system. Two genders (as in Dutch; English 1, High German 3), two plural formation suffixes -en and -s (as in Dutch; English only -s; High German has more than four categories). Just likes Dutch and English, Frisian lost the historical verbal subjunctive. In the common present tense verbal paradigm, Frisian has four different forms to mark the six different person-number combinations (three persons in singular and in plural). English has two, Dutch three and High German five different forms. Distinct among other West Germanic languages is the

\(^1\) Cf. Århammar 1969 for the complex developments of the Proto-Germanic /eu/ in Frisian.

\(^2\) Interpretation of the dialects by the author.
In the last three to four decades, there has been a heavy tendency to conform to the Dutch pattern.

1.2.2 Main relevant sources

Apart from some scarce runic evidence of Frisian from as early as the 6th century, the oldest attestations to Frisian are from the 12th or 13th centuries. Since the 14th century, there has been a continuous, albeit sometimes limited flow of written West Frisian. The oldest texts are written in a fairly archaic language. This archaic language stage of Frisian is found in parts of two late mediaeval West Frisian codices, one known as Unia (Siebs 1895) and the other as Jus Municipale Frisonum (Buma/Ebel 1977) as well as a somewhat richer text corpus from adjacent regions towards the east, where Frisian was widely spoken at the time (cf. map 1.1).

From the present West Frisian speaking region, about 1,200 charters, which are written in Frisian, have been preserved. Apart from a single charter originating from 1329, they cover the time frame between 1378 and 1550. They have been published in four volumes of the series Oudfriese Oorkonden (OFO I-IV). The attested charters are formal texts and juridical and administrative writings from private and public sources.

The mediaeval charters present us with an excellent source for diachronic language research. The precise date on the majority of the charters offers the opportunity to monitor both the introduction of linguistic innovations and the decline of older forms. The charters cover most of the transition period from archaic mediaeval Frisian to early Modern Frisian. The charters can only be a suitable source for studying language change, when they reflect the actual phonology of the language. This point will be discussed in § 1.3.2. The charters are not only dated, but can also be connected with specific regions of the West Frisian language area. Miedema (1986) and Vries (1986) have shown that the West Frisian charters can be used for historical dialect-geographical research. This specific aspect is discussed in § 1.3.4.

Many of the processes that are subject to investigation in this research, are well covered by the data from the charters. However, some processes started earlier, in the 14th century. The sole charter from 1329 is too limited to provide a proper insight into the 14th century developments. To extend the scope back into the 14th
century, one has to rely on the mediaeval Frisian codices. The most important of these is *Unia*. The text of this codex, written in 1477, has remained unpublished and has only been preserved in a 17th century copy. The codex was lost in the 18th century. The language of *Unia* contains the most archaic language forms from the region of Fryslân. Most of the texts in *Unia* must have received their linguistic form long before 1477. However, because nothing is known about the time the texts received their attested linguistic form, the dating of the linguistic features can only be done by an indirect method. The language of this manuscript is analysed for the purpose of this investigation in § 1.3.8.

The codex *Jus Municipale Frisonum* (abbreviation *Jus*) seems more complicated in its linguistic structure than *Unia*. The preserved codex is from 1530 but the first part of it was probably written in 1464. *Jus* is most likely from the region around Sneek (Johnston 2001, 575). It is written in a language that in some aspects is as old as the late-14th century charters. For example, the use of the letter *<a>* in the unstressed syllables: *<habbane>, <habbath>* instead of Middle Frisian *<habben>* and *<habbet>*. These archaic forms in *Jus* are accompanied by more recent forms. In spelling conventions, the text is clearly 15th century, with explicit marking of long vowels and frequent rendering of Old Frisian /ð/ and /þ/ (spelled *<th>*) by *<d>* and *<t>* (cf. § 1.3.2). The archaic linguistic traces in some sections suggest that parts of the text from 1464 have been copied from an even older codex from the late 14th century, being partly adapted to the language of the 15th century. This causes a complicated linguistic layering. Therefore, the codex is only used occasionally.

Other mediaeval codices from the West Frisian region are linguistically dated from the 15th century, a time sufficiently covered by the charters. They are not considered further here. Incidental reference will be made to some mediaeval Old Frisian codices, marked with the labels B, E1, F, H, R1, R2 (cf. Johnston 2001).

The advances in the data analysis are the result of the availability of the texts in a digital form. This enables data analysis, using database programs and Geographical Information System (GIS) software.

1.2.3 The Periodisation of Frisian and terminology used

The mediaeval attestations to Frisian are chronologically contemporaneous with for example *Middle English* and *Middle Dutch*. Traditionally they are labelled *Old Frisian*. Sjölin (1969, 16-18) showed a contrast between a ‘classical’ and a ‘post-classical’ Old Frisian. Sjölin’s labels build upon both linguistic and spelling differences. The demarcation lies roughly in the year 1400. The label *Mid Frisian* was up till now, only applied to West Frisian language attestations from 1550 to 1800. See Bremmer (2001) for an overview of the discussion in the last century.
The typological labelling of the older attestations have been a topic of discussion between De Haan (2001, 188ff.) and Versloot (2005, 266ff.). Both agree, to abandon the label *Middle Frisian* for post-mediaeval West Frisian language forms and to use it for hitherto ‘post-classical Old Frisian’ forms of mediaeval Frisian, the language of the 15\textsuperscript{th} and early 16\textsuperscript{th} century. De Haan’s opinion that all mediaeval Frisian deserves the label *Middle Frisian* is contested by Versloot, who provides arguments to call the archaic language type from the 14\textsuperscript{th} century and earlier (late-) *Old Frisian*, in spite of its relatively late attestation (Versloot 2005, 288).

In this thesis, *Old Frisian* is the language from the 14\textsuperscript{th} century and earlier. For West Frisian, the archaic stage is represented in the oldest sections of the codex *Unia*. Other parts of *Unia* and a scarce number of late 14\textsuperscript{th} century charters are written in a form of *late-Old Frisian*. Unless stated otherwise, *Old Frisian* applies to Frisian sources from the current province of Fryslân. The language of the time after around 1550 will be referred to as *Modern Frisian*. The term *early-Modern Frisian* will be used for specific forms of the language from the 17\textsuperscript{th} and 18\textsuperscript{th} centuries. In the context of this study, *Modern Frisian* equals *Modern West Frisian*. The adjective ‘West’ will eventually be used to express a contrast with East and North Frisian. Examples taken from the transitional period between Old Frisian and early Modern Frisian will appear under the label *Middle Frisian*. The bulk of the charters can be assigned to this Middle Frisian period (cf. table 1.1):

<table>
<thead>
<tr>
<th>time frame</th>
<th>- ±1300</th>
<th>± 1300 - 1400</th>
<th>± 1400 - 1550</th>
<th>± 1550 - 1750</th>
<th>± 1750 -</th>
</tr>
</thead>
<tbody>
<tr>
<td>in this study</td>
<td>Old Frisian</td>
<td>(late-) Old Frisian</td>
<td>Middle Frisian</td>
<td>(early-) Modern Frisian</td>
<td>Modern Frisian</td>
</tr>
<tr>
<td>current labels</td>
<td>classical Old Frisian</td>
<td>post-classical Old Frisian</td>
<td>Middle Frisian</td>
<td>Modern Frisian</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.1: Period labels for Frisian

As an example of late-Old Frisian, a citation from one of the oldest original charters is given:

> Wi her *Haia curith* to *Tunawerth* and *Webka Reymsma* vuse gamech duath kuth and kanlik alda ghenim ther thit bref herath oftha siath leza .... (OFO I-4, 1390)

4 “We, master H. *curith* in *Ternaard* and W.R., our countryman announce and let know to all those, that hear or see this letter reading...”
The language of the charters from the 16th century is, with some spelling correction, fairly comprehensible for modern speakers of West Frisian.

In modern spelling, this would be:

Wij trije brooren neys[ioun] blye ende bekenne meij dit ienwyrdige scheedbriefs ende quitanje hoe dat wy synt onder malcender accordiert .... (OFO II-374, 1541)

Apart from the temporal differences, there were of course dialectal variations in Old Frisian. A tradition exists to distinguish between Old West Frisian and Old East Frisian (cf. Bremmer 2001 for an overview). The boundary between these two groups was to be found at the eastern limit of the modern West Frisian language area. Within so-called ‘Old East Frisian’, it is the outstanding character of the Old Frisian dialect of the region Riustringen (cf. map 1.1) that is of special interest within the framework of this study (Boutkan, 1996). Its language is found in two codices from the 14th century, labelled R1 and R2. The labels Old West Frisian and Old East Frisian will occasionally be applied as strictly geographical designations. ‘Old West Frisian’ means: Old Frisian language forms as they are attested in sources from the region where nowadays West Frisian dialects are spoken, mutatis mutandis for Old East Frisian. Old West Frisian and Old East Frisian do not necessarily differ from each other, nor are both groups internally necessarily homogeneous.

1.2.4 Main relevant developments in West Frisian between 1300 and 1600

Extensive changes have taken place in the West Frisian language between the late 14th and the early 16th centuries, changes that are also known to have taken place in many other Germanic languages.

5 “We, three brothers mentioned, confess and confirm with this present letter of arbitration and receipt how that we have agreed among each other...”

6 With the sole difference being that nowadays most Modern East Frisian dialects are extinct.
These are the main developments, relevant for this study:

- Abandoning of case system and reducing grammatical gender contrast from three to two genders (Hoekstra 2001b, 776-778);
- Simplification of inflectional morphology (idem);
- Degemination (Versloot 2001b, 774);
- Lengthening of short root vowels in open syllables (Versloot 2001b, 769);
- Reduction of unstressed vowels (idem).

The reduction of unstressed vowels is the main focus of this study. Lengthening of short root vowels in open syllables and degemination affect the conditions for reducing unstressed vowels. The reduction of unstressed vowels has serious implications and is recursively affected by the inflectional system. That is why these processes are introduced here.

Case system and grammatical gender

Old Frisian had four cases in three grammatical genders, as in Standard High German and Modern Icelandic. This system was gradually abandoned during the Middle Frisian period. So, Old Frisian distinguished for example:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>thi stên</td>
<td>thes stênis</td>
<td>tha stêne</td>
<td>thenne stên</td>
</tr>
</tbody>
</table>

In the Middle Frisian language of the early 16th century proverbs, written by Reyner Bögerman (De Boer, 1900), we find a subject-object system for three genders. In Modern Frisian the distinction between the grammatical masculine and feminine gender has been abandoned for nouns and adjectives. Functional relations are expressed with the help of prepositions and syntactic structures, as is the case with Modern English and Modern Dutch:

- de stien ‘the stone’, de saak ‘the case’, út brie ‘the letter’
- subject, (in)direct object de stien, possessive construction fan de stien, etc.

1 Only the West Frisian dialect of Schiermonnikoog has three grammatical genders in nouns and adjectives.
The exact process of simplification and final abandonment of the case and gender system is not the subject of this thesis, but could be traced precisely with the help of the charter corpus.

**Inflectional morphology**
Old Frisian had several inflectional classes, partly distinct for gender, so for example:

- nom./acc. pl. masc. **stênan** ‘stones’
- nom./acc. pl. neuter **land_** ‘parcels’
- nom./acc. pl. fem. **seka** ‘cases’, **suster(a)** ‘sisters’

In Modern Frisian, almost all nouns end in -en or -s in the plural, according to phonological patterns: **stienen**, **lannen**, **saken**, **susters**

Less extended reductions have taken place in the inflectional patterns of verbs.

**Degemination**
In Old Frisian, a phonological opposition existed between long and short consonants. This is no longer the case in Modern Frisian, where consonants are always short:

- Old Frisian: **setta** [setə] ‘to put’
- Modern Frisian: **sette** [sete]

Sometimes, when the long consonant was shortened, the preceding vowel could be lengthened, as in Modern Frisian **stôk** [stɔ:k] ‘stick’ < Old Frisian **stok** /stɔ:k/, **bôle** [bɔ:la] ‘bread’ < Old Frisian **bolla** /bol-la/.

This so-called compensatory lengthening is no more than a tendency and is prevalent in the northern part of the language area (Spenter 1968, 16; FAND-database, 1980/95). The degemination of long consonants is discussed in § 2.2.

**Lengthening in open syllables**
Old Frisian short vowels have often been lengthened in open syllables:

- Old Frisian: **sone** ‘son’ with /o/, **dagan** ‘days’ with /a/
- Modern Frisian: **soan** with [oː], **dagen** with [aː]

The lengthening process is not as common as it is in for example Dutch and most of the North Germanic languages. The lengthening in open syllables is the subject of § 2.3.
Reduction of unstressed vowels

The development of vowels in unstressed syllables is the main object of this study. In Old and Middle Frisian texts, the letters <a>, <e>, <i>, and <u>, and in the older periods also <u>, are applied to render vowels in unstressed syllables. Modern Frisian has predominantly /ə/ in this context, regularly written <e>, in some suffixes also <i> such as in -lik [lɪk] (cf. English -ly). The current position on the phonological status of written mediaeval <a>, <e>, <i> and <u> is formulated by Boutkan (2001, 619):

“[...] the bulk of O[ld]fr[isian], attests to the merger of all vowels in unstressed syllables to /ə/.”

Sjölin (1969, 18) describes the spelling in the charters as a “[...] willkürliche Orthographie [...]”. According to Boutkan and Sjölin, Middle Frisian and perhaps even Old Frisian had only /ə/ in unstressed syllables. These opinions should not detract from the fact that in the oldest charters and codices, many forms are written with <a> and <u>. These spellings match the reconstructed historical phonological application. For example:

Old Frisian: (wi) <habbath> ‘we have'; <makia> ‘to make'; <sekum> ‘cases (dat. pl.)’
Modern Frisian: (wy) havwe, meitsje, sake (general plural form), where e = [ə].

Many Old Frisian vowels were not only reduced in quality in the subsequent periods, but could also be subject to syncope and apocope:

Old Frisian: dorg ‘door’, farw ‘sailed (past. part.)’
Modern Frisian: doar, fear

The details of these developments are monitored in chapter two, in particular in § 2.4, and discussed in detail in chapter three. Current opinions are evaluated in detail in § 3.1.

1.2.5 Basic views on the mechanisms of language change

Since the early beginnings of the historical linguistics of Germanic languages, scholars have assumed a causal relationship between the Germanic root stress and
the reduction of subsequent sounds in the word (for instance, Streitberg 1896, 29). In Indo-European, the stress could be both on the root and on inflectional endings. For example, Latin: *vidi* ‘I saw’, *vidit*is ‘you saw’. In Germanic languages, the intensity stress was concentrated on the root. For example Gothic *wait* ‘I know’, *witiþ* ‘you know’. The sounds of unstressed endings were subsequently reduced. The reduction or even disappearance of inflectional endings meant a loss of grammatical function. This inevitably caused the collapse of the Old Germanic inflectional system and a transition towards more analytical languages (Von Polenz 1972, 20).

This phonetically-based intuition may seem fairly logical. But from a synchronic, functional point of view it is difficult to understand. If a speaker from the year 800 was able to acquire a certain phonological grammar that allowed for a wide range of vowels and consonants in an unstressed position, functionally loaded by the archaic Germanic (originally Indo-European) inflectional grammar system, why should this change? The archaic grammar had evolved and existed for several generations. Being functional and possible, there seems to be no evident force to cause a change in the system (cf. for similar considerations Zuraw 2003, 139). Evidence of this perspective can be seen in modern Icelandic. There hardly any Germanic language with such strong and consistent intensity stress on the first syllable of every word, yet at the same time, Icelandic is also the Germanic language with the most archaic inflectional system with three different vowels /i, a, ð/ in an unstressed position, still in full use. Geographical isolation, together with the status of the language and the level of literacy of the population are widely mentioned as relevant factors (Karlsson 1989, 52-54).

Historical linguistics shows that detailed reconstructions of former phonology and morphology can be achieved. An imperative sequence of stages in the development of languages is questionable, but at least retracing seems possible. Neogrammarians argue that the cause of changes was to be found in ‘speech’ (Paul 1920, 32). From a Neogrammarian perspective, the direction of changes was
determined by 'convenience' ("Bequemlichkeit"), i.e. of articulation (idem, 56, 57). Paul contends that it is impossible to give any general indication about the nature and direction of this 'convenience' force. The problem with this view is that what is 'convenient' in one case can remain without implication in another. The case of final devoicing (for example, Frisian bêd [bê:t] and English bed [bê:d]) provides a good example. Final devoicing is a strong tendency in many stages of most Germanic languages. If it is for the sake of 'convenience', how can it be that Modern English can easily do without it?

The concept of 'convenience' is reformulated as 'markedness constraints' in Optimality Theory (OT). OT-constraints often express some kind of articulatory inconvenience that is to be avoided (Kager 1999, 5). In an OT framework, differences between languages or between two stages of one language can be the result of different constraint rankings. According to the theory, there is no natural ranking of constraints. The constraints are universal, but their ranking is language-specific and the principle of 'strict domination' favours one constraint above the other (idem, 11, 12). This implies the same uncertainty as in the Neogrammatically approach, where language change is somehow linked to 'convenience', but no definitive indicator can be found to test the theory. Thus every account of the past becomes a 'just so' story. Things happened as they did for a reason, not imperatively, but somehow logically. If developments are not imperative, the likelihood of reliable reconstructions of the history of languages diminishes and predictions for the future become impossible.

This study considers language\textsuperscript{11} as a deterministic dynamic system, governed by self-organisation. Such systems can exhibit chaotic behaviour. To my knowledge, this is not a fully established theory. This study attempts to contribute to the development of such a theory. Basic concepts of this approach are outlined in § 1.4. Application of these concepts is found in the models in § 5.1 and 5.2. The data analysis of chapter two preceding the elaboration of the theoretical concepts is in itself a contribution to Frisian linguistics. Section 1.3 constitutes the first stages of this study. How is written language material of late mediaeval Frisian used to investigate the phonological and phonetic changes in the language over that period?

\textsuperscript{11} In this approach, 'performance' and 'competence' are not separated; every utterance (in historical material 'spelling form') is treated equally. There is no a priori contrast between a grammatically correct form and a 'slip of the pen'.

-14-
1.3 Methodological preliminaries

In this section, several methodological preliminaries will be discussed. Firstly, a purely technical description is given on how the data was collected from the digital charter edition (§ 1.3.1). The question is then raised about spelling and spelling tradition. What are the implications of extracting phonological or even phonetic information from the written sources (§ 1.3.2)? The outcome of that section forms the basis for a closer definition of the grapheme, phoneme and phone concepts for the purpose of this study (§ 1.3.3). The step towards mapping geographical differences is explained in § 1.3.4, followed by a practical instruction on how to read the graphs and maps (§ 1.3.5). The differences between the language forms in original charters and copies, and their implications for the use of the sources is addressed in § 1.3.6. Apart from the relationship between spelling and phonology, the size of the corpus also poses a methodological problem. Is it large enough to warrant general statements? Are observed trends, for example, statistically reliable? Such questions are discussed in § 1.3.7. In the final section (§ 1.3.8) the use and interpretation of the data from the codex Unia are outlined.

1.3.1 Collecting the data

At this moment, the entire corpus of Frisian charters, city administration books and letters from the period before 1600 are being digitally lemmatised at the Frisian Institute of the University of Groningen. The information is already available on card indexes, kept in the institute, and is being converted into a digital database. This digital database will provide the opportunity to quickly find any form of lemma. A complete digital text edition of the charters are a prerequisite for this lemmatisation. Such a digital edition, created at the same institute in Groningen, is already available and has kindly been made available for this project. This digital text edition follows the printed edition of the charters (OFO I-IV), improved with corrections from Vries (1984).

Because the digital database with lemmatised information was not available when the research was carried out, specific spelling forms of a given lemma were selected, using regular expressions\(^\text{12}\) with the aid of a Python computer script. This method can be used as long as the lemma in question does not have too many homonyms. To allow for relevant corpus counts, the lemmas have to appear with substantial frequency. The question of absolute frequencies and statistical reliability is discussed in § 1.3.7.

\(^{12}\) Regular expressions are character combinations extended with additional options for wild cards, alternative characters, etc. To give an idea of the possibilities, the regular expression '\[ae\]c?k[ea]' will return positive matches including: seke, secke, sek, sak, saka, sack, sacke, etc. forms of Old Frisian *sake* 'case', but not, for instance, *saeke*. Etymological cognates are English *sake*, Dutch *zaak* and German *Sache*.
Most of the percentages in this study have been rounded off on integer values (whole numbers). This research uses around 100 data tables, some containing several thousand records. Including all such data in this publication, would exceed a reasonable size. The data underlying this study will be made available on the author's personal website at the Fryske Akademy: www.fa.knaw.nl.

As an example of a data table from the charters, table 1.2 shows the beginning of the working table of *seke* ‘case’. The complete table contains 876 records, including 443 from original charters. The legend of the table is given underneath the table. Tables of this kind are the basis for all the graphs, tables and maps presented in this publication.

<table>
<thead>
<tr>
<th>Chnr</th>
<th>Root</th>
<th>Ending</th>
<th>Case</th>
<th>Text</th>
<th>Year</th>
<th>Loc</th>
<th>o_c</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>20001</td>
<td>seck</td>
<td>apl</td>
<td></td>
<td>lave wra sindstall, ende alle seeke by riechta in da sind stalle,</td>
<td>1378</td>
<td>Fra</td>
<td>c</td>
<td>-</td>
</tr>
<tr>
<td>20001</td>
<td>seck</td>
<td>apl</td>
<td></td>
<td>deghen II riecht deghen. alle seeka to bij riechten, after deem</td>
<td>1378</td>
<td>Fra</td>
<td>c</td>
<td>-</td>
</tr>
<tr>
<td>20001</td>
<td>seck</td>
<td>nsg</td>
<td></td>
<td>engh-seke were, deir dat sind sinds naer fan seyd, dat to</td>
<td>1378</td>
<td>Fra</td>
<td>c</td>
<td>-</td>
</tr>
<tr>
<td>20002</td>
<td>see</td>
<td>nsg</td>
<td></td>
<td>hin ne (x) also wore seea thema 2) se inda lande nat birecba ne maybe</td>
<td>1379</td>
<td>Wun</td>
<td>qe</td>
<td>-</td>
</tr>
<tr>
<td>20002</td>
<td>see</td>
<td>in</td>
<td>x</td>
<td>jefhia elles 1) sekin seka hem tha heren fan Vraacht jefhia fan</td>
<td>1379</td>
<td>Wun</td>
<td>qe</td>
<td>-</td>
</tr>
<tr>
<td>30001</td>
<td>sak</td>
<td>nsg</td>
<td></td>
<td>ende is that sake that theese dijk wande mit maeche dullen and</td>
<td>1381</td>
<td>Lit</td>
<td>c</td>
<td>-</td>
</tr>
<tr>
<td>30001</td>
<td>sak</td>
<td>nsg</td>
<td></td>
<td>that sake that his thit naut an dwie al hir foret preyan is als that</td>
<td>1381</td>
<td>Lit</td>
<td>c</td>
<td>-</td>
</tr>
<tr>
<td>30001</td>
<td>sak</td>
<td>nsg</td>
<td></td>
<td>tiada schellet ene tila halda anda hoda tot Termnaa, and is that</td>
<td>1381</td>
<td>Lit</td>
<td>c</td>
<td>-</td>
</tr>
<tr>
<td>10002</td>
<td>seck</td>
<td>apl</td>
<td></td>
<td>mantghafde spreke, ende wy thine fowysya seka babbath</td>
<td>1386</td>
<td>Lit</td>
<td>qe</td>
<td>-</td>
</tr>
<tr>
<td>10002</td>
<td>seck</td>
<td>dpl</td>
<td></td>
<td>hir fore fan alle tha sekim ther sind seheen er Doza dathe, jefta</td>
<td>1386</td>
<td>Lit</td>
<td>qe</td>
<td>-</td>
</tr>
<tr>
<td>10002</td>
<td>seck</td>
<td>dsg</td>
<td></td>
<td>friendelike soneman in that seke twincba Heerta Heningha to</td>
<td>1386</td>
<td>Lit</td>
<td>qe</td>
<td>-</td>
</tr>
<tr>
<td>10007</td>
<td>seck</td>
<td>dsg</td>
<td></td>
<td>seke ende fen tha lande scales ende sonder seha to haldeane.</td>
<td>1397</td>
<td>Ldn</td>
<td>c</td>
<td>-</td>
</tr>
<tr>
<td>40001</td>
<td>seck</td>
<td>dsg</td>
<td></td>
<td>seke ende fen tha londe scales ende sender seha to haldeane.</td>
<td>1397</td>
<td>Ldl</td>
<td>o</td>
<td>-</td>
</tr>
<tr>
<td>20005</td>
<td>seck</td>
<td>nsg</td>
<td></td>
<td>dio seeke datter emgher handa clage jefha spreke fan comme dor</td>
<td>1402</td>
<td>Lit</td>
<td>oc</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1.2: Example of a part of a data table from the charters: Old Frisian lemma *seke* ‘case’

Chnr = chart number; e.g. 20001 = OFO II, nr. 1, 30001 = OFO III, nr. 1, etc.;
Root = root of the token (token = single appearance of a lemma); Ending = ending of the word; Case = number and case, e.g. dsg = dative singular; apl = accusative plural; Text = fragment from the charter containing the token; Year = year that the charter was written; Loc = location of the charter, based on the toponyms, supplemented with information from identified scribes (cf. § 1.3.4); o_c = original (o) or copy (c); qc = 17th century edition from the original, see § 1.3.6; Author = identified scribes (not in the example, because authors of the oldest charters are unknown).
In the fourth column, the paradigmatic interpretation is given. This is necessary to determine the status of the endings. Is it a dative singular, a nominative plural, etc.? Different numbers and cases can be expressed with different endings, especially in classical Old Frisian. Later, in the 15th century, the case system was abandoned and historically ‘incorrect’ endings were applied in several instances. The paradigmatic classification given is primarily of a functional nature. That means that a form is categorised as a dative singular if it corresponds to the syntactic and semantic function, even if the ending does not comply with the historical dative singular ending. This can be of great importance in determining a correct judgement of the development (cf. § 2.4.3.4).

<table>
<thead>
<tr>
<th>Root</th>
<th>Ending</th>
<th>Line</th>
<th>Text</th>
<th>Class</th>
<th>Gnr</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>mak</td>
<td>ade</td>
<td>885</td>
<td>were ther hi ther elath of makade, so seen it swera suene sirra</td>
<td>A</td>
<td>855</td>
<td>Older ‘Skeltenriuht’</td>
</tr>
<tr>
<td>mac</td>
<td>ad</td>
<td>2224</td>
<td>ther fan tha shoopum gat also maekad se mit tre and</td>
<td>A</td>
<td>869</td>
<td>Formulary for Opening a Thing Session</td>
</tr>
<tr>
<td>mak</td>
<td>aden</td>
<td>194</td>
<td>ther tha bugh makeren ti Rome, and riecht sete akker ser rie&lt;</td>
<td>A</td>
<td>874</td>
<td>What is Law? (Haet is riecht)</td>
</tr>
<tr>
<td>mak</td>
<td>ad</td>
<td>1377</td>
<td>‘heud sto’ ther thi expelle ust makad is. Tha thi panningen sath thi ‘fria Fresa sinn’ seulta sti</td>
<td>A</td>
<td>627</td>
<td>The Seventeen Statutes</td>
</tr>
<tr>
<td>makc</td>
<td>at</td>
<td>5777</td>
<td>asert hine van gode saet and maekat annen afgod fan</td>
<td>A</td>
<td>876</td>
<td>Authentica Riecth</td>
</tr>
<tr>
<td>mak</td>
<td>ad</td>
<td>5228</td>
<td>hot so wise loode. habbat makad and hrocht in ean</td>
<td>A</td>
<td>874</td>
<td>What is Law? (Haet is riecht)</td>
</tr>
<tr>
<td>makc</td>
<td>at</td>
<td>5175</td>
<td>ment to maeken.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>makc</td>
<td>ad</td>
<td>3695</td>
<td>Thio hispinge seil maekad wass ma seritu under thiisa</td>
<td>A</td>
<td>647</td>
<td>Veredung this riechts (Processus Judicii)</td>
</tr>
<tr>
<td>makc</td>
<td>at</td>
<td>3616</td>
<td>se an eind maekat van ther seke ther ma om kyfis.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>makc</td>
<td>ad</td>
<td>4828</td>
<td>maekad and reeknat. and hi Wainbrir</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>makc</td>
<td>at</td>
<td>5107</td>
<td>forwert dere hen hem maekat senth by anne preester teo</td>
<td>B</td>
<td>885</td>
<td>Statutes of the Leppa</td>
</tr>
<tr>
<td>makc</td>
<td>at</td>
<td>4368</td>
<td>Alle deste bots sent maekat to gane to riechts ly</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.3: Example of a data table from the codex Unia: Old Frisian lemma makia ‘to make’. The special aim of this table is to trace the transition from single consonant spelling in for example, Old Frisian makad (past. part.), towards double consonant spelling, Modern Frisian makke (past. part.) c.f. § 2.2.

Root = root of the token; Ending = ending of the word; Line = line in the digital text edition used; Text = fragment from the codex containing the token, corresponding to the line given; Class = grouping of texts in three classes according to linguistic character, see. § 1.3.8. The texts of group A are ordered according to a tentative sub-classification, not shown in the table, reflecting the linguistic age of the text; Gnr = the so-called ‘Gerbenzon’ number: the Gerbenzon-index is an index on the entire mediaval Old Frisian corpus where every text has a unique number. Texts can appear in different codices in different editions; Title = English title of the text.
The text of the codex *Unia* has not yet been published. A preliminary digital text edition is available at the Fryske Akademy. This text version was available for this research. It should be noted that the transcription is not yet complete, but the parts available are assumed to be well transcribed. This text was also searched with regular expressions. An example of a data table from *Unia* is given in table 1.3. The question of how the text was used and dated is presented in § 1.3.8.
1.3.2 Spelling and spelling tradition

Written charters from the period 1378-1550 have been used to obtain information about phonological and phonetic features of the Frisian language and their variation in time and space. The first issue that has to be clarified here is whether the charters are suitable material for answering questions on pronunciation at the time of writing. Are written texts embedded in some form of spelling tradition that filter and mask the phonological details and the phonetic features of the spoken language of the time? What was the relationship between spelling and pronunciation in Frysln in the 15th century?

The attested charters are primarily juridical and administrative texts. They belong mainly to formal registers. They are definitely not transcriptions made by a trained linguistic scholar. The spelling used in the charters is the result of several factors. Written pieces were prepared and drafts were copied into neat versions. For archiving purposes, copies could be made from the original. There are some instances where multiple versions of one charter have been preserved. For example, OFO II-3 with two parallel old copies. Was the text dictated by one of the participants in for instance a peace agreement, or did a clerk prepare the piece, using his private formulations and spelling habits? In cases of agreements between multiple parties, representatives of which party wrote the text? All kinds of situations might have occurred (cf. Rem 2003, 52-53). As the charters are formal pieces of writing, it may be assumed that the authors were careful in their writing. What we see must have been a reasonable and acceptable spelling.

Established spelling and loose spelling traditions

Was there a spelling tradition mediaeval authors could rely upon? Were authors looking at a previously applied spelling in older texts or were they trying to write as they spoke, as far as that was possible? What would be the pattern of spelling and spelling changes in the charters if spelling traditions existed at that time? According to the qualification of ‘random spelling’ as applied by Sjölin (1969, 18, cf. § 1.2.4), there would be little hope in expecting any meaningful conclusion from the spelling in the charters subject to a reconstruction of sounds actually realised.

In the modern situation of established languages, spellings are fixed and then implemented for long periods. Texts are almost entirely homogeneous in spelling. Individual, generational and geographical differences in pronunciation and grammar are (almost) entirely obscured by the spelling conventions. Spelling changes are discrete and fairly consistent in their application. This was definitely not the case in 15th-century Frysln.
In a less centralised society, such as the mediaeval county of Holland\textsuperscript{13}, the spelling might be somewhat looser. Spelling is relatively homogeneous for clusters of authors. For instance, from one scriptorium or chancery. Changes in the pronunciation are only shown with delay, if at all. Established traditions may be responsible for the appearance of archaic forms. Geographical peculiarities are filtered and replaced by more traditional forms. As authors are explicitly reflecting on their spelling, their choices are more categorical than incidental. Personal variation will be the result of ‘mistakes’.

The possibility of an Old Frisian spelling tradition
Was there anything like a Frisian spelling tradition in the 14\textsuperscript{th} and 15\textsuperscript{th} century? Meijering (2005, 200) asks in his review of Bremmer 2004:

\begin{quote}
“Hoe verklaart [Bremmer] dat reeds de vroegste friestalige documenten, ja zelfs losse woorden in een Latijnse context, zo’n consistente spelling vertonen? Waar komt die vandaan?”\textsuperscript{14}
\end{quote}

This suggests something like an Old Frisian spelling tradition. Traditions are per se the result of a gradual evolution. Bremmer (2004, 78) suggests something similar when writing about a 13\textsuperscript{th} century author being: “vertrouwd […] met het geschreven Oudfries van zijn tijd.”\textsuperscript{15} One can only be ‘familiar’ with ‘the written Old Frisian’ if a substantial number of texts, exhibiting some kind of spelling tradition is assumed.

This so-called classical Old Frisian spelling tradition comprises the following spelling conventions (Sjölin1969, 17, 29-30; Bremmer 2004, 78):

- Vowel length is not indicated with an additional subsequent vowel;
- Geminate consonants are written with a double consonant sign;
- Historical /ð/ and /þ/ are written <th> and not <d> and <t>;
- Historical unstressed full vowels are written, so: <alra> instead of later (spelling) forms like <alre>, <aller>, <alder>.

\textsuperscript{13} Unlike the international practice, Holland covers only the western part of the current Netherlands (cf. England not to be confused with Great Britain).

\textsuperscript{14} How does [Bremmer] explain that the earliest documents written in Frisian, even isolated words in a Latin context, already show such a consistent spelling? Where does it come from?

\textsuperscript{15} Being familiar with the written Old Frisian of his time.
It is not necessary to postulate on an independent Old Frisian spelling tradition as the reason for the existence of those spelling practices in the Old Frisian texts. These practices comply with the spelling conventions of Latin: No vowel length indication, double consonants where appropriate and circumscribing the Greek letter <è> with <th>. There is no problem with full vowels in either stressed or unstressed syllables. Every mediaeval writer of Old Frisian, familiar with Latin, would ‘re-invent’ the so-called ‘classical Old Frisian spelling’.

There is little evidence that a substantial corpus of Frisian texts existed, to establish an independent Old Frisian spelling tradition. Bremmer (2004, 85-86) states that Frisian was hardly written before 1200 and Latin remained the main written language in Fryslân in the 13th century and even later (idem, 71).

Spelling practices in the Middle Frisian era

Also, in the 15th century, the circumstances for the development of an independent Frisian spelling tradition were not favourable. There was no leading writing centre in Fryslân, such as a feudal court or any form of central government. Scriptoria were dispersed over monasteries and city chanceries (Vries 1993, 65). Initially, professional writers were literate in Latin and Dutch and later in Frisian. Several attended universities in the Low Countries or adjacent regions, including Leuven and Cologne (idem, 68). There were no universities in Fryslân at that time. Every literate person in Fryslân was acquainted with Middle Dutch (idem, 66). Several text types were exclusively available in that language and none at all in Frisian (idem, 61 ff.). The importance of Dutch as a written language is also illustrated by the fact that Dutch was a serious alternative for several individuals in the 15th century, even for private use (idem, 68-69). So if there was any spelling tradition that writers were familiar with in Fryslân in the 15th century, it was Dutch.

The conclusion is that there was no socio-cultural basis for an independent Frisian spelling tradition that could mask personal, temporal or geographical variations. The number of Frisian texts was too limited and the practical confrontation with non-Frisian texts too intense in the period 1200 to 1550. People writing in Frisian were familiar with spelling conventions, particularly Latin in the earliest times, and Dutch from the 15th century onwards. People knew how to write a language, they did not know how to write ‘standard(-like) Frisian’. The solution lay in writing the language as they spoke it, with the graphical means they were familiar with.

The word choice and syntax of the charters, for example, in the standard formulas, may be highly defined by the genre of formal, juridical texts, that were often influenced by Latin and Dutch examples. In choices between morphological endings and the quality of unstressed vowels, the authors had nothing to rely on other than their own competence of Frisian and their familiarity with written
Dutch and Latin. The choices may have been somewhat conservative, due to the careful writing process, but the outcome reflects existing contemporaneous phonological and morphological variations in the spoken language.

Section summary:

- There was no independent spelling practice for Frisian during the Middle Ages;
- Authors of Old and Middle Frisian used existing conventions from other languages they were familiar with;
- The presumed ‘Old Frisian spelling tradition’ is the result of an ad hoc application of Latin spelling conventions;
- Middle Dutch spelling practices were used for Middle Frisian.
1.3.3 Phones, phonemes and graphemes

Since the introduction of structuralism into linguistics, a sharp distinction is made between:

- The sound or phone - usually considered the field of phonetics;
- Its categorical interpretation in a semantically contrastive context as a phoneme;
- Its representation in conventional spelling systems, grapheme.

This was a large methodological improvement. Within that framework, one is able to raise questions for instance on the relationship between spelling and pronunciation or the interaction between allophony and sound change. The methodological separation between grapheme, phoneme and phone induced scepticism about their interrelation. Some scholars had become reluctant to formulate any positive statements on pronunciations in the past and preferred a strictly graphematical analysis.\(^\text{16}\)

In this study, spelling is assumed to be phonemic in principle. In some cases allophonic variations and phonetic details are expressed. This resembles the practice of broad phonetic transcriptions. The default interpretation of the mediaeval spelling in this study is that <a> represents /a/ which is expressed in speech by [a], etc. In several aspects, the Latin and Middle Dutch spelling conventions were different. These differences must be taken into account.

**Reading Old and Middle Frisian phonemes through Latin and Middle Dutch spelling conventions**

The spelling conventions of Latin and Middle Dutch result in some deviation from the default hypothesis. Latin and Middle Dutch have different systems to represent (or not represent) the length of vowels and consonants.

There was a transition in the early 15\(^{th}\) century from Latin-based spelling practices towards a spelling practice that kept pace with that of Middle Dutch. In the Latin-

\(^{16}\) Sjölin (1970, 104) formulates a strictly theoretical point of view: “Eine ‘graphonemische’ Interpretation der graphematischen Daten ist in diesem Zusammenhang strenggenommen nicht relevant, da wir nicht wissen, inwieweit der Text der Hs. auch als Lautung realisiert wurde.” (A ‘graphonomic’ interpretation of the graphematic data is, strictly speaking, irrelevant, because it is unknown to what extent the text of the codex was realised as acoustic sounds). However, this is the introduction to a section with the title: “Schriftbild und Lautung” (‘spelling and sounds’). Following this opening statement, Sjölin articulates several observations on pronunciation in mediaeval Frisian, based on spelling contrasts or, on the contrary, the absence of them. So despite Sjölin’s theoretical renouncement, he does draw phonemic and phonetic conclusions from written mediaeval texts.
based spelling practices, vowel length is not indicated. So, late 14\textsuperscript{th} century <herath> represents /h\textipa{@}ra\textipa{\textbackslash n}/ ‘(we) hear’, while <seke> represents /s\textipa{\textbackslash k\textbackslash o\textbackslash n}/ ‘case’.

In Middle Dutch vowel length is indicated, applying the following system (examples from Middle Dutch):

- Short vowels in closed syllable are written with a monograph, e.g. <dat> = /dat/ ‘that’;
- Long vowels in closed syllables are written with a digraph, e.g. <doer> = /d\textipa{\textbackslash r}/ ‘through’;
- Long vowels in open syllables can be written with either a monograph or a digraph, so both <zonen> and <zoenen> represent /zo:n\textipa{\textbackslash n}/ ‘sons’.

The shift in orientation from Latin spelling practices towards Dutch practices can be illustrated by the spelling of the Old Frisian word <fore>, Modern Frisian <foar> ‘for’. Originally the word has a short root vowel: /\textipa{\textbackslash r}/. The short root vowel is lengthened in open syllable in the 14\textsuperscript{th} century, making /\textipa{\textbackslash r}/. After 1390 the final schwa is dropped, making /\textipa{\textbackslash r}/.

<table>
<thead>
<tr>
<th>time frame (approximation)</th>
<th>phonological form</th>
<th>spelling</th>
<th>spelling tradition</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 1300</td>
<td>/\textipa{\textbackslash r}/</td>
<td>&lt;foar&gt;</td>
<td>Latin</td>
</tr>
<tr>
<td>14\textsuperscript{th} century</td>
<td>/\textipa{\textbackslash r}/</td>
<td>&lt;foar&gt;</td>
<td>Latin</td>
</tr>
<tr>
<td>&gt; 1390</td>
<td>/\textipa{\textbackslash r}/</td>
<td>&lt;foar&gt;</td>
<td>Latin</td>
</tr>
<tr>
<td>&gt; 1430</td>
<td>/\textipa{\textbackslash r}/</td>
<td>&lt;foar&gt;</td>
<td>Middle Dutch</td>
</tr>
</tbody>
</table>

Table 1.4: Interpretation of vowel length according to Latin and Middle Dutch spelling practices.

Due to the process of degemination of long consonants, a reverse interpretation is applied to the consonants. In the older parts of Unia and the oldest charters, double consonants mark phonologically long consonants, complying with Latin spelling practices. In more recent texts and charters, double consonants were used to show the shortness of preceding vowels in accordance with Middle Dutch writing practices (cf. Hofmann 1969). The interaction between the actual phonology of geminate consonants and the spelling of consonant digraphs is discussed in § 2.2.3.
The 'prototype' of a sound comes close to, but does not equal the generative concept of underlying phoneme. Allophones for example, can represent two prototypes, but one underlying phoneme.

---

Table 1.5: Interpretation of consonant and vowel length according to Latin and Middle Dutch spelling practices.

<table>
<thead>
<tr>
<th>Time Frame</th>
<th>Phonological Form</th>
<th>Spelling</th>
<th>Spelling Tradition</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; ± 1400</td>
<td>/sɛt-/ta ~ lɛza ~ dedø/</td>
<td>&lt;setta ~ lessa ~ dede&gt;</td>
<td>Latin</td>
</tr>
<tr>
<td>&gt; ± 1450</td>
<td>/sɛpta ~ lɛza ~ dedø/</td>
<td>&lt;setta ~ lessa ~ dede&gt;</td>
<td>Middle Dutch</td>
</tr>
</tbody>
</table>

So, information about vowel length in the Latin spelling tradition, and consonant length in the Middle Dutch spelling tradition can be obtained solely from indirect evidence.

Schwa (ə)

A special case is <e> and <i> in unstressed position. In older stages of the Germanic languages, there was no /a/. Written <e> and <i> in unstressed syllables reflected something like /e/ or /ɛ/ and /i/ or /ɪ/. But in Old and Middle Frisian, the letters <e> and <i> alternate in positions where historical phonological reconstructions suggest /o/. Middle Dutch had /o/ and this sound was most often rendered as <e>, sometimes also <i>, <a>, <o> or <u> (Pijnenburg 1997, 80). For <e> in unstressed syllables, the default interpretation will be: <e> ~ /o/ ~ [ø]. In the data analysis in § 2, the tokens spelled with <e> and <i> are kept apart. The case of alternation between <e> and <i> in unstressed syllables is analysed in detail in § 3.6.

Phonetic notation and real-world realisations

The phonetic sign [a] is a container for a theoretically endless range of actual realisations. In real performance, there are measurable differences in formant contrast and length between [a] in word-final position, syllable final position, and protected [a], in isolated words or in allegro speech, etc. even for one speaker. This variation goes beyond the possibilities of the International Phonetic Alphabet. As an example, the IPA allows for three different length categories, but in reality, the length can be measured on a continuous scale in milliseconds with acoustic equipment. The IPA sign [a] represents an abstract realisation of /a/: the prototypical realisation.17

---

17 The 'prototype' of a sound comes close to, but does not equal the generative concept of underlying phoneme. Allophones for example, can represent two prototypes, but one underlying phoneme.
The following analysis will show that the key to the answer about the cause, order and direction of changes can be found in the variation that lies in performance variation. Some spelling practices even seem to be an expression of it. Therefore, the interpretation of sound features will benefit from an approach where actual realisations of one phoneme are not considered as uniform members of a discrete group, but as clusters of appearances, to be measured on continuous scales. A fluid interpretation of the concept phoneme results in a blurred transition between phoneme, allophone and phone (cf. Pierrehumbert 2003, 183 & 192).

Prototypical realisations and some aspects of performance variation are language-dependent. On the other hand, some of the performance variations seem to be universal, because they are controlled by articulation constraints. These constraints are the consequence of general characteristics of the human speech organ and its articulatory control. They are bound by the biological and physical substratum. For example, the formant contrast of vowels produced in fast speech are less than the formant contrast of the same phonological vowel in slower speech. This phenomenon is for example observed both in Frisian and Japanese (De Graaf 1986, 17-19). Because physical laws are universal both in place and time, and the human species has not changed significantly during the past 1000 years, one may assume that many phonetic performance effects were the same in 1400 as they are now. For example, when modern measurements show that the intensity volume of [a] is larger than that of [ə], that fact is taken to be valid for the 15th century as well. After all, all notations with ‘[ ]’ concerning historical data are, strictly speaking, hypothetical.

Notation legend
Within the framework of the default hypothesis of identification of grapheme with phoneme and prototypical phonetic realisation, it is sometimes difficult to choose an appropriate notation. Below the reader finds the interpretation that is given to the several notations in the course of this study, beginning with the written characters:

- `<a>`: The character ‘a’ on paper; the basic idea is that the author wrote that character because its sound value in Latin and/or Dutch was the closest match with the sound that he perceived in his own native language intuition;

- `/a/` A meaningful sound category in the language, distinct from /o/, /ɛ/ or any other phoneme by creating different meanings in (semi-)minimal pairs. This phoneme has a prototypical realisation;

- `[a]` The prototypical realisation of `/a/`: the most open sound of all language sounds and of relative short duration, etc. The phonetic
notation [a] is still covering a wide range of performance variation in duration, formant contrast, etc.;

- This is meant to be a diachronic notation. It is, for instance, used to mark the infinitive ending of a verb, expressing the following implicit assumptions:
  - <a> is the almost exclusive notation for infinitives in the Old Frisian sources,
  - it probably reflected the phoneme /a/,
  - it was prototypically realised [a],
  - after 1490 it was increasingly written as <e>
  - in early Modern Frisian it was finally pronounced as [ø],
  - etc...

Section summary:

- The spelling of Old and Middle Frisian was phonemic in principle. In some cases allophonic variations and phonetic details are expressed;
- The default interpretation of the spelling is a one-to-one match between grapheme, phoneme and phone;
- This interpretation needs corrections for diverging practices of Latin and Dutch in the representation of vowel and consonant lengths;
- Phonetic variation is measured on a continuous scale;
- Bio-physically induced sound patterns are universal in time and space.
1.3.4 Mapping the charter language

When the spelling reflects the major traces of phonology, geographical variations may also be reflected in the spelling. In other languages, extensive research has taken place into the methodological possibilities and constraints of historical dialect maps. In Germany the key word in this discussion is *Schreiblandschaft*, meaning ‘spelling landscape’ (König 2001, 95). This term indicates observed geographical variations in written sources, while leaving open the question about the precise relation between written and spoken language. For late mediaeval Frisian, an example of an historical dialect map can be found in Miedema (1986), whereas Vries (1986) discusses dialectal features of one individual author. Miedema relies on text content to locate the sources. Vries focusses on the author’s biographical data.

Where does one locate the linguistic information on the map? A charter can be the result of an interaction between two or more parties which do not necessarily come from the same region. Was the text dictated, was it a prepared written concept, who dictated it and whose clerk wrote it? (Rem 2003, 24-28)?

Rem (idem, 23) formulates several criteria to decide on the suitability of mediaeval Dutch charters for language geographical investigations. She suggests that “the local dialect” can be found in charters where:

- The parties involved come from the same place or region;
- There is no-one with a high social status involved. A count or duke for instance, may employ clerks from a wider region and his chancery might develop a non-local writing tradition whose legibility was aimed at reaching a wider audience.

In a Frisian context, everything is basically from the ‘same’ region from the perspective of the Low Countries. In § 1.3.2, it was shown that there was no feudal court or dominating cultural centre in Fryslân at the time that could develop a uniform language.

A combination of location methods was used for the Frisian charters. To identify the origin of the charters, place names in the text are particularly suitable. For example, names of dwelling places of the people involved, geographical objects mentioned in an agreement or will, etc. These give a clue to the background of the parties, hopefully including the person that defined the linguistic character of the text. For several charters, the names of the authors are known (Vries 1984 & 1993) as well as parts of their biographical data.
In the onomastics department of the *Frysk Academy*, a database has been prepared with all the toponyms from the charters. Every object was located and, where possible, identified with modern objects. For location reference, modern municipalities are used. The default approach was to allocate a charter in the centre of the municipality most frequently, not necessarily as an absolute majority, mentioned in it. This association was carried out with a simple database operation on the charter toponym database. There are several uncertainties connected with the outcome. The author or clerk may have come from another region or may have lived there but originated from another region. The author is most likely, but not necessarily the person who developed the linguistic character of the text. As many factors are uncertain, the linking by toponyms was the only option available for the majority of the charters.

Map 1.2: Base map for dialect maps based on mediaeval charters.

---

18 The author would like to express particular gratitude to Karel Gildemacher for providing this database at an early stage of this research (published in Gildemacher 2008). The geographical method presented in this section was in fact the key to the analysis that resulted in this thesis.
The application of the toponomy location method provided an opportunity to test the results. If the method were too fuzzy and unsure, no consistent maps could be expected from the data. But test maps drawn after the application of the toponomy localisation already appear to be consistent in their geographical distributions. This is implicit proof of the validity of the method.

The topographic base is shown in map 1.2. The region north-west of Leeuwarderdeel was ‘undyked’ marshland at that time. It is in the modern municipality of Het Bildt. The lakes and moors are shown to illustrate the geographical limitations of the cultivated area. They are not repeated in the legend of other maps. The south-east was particularly scarcely populated. There are very few charters from those regions. The borders correspond to the communities (deelen) and cities of the late Middle Ages. It should be kept in mind, that the allocation is done on the basis of the modern municipalities. The majority of the modern municipalities correspond to old ones. Some are the result of merging two old ones. For example, the modern municipality of Skarsterlân is the result of amalgamating Doniawertal and Haskerland. The point ‘Skarsterlân’ is on its capital Joure, lying near the common border of the two former municipalities, in the heart of the new one.

To improve the statistical reliability of the maps, in some cases the data has been aggregated from municipalities to six larger regions. The regional division can be found in map 1.8 (cf. § 1.3.7.4).

This toponomy allocation method was checked with the information obtained from identified authors (Vries, 1984 & 1993, and some additional non-published information provided for this research by Vries). The content of information about these authors varies. Sometimes it is simply an anonymous handwriting that has been recognised on several charters. Sometimes the name is known, while sometimes only the function is known, without any additional biographical data. For example, one of the identified authors is a ‘citizen in Harlingen’. The interpretation applied in this study is that he came from Harlingen or at least had been living there for so long, that he had assimilated his speech to that of other citizens of the city. But we can’t be certain.

The charter database table contains 1,247 localised items. For 289 of them, the writer’s hand has been identified in at least a part of the main text (excluding the cases with identified handwriting on the dors). This means that known handwritings can only be used as a partial check. In 53% of the overlapping cases, the localisation based on the toponyms, complies with that based on the authors. That is a majority, but it would imply that almost half of the locations are wrong. Also, instances where charters were assigned to adjacent municipalities were
initially counted as ‘wrong’, like Leeuwarden instead of Leeuwarderadeel, Bolsward and Harlingen instead of Wûnseradiel and Sneek instead of Wymbritseradiel (or vice versa). These errors do not significantly affect the general picture. Assuming those cases are ‘correct’, raises the overall level of correctness to 62%.

Of the remaining 38% of mismatches, 20 per cent points are due to the wrong allocation of texts written by Hemma Odda zin. He is the most productive author in the corpus. Hemma Odda zin is (or is one of) the author(s) of 116 charters in the database, or 9% of all the charters. He was city clerk of Leeuwarden and deputy-mayor of Tytsjerksteradiel, but was probably born in Bolsward. He owned estates and houses in neighbouring municipalities that are mentioned in the charters with toponyms. As city clerk of Leeuwarden, he often assisted in compiling charters. Using the toponomy method, texts written by him were allocated in the municipalities of Ferwerderadiel, Bolsward, Dongeradeel, Dantumadiel, Franekeradeel, Tytsjerksteradiel and Boarnsterhim. Only 41% of the charters written by Hemma were assigned to Leeuwarden or Leeuwarderadeel using the toponomy method.

Looking at the reliability from the perspective of the individual municipalities, it appears that municipalities towards the south and west of Fryslân, or those further away from Leeuwarden, had a match of around two-thirds between the toponomy method and the writers’ method. Due to the impact of Hemma Odda zin, the match between these methods was only about a quarter in municipalities directly surrounding Leeuwarden and further to the north-east. (For example Dongeradeel, Tytsjerksteradiel and Littenseradiel). The language of the city of Leeuwarden (and on a lower level for the cities Harlingen, Bolsward with Wûnseradiel) distorts the data from the neighbouring countryside regions. In instances where it was known, the locations were corrected for information from the known authors. The information about the texts written by Hemma Odda zin correct the large distortions in the regions surrounding Leeuwarden. After this correction, the geographical reliability of the countryside regions around Leeuwarden is much better now than the aforementioned one-quarter. It is unlikely that another highly productive writer was present among the authors of the anonymous charters who has not been identified by Vries. Still, some individual mismatches will remain.

Looking at the city of Leeuwarden itself, 93% of the charters assigned to the city by the toponomy method are also linked to Leeuwarden according to identified authors. This means that typical peripheral dialectal features rarely appear in the centrally located city by incorrect geographical assignment. Such peripheral features will appear where they ‘belong’, but their proportion might be

\[\text{\textsuperscript{15} cf. Vries (1993, 190) and \$2.3.4.2.}\]
underestimated due to the influence of city writers. Given the combination of
toponyms and identified handwritings along with the assumption that no other
highly productive writer will be found among the anonymous charter authors, the
overall reliability of the locations can be estimated at over 70%.

In practice, linguistic information in peripheral regions is very consistent in the
maps. The default geographical opposition in Fryslân in the 15th century is
between the south-west and north-east. In extreme positions, ‘typical dialectal’
features very often constitute over 75% and sometimes 100% of the attestations
allocated in that region. This supports the reliability of the applied method.

Section summary:

- The West Frisian charters are localised using dominant
toponymic references in the texts;
- This toponomy method was corrected for instances where
the author has been identified and his place of origin is
explicitly known;
- The overall reliability of the localisations is over 70%.
As mentioned in § 1.3.1, the classification 'dative plural' is a functional-syntactic one. After 1490, there is no longer a formal contrast between dative plural and nominative/accusative plural.

1.3.5 Reading instruction for maps and graphs

Graphs and maps are of great importance in this thesis. They are not merely illustrations, but form an integral part of the text. In this section, the reader will find explanations on how to read and interpret the graphical representations.

Variants from tables such as 1.2 are categorised into classes, counted and depicted in graphs. Making a classification implies neglecting irrelevant facts and aggregating variants that belong together from a specific perspective. Table 1.6 shows the changes in the ending of the dative plural in the noun seke 'case' in absolute numbers. The graph’s focus is the transition from the archaic ending <um> with secondary forms <em> and <im> against the modern plural form <en>. The archaic ending of the nominative and accusative plural <a> is attested in the function of a dative plural in a few instances. Irrelevant in this perspective is for instance, the spelling of the root consonant as <k> or <ck>, as is the quality of the vowel before the <m>. In graph 2.1 for example, it is the difference between <k> and <ck> that counts. The data in the table and graphs are divided into time frames. The limits of the time frames are a result of pre-analysis of the material. The first time frame starts in 1378 (plus the single charter from 1329). The most common time frames are those used in table 1.6.

<table>
<thead>
<tr>
<th>case ending</th>
<th>1430</th>
<th>1430-1460</th>
<th>1460-1490</th>
<th>1490-1510</th>
<th>1510+</th>
</tr>
</thead>
<tbody>
<tr>
<td>-Vm</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>-a</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>-en</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>12</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 1.6: Number of attestations to the dative plural of seke showing the distribution of different endings in several periods, using original charters.

Reading table 1.6 in the time frame up till 1430, there are two instances of a dative plural of seke in the original charters. These are OFO 1-2, 1386 <sekim> (cf. table 1.2) and <seckum> (OFO 1-28,1418). The form was categorised as -Vm, meaning vowel + <m>. That implies that 100% of the attestations from that period belong to the class -Vm. Between 1430 and 1460 there are eight instances of a dative plural of seke in the original charters. Five times <um> and once <em> making a total of six for the class -Vm. The forms <secken> and <secka> are both found once. The latter is the form of the accusative plural, appearing in a dative context. The spelling -Vm covers six out of eight cases or 75% in the time frame 1430 to 1460. The figures are shown in graph 1.1.

---

20 As mentioned in § 1.3.1, the classification 'dative plural' is a functional-syntactic one. After 1490, there is no longer a formal contrast between dative plural and nominative/accusative plural.
Graph 1.1: Graphical depiction of the data in table 1.6.

Map 1.3: Distribution of cases of plural of ‘son’.
The vowels of the word ‘son’ in Old Frisian have a complicated history. It is an old **u-stem**, cf. Riustringen Old Frisian *sunu*, other Old East Frisian texts *sune*. The palatalisation and delabialisation in the Middle West Frisian plural looks like an instance of *i*-mutation, cf. Icelandic *has* - *synir* (sg. - pl.). Old Saxon and Old High German tend to level all the *u*-stems to the *i*-stems, with *i*-mutation, cf. Modern High German *Sohn* - *Söhne*. But the Middle Frisian vowel might also fit in a wider pattern of spontaneous palatalisation, like Modern Frisian *sinne* ‘sun’ < Old Frisian *sunne* (Hoekstra 2007). This raises the question of where the vowel /o/ in the singular comes from. It might be Dutch (Modern Dutch *Zoon*), but the archaic Older ‘Skeltenariucht’ in the codex *Unia* (cf. § 1.3.8) has only <*sone*>. There is a general tendency in West Frisian to lower /u/ to /o/ (Hoekstra 2001a, 723), for example, Modern Frisian *komme* ‘to come’ [komə] < Old Frisian *kuma*. During this transition, a Vowel Harmony effect became apparent in the Older ‘Skeltenariucht’. The root vowel /u/ is dominant in the infinitive <*kuma*>, and the third person singular present indicative <*kum(ə)*th>, but in the conjunctive present, the /o/ prevails: <*come*>. The same effect will be the cause for the early transition from /u/ > /o/ in the word *innu*/*sone*.

---

The plural of ‘son’ appears in the charters as <z/so(e)nen>, <zonnen>, <z/sennen>, <z/sinnen>. The focus of categorisation in this example is the presence of <o> in the root. Forms without <o> represent phonological /sønən/ or /sɪnən/. The alternation between /e/ and /i/ is not relevant for this categorisation. Forms with <o> represent phonological forms like /sønən/ or /sønən/. The difference between long and short vowels may be important from an etymological point of view, but this map series is concerned with the opposition between front - back: /e/ /i/ ~ /o/. Map 1.3 shows the geographical distribution of the data, against the background of the topography of Fryslân of the 15th century (see map 1.2). In charters linked to municipalities like Dangeradeel, Heerenveen or Wymbritseradiel only one plural form is found. In Franekeradeel there are 11 tokens designating the plural form of ‘son’ and 17 in Leeuwarden, etc. The circle size for one and two attestations is the same. The map caption includes the period covered by the data. In this case it is quite a long period. Graph 1.2 shows the temporal distribution of the variants of ‘son’. At first glance, it shows that the <e>/<o>-ratio does not really change through time. The overall percentage of forms with <o> is 29%. The idea of a temporal trend can safely be dismissed and the data from the entire period can be used to draw a map. This is different to a distribution as found in graph 1.1. When the diachronic changes are large, not all the data may be shown on one map, because temporal and geographical variations could interfere.
Graph 1.2: Ratio of plural forms of ‘son’ with <e> and <i>, counted as <e>, or <o>, between 1431 (oldest plural) and 1547 (youngest plural).

Map 1.4: Categorised data; the numbers represent the percentage <e>.
Following the categorisation of the data, the categorised data can be shown on the map at the municipality's central point with pie charts. This is shown in map 1.4. The size of the circles still reflects the number of attestations per municipality. The pie charts show the ratio between the forms of <o> and <e>. The dark portions reflect the relative number of forms with <e>, while the white portions show the forms with <o>. The figures next to the pie charts show the percentage of forms with <e>. For instance, in Franekeradeel there are 11 attestations to plurals of 'son' in original charters: one time <zoenen>, the rest <sinnen> or <sennen>. A portion of 91% (= 10/11) of the pie chart is grey, representing the forms with <e>. The number ‘91’ is shown next to the pie chart.

The map reveals that forms with <o> are only found in the centre and north-east of the province. The south-west has only <e>. Two points east of Leeuwarden have only <o>, suggesting a trend from south-west to north-east.

This trend is illustrated using cartographical interpolation techniques. A trend surface is calculated using the percentages in the map points. The trend surfaces in the maps in this thesis are calculated with the interpolation technique called moving average.\(^{22}\) The area of the province Fryslân is divided into a grid of very small points of 100 x 100 metres, like the pixels in a digital camera. For every pixel in the grid, a value is calculated on the basis of the point values in the vicinity. Municipality points exercise influence within a maximum radius of 25 kilometres. The closer the municipality point, the larger the impact.

Map 1.5 shows the trend surface that was computed from the percentages in map 1.4. The applied method may result in a difference between the computed surface value and the original percentage of the municipality point. For example, at the point of the municipality of Franekeradeel, with 91% <e> forms in the plural, the trend surface has a value of 71%. Measuring the map reliability is discussed further in § 1.3.7.7. The outcome of that discussion is that reliability is lower in the far north-east, with 0% <e> in Leeuwarderadeel and Tytsjerksteradiel and 100% in Dongeradeel. The general trend is reliable. The detailed pattern with a secondary high in Dongeradeel is questionable.

\(^{22}\) All historical dialect maps are made using the Geographical Information System software ILWIS: www.itc.nl/ilwis. The description of the moving average technic is given in the appendix 2.
In some municipalities there is no data at all. This is why regions have been marked with a hatch. This is not necessary for a municipality such as Littenseradiel, despite the lack of data. The applied calculation technique is an interpolation method. The algorithm was designed to compute values between points. Littenseradiel is surrounded by points, providing a reliable interpolation. At the edges, there is no interpolation but an extrapolation. This is less reliable and therefore those regions have been marked with a hatch.

The value of a trend surface is relative. All kinds of parameters can be adjusted to enhance the map image. It is possible to change the focus of the map by using a visualisation technique. Map 1.6 is based on exactly the same data as map 1.5, including the trend surface. The only differences are the settings of the grey tone scale. The different settings of the grey tone are accounted for in the legend. The grey tone scale in map 1.6 runs from fully black for 100% \(<c>\) to fully white for \(\leq 80\% \,<c>\). In map 1.5 the boundary values are 90% and 10%. The focus of map 1.6 is therefore on the region with almost exclusively plurals of ‘son’ without an \(<o>\). The less significant contrast in the north-east is thus encompassed.
Map 1.6: Adjusting the trend surface’s grey tone scale.

Map 1.7: Complete map of the plural of ‘son’ in the 15th and early 16th centuries.
Map 1.7 shows the combination of the maps 1.4 and 1.5. The adjustment made here is that the grey tone range is the same as in map 1.4, but the most extreme value has a lighter grey tone, to make the pie charts more legible. The trend surface shows the general tendency of the feature, in this case the root vowel. The pie charts display the underlying data, including the absolute number of attestations, reflecting an indication of the reliability.
1.3.6 Original charters and copies
When dealing with spelling practices on the assumption that they somehow reflect the phonological system of the authors, the original documents are used for initial evaluation. Old and contemporaneous copies (copies made at the same moment as an original charter for administrative use) have also been considered as ‘originals’. In a few instances where very little material is available, other copies have been studied to get an impression of the possible variation. This problem is particularly apparent in the oldest period.

Christian Schotanus
In 1658, Christian Schotanus published several charters in the Tablinum of his “De Geschiedenissen Kerkelyck ende Wereldlych van Friesland Oost ende West.”. Three of the twelve charters from the 14\textsuperscript{th} century are found in his material, including the oldest charters: OFO I-1 from 1329, OFO I-2 from 1386 and OFO II-2 from 1379. Three more charters from the time frame before 1430 are known from the text editions provided by Schotanus. These charters were not copied by a more or less contemporaneous clerk who took an interest in the juridical content. For Schotanus, who lived more than 200 years later, adapting the language of the charters to his own language was not a serious option, as was the case for a copyist in for example 1430, when copying a piece from 1390. The Schotanus text editions are rather philological. The consistent archaic language form in the texts, without any hyper-correct exaggeration, supports this point of view. Unfortunately, copies made by Gabbema in the 17\textsuperscript{th} century do not show the same reliable character. The information from the charters published by Schotanus are regularly included in the tables and graphs.

Copying practices
Comparing spelling in the originals with that of the copies tells us something about the way spelling tended to be altered during the copying process. When studying the spelling of a copy, one has to wonder whether it was the spelling / language of the original author or was it adapted to the language of the copyist? Or a mixture of both? The following example illustrates the strategies of copyists towards archaic language forms in the originals. Before 1460 and after 1490 the common endings of \textit{seke} ‘case’ were:

<table>
<thead>
<tr>
<th>Type</th>
<th>1460</th>
<th>1490</th>
</tr>
</thead>
<tbody>
<tr>
<td>singular</td>
<td>-e</td>
<td>ø / -e</td>
</tr>
<tr>
<td>nom./acc. plural</td>
<td>-a</td>
<td>-en</td>
</tr>
<tr>
<td>dative plural</td>
<td>-um / -em</td>
<td>-en</td>
</tr>
</tbody>
</table>

Table 1.7: Case endings of \textit{seke} in the 15\textsuperscript{th} century.
The difference between the plural forms on the one hand and the singular on the other is:

- The endings -um and -a were replaced by -en after around 1460/1490;
- The ending -e remained a geographical variant of [ø] in the 16th century: *seek / secke*. In other singular nouns, the ending -e also remained very common.

Graph 1.3 shows the spelling of the dative plural, the nom./acc. plural and the singular forms of *seke* in originals and copies. For a copyist in the period after about 1470, the archaic plural forms were odd and perhaps even confusing. The proportion of these archaic endings in the copies of old charters is half, or less that of the original ones. That means that copyists tended to replace them by modern forms. In the singular, the old form <secke>, instead of the modern <seck>, might have been noted as ‘uncommon’ or dialectally marked, but it was an acceptable contemporaneous form. Therefore, there is no systematic replacement of the old form by the new one. So copyists seem to have had different (unconscious) strategies towards features that were outdated at the time of copying.

Graph 1.3: Archaic endings in original charters and copies. The dative plural of *seke* was written with the archaic ending -um (-e/im) in 60% of the relevant tokens from the period 1430 to 1460 in original charters, but in only 30% of copies made of charters from that same period (note: the copy was made later!), etc.

---

23 Note that [ø] and /ø/ represent a rounded, mid-high, fronted vowel, but a single ø equals ‘zero’/empty/nothing.
and those, that may have been marked in specific cases, but still fitted into the morphological and phonological system of the language.

An extensive study of the language in originals and copies will probably produce a more detailed impression of the impact of copying on the alternation of the spelling. One possibility of such an investigation is that some copies could be included in the counting of the ‘originals’, broadening the basis for quantitative analysis.

Section summary:

- For historical analysis, tokens, preferably from original charters, are used;
- Some old copies and unique text editions from the 17th century are also used;
- The linguistic characteristics of a copy can be the product of both the original and the characteristics of the language at the time of copying.
A.P. Versloot: Mechanisms of Language Change

1.3.7 Data size and statistical reliability
1.3.7.1 The implications of counting features from an historical corpus

The charter corpus has a fixed size. Historical data cannot be expanded by additional surveys. It is, therefore, extremely important to estimate the reliability of the patterns observed, especially when the absolute number of attestations is limited. In § 1.3.2 and subsequently 1.3.4 it was claimed that the original charters can be used to make historical and geographical reconstructions of 15th and early 16th century Frisian language. This section is concerned with the statistical aspects of the data’s reliability.

Corpus linguistics is concerned mainly with frequencies, the relative proportion of variants at a given time, place and in a certain style genre, the shift of the proportions through time, place and styles, etc. The strength of this approach lies in the mass of the data. Ask one person on the street and he has a personal opinion. Ask 200 people and the sentiment of the country is known. If the selection of people who answer the question is random, there is no need to consider an individual’s background. That is the power of statistics, and thus the power of corpus linguistics.

This is not the first quantitative linguistic investigation, and there are other Frisian language studies using corpora. So far historical investigations have tended to be philological, like the Gosses (1928) and Boersma (1939) analysis of a part of the Frisian charter language. In historical Frisian language research this study is innovative given its size and detailed reconstructions of variations in time and space.

This section deals with questions such as: Which characteristics of the data, apart from the philological ones discussed in § 1.3.2, are important for assessing the reliability? How are these expressed with statistical methods? Can the reliability of the corpus be measured, and how reliable are the findings? The reliability of random samples increases with a larger numbers of responses. How many attestations are needed to make valid observations about the variation of language.

---

24 It would be an underestimation of the efforts made by Gosses and Boersma not to mention a few words on their research. Their research is compatible from a conceptual point of view. The authors are very much aware of the historical dynamics of their material and this author shares Gosses’ opinion on the phonological reliability of the data (Gosses, p.11-12). Where this study covers the total language area, they consider it to be an advantage to have only charters from a limited radius to avoid dialect mixture. Whilst this author agrees with many of their observations, extensive reference to their results would take considerable effort (their publications make rather difficult reading) and the harvest would be limited. This study relies on observations carried out during the course of this particular research. Comparison with the work of Gosses and Boersma is left to others.
A.P. Versloot: *Mechanisms of Language Change*

in time and space? An example of these questions is as follows:

In § 2.4.3.2 (graph 2.6), the example of *habbath* is used. In the first time frame 1379 to 1400, 50% of the relevant tokens use the ending `<at(h)>` while 50% take `<et(h)>`. The percentages are based on eight tokens from five original charters. What is the distribution of tokens with an `<a>`? Are they all concentrated in one or two charters and what would it mean to the overall picture if precisely those charters were missing by a fate of history? Is the data evenly dispersed over the charters? Should the tokens be counted per charter or per token? What does it mean for the statistical reliability? And what do the observed patterns of distribution tell us about the linguistic reality behind the tokens?

1.3.7.2 The concept of ‘random’

The first prerequisite of reliable statistical outcomes is that the data form a random selection and hence a representative selection from reality. The second one is the demand for a sufficient number of observations. Even when the selection is random, the number of observations must be enough to capture all variations. Throwing a die once is a completely random act, but this single act is not enough either to capture all the variants (‘1’ to ‘6’) or to get a realistic impression of their relative importance (each 1/6). Does the charter corpus represent a representative selection for the purpose of linguistic reconstruction and does it offer a sufficient number of observations?

What does the concept of ‘random’ imply? Random means that the way the selection of examples were drawn from the total population did not systematically favour any of the available variants. The question can be split into two sub-questions:

1) Is the charter corpus a random and representative selection of all the charters that ever existed?
2) Are the tokens in the charters a random and representative selection of the language at the time?

1.3.7.3 Random preservation

The first question is difficult to answer, because it is not known how many old charters were lost.\(^{25}\) However, decay and fire, the main enemies of old charters, are

---

\(^{25}\) It may be possible to get some indication of it, by looking at inventories of charters from the earlier centuries, some of which have been preserved. A systematic comparison has not yet taken place, but the overall impression of O.Vries, who identified some of these lists, is that the preserved charters are only a fraction of those which once existed (kindly communicated to the author by O. Vries).
not interested in spelling variations, i.e. there seems to be no reason why some spelling forms would be favoured by the loss of charters. The only possible example are three of the oldest charters, published by Christian Schotanus in 1658. They were (indirectly) preserved because they were old. Old means linguistically archaic and from that point of view, the total of the corpus has become slightly more archaic because of this special antiquarian interest of Schotanus. One may argue the language (Frisian, Latin, Dutch) affected the preservation or destruction of the charters. However, this influences the numbers of charters in different languages. There is no apparent reason why charters, for example, spelling <habbath> would be destroyed and those spelling <habbet> preserved. The fact that <habbet> is a more recent form and that time influences the number of charters preserved is another issue (see later).

One sub-aspect of preservation which does influence spelling forms is copying. Many charters have only been preserved as copies, often from the same century. Older charters whose content was of special interest to users could be copied for further use. A copyist, for instance, could favour a given spelling or adjust them to his own language practice (cf. § 1.3.6). For copyists from the 15th and 16th century, the language of the older charters was basically regarded as the same language as their own. This meant that re-spelling the text was an improvement rather than a distortion, for example, eliminating archaic forms that could affect the comprehensibility of the text. The old charters edited by Schotanus are not considered to be ‘copies’ but rather as old text editions. Using (text editions of) original charters avoids the problem of selective re-spelling by copyists.

On the whole, there seems to be no reason why only spelling forms would show any relation with the probability of preservation. This means that in this respect the corpus of original charters may be regarded as a random selection.

1.3.7.4 Random selection from language forms
Putting all forms from the corpus together would still not constitute a representative selection of the language from the charters. The corpus exhibits major changes over time and space. In graph 1.1, there are only two tokens of a dative plural of seke in the first time frame up to 1430, but 13 between 1490 to 1510. In the total corpus, the more recent language forms are over-represented. To overcome this, the corpus is cut into shorter time frames, as was done in graph 1.1 and 1.2. The total number is 100% per time frame and the different variants are computed as a proportion of that. Graph 1.1, depicting the temporal development of the dative plural ending -um (-em / -im) reveals a clear temporal skewness, graph 1.2, showing the ratio of plural forms of ‘son’ with <c> and <i> or <o> as the root vowel, shows stability over time.
Also, the geographical distribution is not balanced. No less than 30% of the original charters are connected to Leeuwarden while the next 11% of the charters are from Bunnik (cf. for references to municipalities here and in the forthcoming sections, map 1.2). There is no direct link to population density, even though charters from scarcely populated south-eastern regions are rare.

A third source for distortion is the combination of time and location. The oldest preserved charters are predominantly from the centre and north-east of Fryslân. The oldest original charter from the north-east region is from 1390. The oldest one from the South-West (Gaasterlân, Nijefurd, Lemsterland) is from 1443. For most municipalities, there are original charters from the 16th century, so geographical distortion is especially prevalent in the earliest periods. The number of original charters per region up till 1430 ranges from 15 in the North-East to nil in the South-West (map 1.8).

Map 1.8: Mapping regions and number of original charters until 1430. In the text, for example, 'South-West' (with capitals) refers to the specific region delineated in this map; 'south-west' is a general geographical reference.
Dialect maps can correct unbalanced geographical distributions. However, maps can suffer from temporal imbalances in the distribution of the charters over the regions. How this is counterbalanced is illustrated in map 1.9, showing the spread of <a> spelling in the ending of *bitaliane* 'to pay (gerund)' and *bitalad* 'paid (past part.)'. Variation can only be studied over the period it occurred. The last attestation to <a> in the ending in an original charter is in OFO IV-33 (1468): <bitellath>. Therefore, the selection is limited to the period before 1470. To control temporal distortions in geographical distribution, an average year of attestations per municipality is calculated. These are the years shown in map 1.9. The average is calculated over the tokens, not over the charters amounts. Map 1.9 shows the average date of the attestations from the pie chart for Nijefurd, with relatively young charters, as being 1456. The pie chart for Dongeradeel is based on tokens with an average date of 1454: one <a> from 1431, one <a> from 1465 and one <e> from 1467. It is no surprise that the oldest attestation is with an <a>, but the two tokens from 1465 and 1467 are entirely contemporaneous with attestations.
from the south-west. On average, most points lie between 1450 and 1460, so the map can be regarded as reflecting the situation from the middle of the 15th century. The data has also been checked for average years of spelling <a> or <e>. In Leeuwarden for example, the average for tokens with an <a> is 1454 and 1452 for the tokens with an <e>/<i>. This may serve as an example that temporal differences play no role in map 1.9.

Graph 1.2 is an example where the spread of the data had no influence on the variation. In that case the map can safely be drawn from data from the entire period. Assume a situation with variant A in the South-West and B in the rest, with no shift through time. The under-representation of charters from the South-West in the oldest time frames would suggest an increase of variant A in graphs of type 1.1 and 1.2. This can be encompassed by a proper dialectal evaluation of the data. The South-West is represented in the later time frames. Wûnseradiel and Bolsward, not entirely in the south-west, but often 'south-western' in their language, are both well represented, Wûnseradiel with even the next to oldest charter.

In the relation between changes in time and space, the former appears to be dominant. The tendency of the changes is similar in the whole of Fryslân: habbath ‘(we/you/they) have’ finally becomes babbe and door ‘door’ changes to doer. Only in a few cases do the geographical differences become (semi-)permanent contrasts, but the dialectal variations are mostly a matter of differences in timing. In one area a change may appear earlier or later, but the tendency is the same. In a few cases the trends are divided up per region, such as the case of replacing -a with -en in the feminine plural form in sekä (map 2.10).

The conclusion is that temporal and geographical variations have to be taken into account in order to balance distortions in the charter corpus.

1.3.7.5 Central concepts: variation and variance

In the previous sections the following conclusions were drawn:

- Given the absence of an independent Frisian spelling tradition in the 15th and 16th century, the spelling is likely to be highly phonetic/phonological, applying Latin or Dutch spelling conventions of that time (§ 1.3.2);
- The corpus of original charters is probably random for the aspect of spelling (§ 1.3.7.3);
- The corpus is not random for temporal and geographical variation, but can be corrected for skewness that is the result of such variation (§ 1.3.7.4).
The next step is the question of statistical reliability. Selection of tokens from the corpus may be ‘random’ and the spelling may represent something ‘phonological’, but are the trends in the material also ‘real’, or is the corpus too small and the observed trends only a result of random variations? Key concepts in this discussion are variation and variance.

1.3.7.6 Variation in time and space: testing the time trend

For both the variation in time and space, statistical reliability can be computed using standard statistical tests. The χ²-test of goodness of fit (further referred to as the ‘χ²-test’) provides the probability that:

- A given distribution deviates from an expected one by chance;
- Two distributions deviate from each other only by chance.

Further reading on the χ²-test of goodness of fit can, for example, be found in Moore & McCabe 2003, 620 ff.; Field 2005, 682 ff.; online information, plus several practical on-line calculators are found on: http://faculty.vassar.edu/lowry/webtext.html.26 Whenever the χ²-test is used to detect significant patterns in data, the details are presented in the appendix 1.

To illustrate how the χ²-test works, the following example is provided: Ten observations distributed over two variants are provided with a default hypothesis of 50% chance for the occurrence of either variant, for example, two sides of a coin. That means the most likely number of observations is five per variant. Assume the following observations of variant A and B:

<table>
<thead>
<tr>
<th>Variant</th>
<th>Expected</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>3 (30%)</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>7 (70%)</td>
</tr>
</tbody>
</table>

The χ²-test outcome of $p = 20.6\%$ expresses the probability that this skewed observation sample (i.e. 3 vs. 7 instead of 5 vs. 5) or an even more distorted one may occur, given the assumption (null-hypothesis) that the underlying probability is 50-50% (Moore & McCabe 2003, 441). This probability, expressed in

26 Another useful calculator is the built-in function of Microsoft Excel: =CHI.TEST (observed;expected), returning the p-value of the observed distribution.
percentages or in decimals, is the \( p \text{(robability)-value} \) of the test. The threshold of the \( p \)-value, the so-called \textit{significance level}, is usually positioned at 5\% or lower, depending on the aim of the study (Moore & McCabe 2003, 462). The computed \( p \)-value of 20.6\% is not low enough to interpret skewed observations as a significant deviation from the null-hypothesis. Because there are so few observations, the deviation between ‘observed’ and ‘expected’ is not significant in this example.

In the following example, the number of observations is three times larger (30 instead of 10), but the relative proportion of the observations A and B is still 30\% and 70\%:

<table>
<thead>
<tr>
<th>Expected</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variant A</td>
<td>15</td>
</tr>
<tr>
<td>Variant B</td>
<td>15</td>
</tr>
</tbody>
</table>

\( \chi^2 \)-test: \( p = 2.8\% \)

Due to the larger sample set, the observed distribution is now significantly different from the expected 50-50\% distribution. With the \( p \)-value < 5\%, the test tells us that it is unlikely that the underlying probability on variant A or B is indeed 50\%. However, it does not tell us what the underlying probability is, nor how great the probability is that it would be 30-70\%. In the rest of this study there are two \( \chi^2 \)-test applications:

- A specific, observed distribution is evaluated against a probability established in another dataset. This is the same type as the previous example;
- A set of observations on for example, variation at the ending of the dative plural in the charters, presented in table 1.6. and graph 1.1, can be split according to a relevant linguistic feature, such as ‘long root’ versus ‘short root’, or according to time, such as ‘data older than 1460’ versus ‘data younger than 1460’. Taking the latter example, it is possible to pose the question: is the difference in the relative distributions of the different endings for the two sub-sets the result of chance (the null-hypothesis), or is the variation, at least partly, the result of the factor time? The null-hypothesis states that both endings appear in a similar proportion in both time frames. The actually observed attestations for the contrasting endings are compared with each other and with the average, based on the sum of both amounts. Also here, the \( \chi^2 \)-test returns a \( p \)-value that can be
A.P. Versloot: *Mechanisms of Language Change*

compared with the defined significance level of 5%. In this case, the p-value is 0.0002% (see the appendix 1 for the exact figures).\(^{27}\) In this case it is possible to conclude that time plays a role, but the p-value is not a measure for the impact of the factor time. There are different measures for this. One of these will be discussed later in this section.

The latter type of comparison with the help of the chi\(^2\)-test is predominant in this study.

To use the chi\(^2\)-test properly, several limitations need to be taken into consideration. Two of them are:

- Number of attestations / observations per cell;
- Number of cells.

There is a problem with the reliability of the chi\(^2\)-test for very small numbers of observations, which is regularly the case while working with historical data. The average of all cell values should be five or more and the smallest expected count should be one or more (Moore & McCabe 2003, 626). For small numbers, there is an alternative test, the *Fisher Exact (Probability) Test* (Nijdam & Van Buuren 1980, 481; http://faculty.vassar.edu/lowry/webtext.html, sub-chapter 8a, including an on-line calculator). Where relevant, the chi\(^2\)-test and Fisher’s Exact Test were both applied, mostly with the same conclusions regarding statistical significance. The outcomes are included in the appendix 1.

For 2 x 2 tables, the measure of \(\chi^2\) is structurally too high, according to some statisticians. They suggest a correction of the \(\chi^2\)-value. This is the so called Yates’ correction. This correction is discouraged by others (Field 2005, 686). In the appendix 1, the test parameters of \(\chi^2\) and p-values are provided both with and without Yates’ correction. In most instances it does not affect the interpretation of the test.

The statistical measure of correlation is illustrated with the same dataset of the dative plural of *seke* ‘case’ (table 1.6. and graph 1.1). The chi\(^2\)-test suggested that time plays a role in the distribution of <um> (<<em>/im>>) versus <en> or <a> with a p-value <0.1% for the contrast before or after 1460. But how strong is this time effect? The chronological trend can be expressed in another way, by showing the correlation with a regression line that symbolises the trend (Moore & McCabe

\(^{27}\) When the p-value < 0.1% (0.1% = 1:1000), the exact value is not given, but simply “<0.1%”
In this example, a linear regression line is computed, shown with a dashed line in graph 1.4, i.e. a straight line from 100% to 0%. The regression line expresses the assumption that the application of the ending <um> (<em/im>) declined steadily between 1420 and 1500. The $R^2$ (the coefficient of determination) expresses the level of variation in the real data that is explained by the regression line (Fields 2005, 148; Moore & McCabe 2003, 144). The possible value ranges from 0 = no trend, to 1 = exact match. The value of 0.94 is very high and confirms the direction of the chronological trend.

But this is not the whole story. The coefficient of determination may be very high, but we must consider the option that the observed (high) correlation is the result of chance (Field 2005, 126). In particular in the case of a low number of points, this is a crucial aspect. In this case, we may assume that the Old Frisian ending -um is gradually replaced by the reduced ending of -en, so the line in graph 1.4 is expected to go down. In other instances it may be of interest to know if there is any trend at all. The former assumption asks for a so-called one-tailed test, the latter, a two-tailed one (cf. Field

---

28 In the extreme case of only two points, there is always an $r^2$ of 1. But the reliability of the correlation is very low given a random positioning of the first point, it depends only on the second point.
In the material presented in this study, the choice between one or two-tailed test rarely resulted in a different conclusion regarding the significance of a correlation. In the appendix 1 both p-values are given. This is not covered in any further detail here.

In the case of the dative plural ending in graph 1.4, the p-value for a one-tailed test is 1.6%, so it may be assumed that the observed trend is not the result of mere chance. The high coefficient of determination ($r^2$) of 0.94 shows that 94% of the observed variation in endings is explained by the factor time.

There is one more aspect of this type of correlation that should be mentioned here, namely that this correlation is based on averages. In Moore & McCabe (2003, 165-166) this topic is mentioned in an example about the correlation between age and the mean height of children. As individual children of the same age show variation in height, the correlation between age and the average height per year group will be higher than the correlation between the age and height of every individual child. In the data in this study, this problem is more complex to overcome. The underlying corpus data is discrete and often binary. A token has the ending <um> (or <e/im>) or it does not. As a test, the underlying binary data of graph 1.4 is plotted with the year of the charter date on the X-axis and the binary score 1 = <um> (or <e/im>), 0 = other on the Y-axis.

Graph 1.5: The correlation between binary raw data and time. Data from the dative plural of seske (cf. graph 1.4).

When using the individual data, the coefficient of determination falls to 0.51, which is still a strong correlation (cf. Field 2005, 32, where $r^2 = 0.25$ is already referred to as a “large effect”). On the other hand, the number of observations increases from 4 to 39. The p-values (one- and two-tailed) are both < 0.1% (Note: this example is not in the appendix 1). This has little effect on the interpretation. There is still a significant and substantial correlation between time and the reduction of the dative plural ending in the word seske ‘case’. Also, here, the data
suggests a transition period between 1420 and 1500. The problem of averages versus individual data is not covered further here.

In summary, the chronological dispersion of the tokens is not merely a coincidence (confirmed by the chi²-test). The trend is definitively explained by a linear downward movement ($r^2 = 0.94$). As a result it is likely that the dative plural ending *-um* (*-em, -im*) disappeared from (written) West Frisian language between 1420 and 1500.

For the root vowel of ‘sons’ (graph 1.2), the probability that the observed deviations per time frame are purely coincidental is very high, indicated by a p-value of 94% (chi²-test). The regression line gives an $r^2$-value of 0.1. The likelihood of this weak trend was tested with a two-tailed test. The p-value is 90.1%. Therefore, the computed correlation of 0.1 is most likely the result of chance. Both statistical tests confirm the absence of a temporal trend in the distribution of <e> and <o> in the root vowel of ‘sons’, at least during the period 1431-1547.

1.3.7.7 Testing the spatial trend
It is possible to verify the statistical reliability of not only time trends but also geographical trend surfaces (cf. map 1.7 and 1.9). Map. 1.7 is based on the percentages of plural forms of ‘son’ with an <e> or <i> per municipality. In the following table 1.8, the observed values are compared with computed surface values:

---

29 Note that the oldest attestation in the data of an other ending than <um> (or <e/im>) is from 1456: <secka> OFO II-40: “...ende hierra frionden toe wessen in needlika secka.” This illustrates a disadvantage of grouping data in classes. This attestation and another one from 1458 fall within the class ’1430-1460’. Different class boundaries would give different averages. What is important here is the fact that both the correlation based on class averages and the one based on individual data provide similar results. The outcome of the analysis based on the individual data also gives another clue: if the reduction of the dative plural ending was a linear process, it seems to be mere chance that there are no attestations from an earlier date than 1456. A more common model for transitions in languages than a linear trend is the so called S-curve (cf. Ke 2004, 216-218).
Table 1.8: Deviation of actual point values from the trend surface value for the ‘sons’ map. Rows with only one token are in italic, rows with high deviation from the surface value are in grey. Table ordered by Surface value.

<table>
<thead>
<tr>
<th>Municipality</th>
<th>Point values</th>
<th>Surface value</th>
<th>Point - surface difference</th>
<th>Within standard deviation</th>
<th>Number of tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyt</td>
<td>0%</td>
<td>34%</td>
<td>-34</td>
<td>no</td>
<td>1</td>
</tr>
<tr>
<td>Ldl</td>
<td>0%</td>
<td>39%</td>
<td>-39</td>
<td>no</td>
<td>2</td>
</tr>
<tr>
<td>Ldn</td>
<td>59%</td>
<td>46%</td>
<td>13</td>
<td>yes</td>
<td>17</td>
</tr>
<tr>
<td>Men</td>
<td>50%</td>
<td>52%</td>
<td>-2</td>
<td>yes</td>
<td>2</td>
</tr>
<tr>
<td>Boa</td>
<td>50%</td>
<td>64%</td>
<td>-14</td>
<td>yes</td>
<td>4</td>
</tr>
<tr>
<td>Don</td>
<td>100%</td>
<td>65%</td>
<td>35</td>
<td>no</td>
<td>1</td>
</tr>
<tr>
<td>Fra</td>
<td>91%</td>
<td>71%</td>
<td>20</td>
<td>yes</td>
<td>11</td>
</tr>
<tr>
<td>Hee</td>
<td>100%</td>
<td>84%</td>
<td>16</td>
<td>yes</td>
<td>1</td>
</tr>
<tr>
<td>Bol</td>
<td>100%</td>
<td>92%</td>
<td>8</td>
<td>yes</td>
<td>3</td>
</tr>
<tr>
<td>Ska</td>
<td>100%</td>
<td>92%</td>
<td>8</td>
<td>yes</td>
<td>1</td>
</tr>
<tr>
<td>Wym</td>
<td>100%</td>
<td>95%</td>
<td>5</td>
<td>yes</td>
<td>1</td>
</tr>
<tr>
<td>Wun</td>
<td>100%</td>
<td>97%</td>
<td>3</td>
<td>yes</td>
<td>2</td>
</tr>
<tr>
<td>Nij</td>
<td>100%</td>
<td>100%</td>
<td>0</td>
<td>yes</td>
<td>3</td>
</tr>
</tbody>
</table>

The assumption is that the computed trend surface is an idealised representation of the real variation, whereas individual point values may exhibit a stochastic deviation from that position. The difference between values in the points and trend surface values in column four is an indication of the accuracy of the data and its approximation by the trend surface. The standard deviation of the differences between actual point values and surface values is 21%. The isolated points in the north-east (*Tytsjerksteradiel*, *Leeuwarderadeel*, *Dongeradeel*) fall beyond the standard deviation. These point values are based on a limited number of tokens. This is in part a coincidence, because the computation of the trend surface was not weighted by the number of tokens per point. Weighting the token frequency would not help this map, because the points for *Tytsjerksteradiel* and *Dongeradeel* are far removed from other points with more tokens. Weighting is only useful in a map such as 1.9, where there are points with few attestations and extreme values (for example, *Menaldumadeel* 100%, *Leeuwarderadeel* 0%) near to points with more attestations and more intermediate values. A low token frequency does not necessarily lead to a high deviation from the trend surface. This can been seen in table 1.8 in the points of the south-west, for example, *Wymbritseradiel* and *Skarsterlân*, where both have one token and a deviation from the trend surface of only 5% and 8% respectively.
The reason for different levels of deviation is the relationship between (assumed) real values and token frequency. We can assume that in the reality of the time, the percentage of forms with an <e> was ± 40% in the north-east (cf. Leeuwarden 46%)\(^30\), but 0% in the south-west. With one token, the observed point value can only be 0% or 100%. This is close to the expected values of the south-west, but automatically causes a large deviation (40% or 60%) in the north-east. Points with only one token will therefore show a higher deviation in the north-east than in the south-west (cf. following section on variance).

The correlation between point and surface values is depicted in graph 1.6. The relevance of the spatial trend surface can also be tested against the hypothesis that values are a stochastically defined pattern of spatially homogeneous data with an average of 71% with an <e>. It is possible to test the correlation between the observed points and the surface graph and this constant average value. The

\[ \text{Graph 1.6: Point and computed surface values for the 'sons'-map} \]

\[ \text{(map 1.5 and table 1.8)} \]

\(^{30}\) It is difficult to obtain more accurate information on the situation in the north-east at the time. In the singular, forms with an <e> appear in the entire language area in a small proportion of ± 30%. In the modern dialects, Terschelling has sin, plural sinnen. In the eastern dialect of Terschelling, that form is archaic and competes with soan(s) (Roggen 1976, lemma sin, Knop 1954, 34, who already noticed the different distribution of forms with <o> over singular and plural in the charters). Bogerman (Dongeradeel ± 1540) has a singular <soen, soon> and no plural attested. Modern Schiermonnikoog dialect has seen, pl. seenen and sein, pl. seenen/-s (Visser/Dyk 2002). The latter form can be the modern representation of the Bogerman form (cf. Modern Schiermonnikoog dialect bras < OF braas 'sock'), but also a Hollandic loanword (cf. Dutch dialects from Holland: sein). The Schiermonnikoog form sein shows that forms with an <e> were indeed present in the north-east. The old singular - plural opposition in the root vowel is lost in every modern dialect.
correlation between the point values and the trend surface is 0.88. The correlation between the point values and a constant average is 0.21. Not only is the correlation between the point values and the surface much higher than between the point values and an average value, but also the probability of the former is much greater, $p < 0.1\%$, for the trend surface, compared to $p = 24.3\%$ for the average value. This can be read as: the arithmetic correlation between the point values and the average is only 0.21 (on a scale 0 - 1), while it is most likely (24%) that this correlation is the result of chance. The geographical trend surface is a much better approximation of the spatial variation.

To overcome problems with low frequency in the north-east, data can be aggregated at the level of the regions (cf. map 1.8). This produces map 1.10. The trend remains the same as in map 1.7, but the curious ‘high’ in Dongeradeel, based on strongly dispersed values in the north-east, has now disappeared. The standard deviation for the differences between point and surface values has been reduced from 21% to only 9%. Graph 1.7 shows the relation between the point values and the trend values.

Map 1.10: Preference for the vowel $<$o$>$ or $<$e$>$/i$>$ in the plural of ‘son’, computed from data aggregated at regional level.
The correlation between the surface values and the point values is now 0.94. The correlation with the constant average of 71% has increased to 0.53, due to the fact that extreme deviations in Leeuwarderadeel and Tytsjerkeradiel (-34% + -39%) have now disappeared. For the correlation with the trend surface, \( p = 0.2\% \), for the correlation with the average, \( p = 14.1\% \). Given the observed distribution, a geographical trend is highly likely and the computed surface values are a better approximation of the observed token frequency than the assumption of a monotonous spatial pattern.

The following paragraph briefly examines the geographical pattern of map 1.9. Spelling with an \(<a>\) appears in 33% of the tokens in the relevant time frame.

The correlation of all the individual point values with the surface values is 0.59 (\( p = 1\% \)) and 0.53 (\( p = 2\% \)), with the average value of 33%, (data not in the appendix 1). This is due to the fact that there are several points with one token (6x) or 2 (2x) with extreme point values (0% or 100%). This results in a strong deviation in the trend surface, as described for ‘sons’. Therefore, the trend surface has been based on a weighted calculation, where the influence of the points with few tokens was reduced. When leaving out the points with only one token, as shown in graph 1.8, the correlation between point values and trend surfaces increases to 0.74 (\( p = 1.1\% \)), while the correlation with the average remains on 0.57 (\( p = 5.4\% \)). When aggregating the data per region (n = 6, df = 4), the correlation between observed values and trend surface values is 0.92 (\( p = 1.0\% \)), while the correlation between observed values and the average value is only 0.28 (\( p = 29.7\% \)) (data not in appendix 1).
The conclusion from these two examples is that computed geographical patterns can be statistically significant at the level of both region and municipality. Individual percentages of points with only one or two tokens within regions with an expected percentage not close to 0% or 100% must be viewed cautiously. Clusters of points with low token frequencies, whose values are confirmed by trend surface and the adjacent point values, are a solid indication of a statistically significant trend. This is the case in map 1.7 and 1.9 in the south-west of Fryslân, for example. These two cases are based on 49 (‘sons’) and 61 (‘<a>’) tokens. All the maps presented in this thesis are based on a similar, and often higher number of tokens, increasing the statistical significance. The statistical significance is not explicitly evaluated for every map or temporal graph, unless special circumstances, such as low frequencies or odd distributions, give cause for this.

1.3.7.8 Variation and variance: statistical context
Apart from variations in time and space, a phenomenon can show variations in a time frame and from within a limited region. Map 1.9 shows that speakers from the north-east used an <a> when spelling bitalad/bitalane approximately 50% of the time. When split up into time frames and regions, there are often only a few relevant tokens per item. The impact of chance or a mistake in the searching script increases significantly. One token can make a significant difference to percentage points. What is the best way to count the data from the charters? Per token or per charter? Counting per token increases the impact of multiple attestations in long texts. Counting per text increases the influence of solitary occurrences in texts. The solution to this question lies in the concept of variance.
Variance is an indicator of the level of difference between individual appearances of a variable. This can be best illustrated by the following example: At a given point in time and space - (for example, in the middle of the 15th century in the language of Leeuwarden and close surroundings) two alternative forms compete with each other and have an overall appearance frequency, assuming that this can be determined, of 1:2. There are two models of reality which could have such a proportion of forms as an outcome:

Model I: One-third of the authors uses variant A and two-thirds variant B. This implies that the variance per author is minimal, for instance, completely predictable after the first observation in a text, but for the corpus, maximal: Charters from three different authors are needed to establish the 1:2 ratio, but per author, one form is sufficient. In this model it is preferable to count per text or even per author;

Model II: Every author uses both alternatives in the ratio 1:2 in a random alternation in his writings, (for example, within one charter). If this is applied consistently, the individual variance equals the population variance. With a sufficient number of tokens to track the variation with a significant estimation, it is possible to deduce the variation among the entire population. If the inter-speaker variance is zero, it does not matter whether there is a long text from one author or two shorter ones from two authors, assuming that both are representative for that period and region. In this model, token count is the preferred method.

Traditional dialectology looks for the ‘real, authentic dialect form’. Generally there is only one. Alternation is regularly interpreted as ‘dialect mixture’ or ‘intrusions’ from other dialects or the standard language, factors that ‘disturb’ the ‘pure’ dialect.

A rule-based grammar model generally produces one correct output form. With optional rules or local free rule ordering, the existence of variants can be modelled into language. These kind of variables and optional rules, which produce a stochastically fluctuating performance, where likelihood is for example, controlled by sociolinguistic factors, is not really the core business of traditional rule or constraint-based grammars. In a sociolinguistic framework, variations in speech by one speaker is a rule rather than an exception and may have semantic implications. By choosing a specific level of variation (archaic - modern, low - high, regional - standard, etc.), the speaker can achieve communicative goals. Seemingly stochastic variation can become meaningful in a sociolinguistic framework.
What are the implications of the two models in the way data is interpreted in an historical corpus? Assume a feature X with two alternating forms A and B. In this example both forms show a 50-50% distribution among speakers of one contemporaneous speech community. Suppose the existence of a mini-corpus containing two charters with three tokens per charter each, exhibiting feature X. The total number of tokens is six. Independent of the models, three attestations to A and three to B would be expected.

In model I, each charter shows only one variant. This is either 3x A or 3x B, as each author complies with his own ideal dialect/grammar. In this model, it is necessary to count the features per charter. In fact, it is not relevant how many tokens are found per charter/author, because each author uses only one form. The chance of finding 50-50% distribution in model I is 50%. However, there is also a 50% chance that one finds two charters/ six tokens with only variant A or B (cf. table 1.9).

<table>
<thead>
<tr>
<th>Number of tokens A</th>
<th>Number of tokens B</th>
<th>Model I</th>
<th>Model II</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6</td>
<td>25.00%</td>
<td>1.56%</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.00%</td>
<td>9.38%</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.00%</td>
<td>23.44%</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>50.00%</td>
<td>31.25%</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0.00%</td>
<td>23.44%</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.00%</td>
<td>9.38%</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>25.00%</td>
<td>1.56%</td>
</tr>
</tbody>
</table>

Table 1.9: The probability of finding different distributions of variant A and B in feature X in two texts with three relevant tokens each and a stochastic probability of 50% for each of the variants. In model I, an author only uses one variant, in model II the author varies according to the given stochastic probability in every instance of feature X. Take note: In model I the probability is 25% that variant A is found 0x and variant B 6x. In model II the probability of this distorted distribution in the observations is only 1.56%, etc. The probability of finding the ‘real’ 50-50% distribution (3x A and 3x B) is 50% in model I and 31.25% in model II. The probability of finding a completely wrong impression of only A or only B is 50% in model I and only 3.12% in model II.

Table 1.9 shows that the probability of finding the ‘ideal’ value of 50-50% might be greater in model I. The advantage of model II is that the probability of missing one variant is much lower. In model II the probability of observing a portion of variant A between 34% and 67%, meaning a maximum error of 17 percent points, is 78%, which is reasonably good for a corpus with only two texts and six tokens. Model II is not only closer to the reality of sociolinguistics, if it is a correct model of the corpus, it would increase the reliability of the observations compared to
counting per charter or author, given the same number of historical material.

<table>
<thead>
<tr>
<th></th>
<th>Text 1</th>
<th>Portion A in 1</th>
<th>Text 2</th>
<th>Portion A in 2</th>
<th>Portion A in corpus</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model I</td>
<td>AAA</td>
<td>100%</td>
<td>BBB</td>
<td>0%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Model II</td>
<td>ABA</td>
<td>60%</td>
<td>BAB</td>
<td>40%</td>
<td>50%</td>
<td>2%</td>
</tr>
</tbody>
</table>

Table 1.10: Variance of the corpus according to model I (one variant per author) and model II (variation within the texts).\(^{31}\)

What does this imply for the concepts of variation and variance, assuming the corpus indeed shows the expected ratio between A and B: 50%-50%? In model I the variance is much higher and as a result so is the probability of finding biased results.

Corpus variance influences chances of coming to the correct conclusion based on random selections from the attested corpus. There are two additional factors:

- Number of variants;
- Ratio/proportion of the variants.

The greater the number of variants involved, the larger the selection needed to get a reliable impression of the actual variation. To have three variants represented in the material, at least three tokens are needed. To get a valid impression of their distribution ratios, more data may be necessary. This research is mostly concerned with binary variables: \(<a>\) or \(<e>\), one consonant or two, etc. This increases the validity of the observations.

The skewness in variant ratios has two consequences. Getting an accurate impression of the presence of low frequent variants is difficult with a limited number of tokens. If a variant appears in 10% of the cases in reality and there are only three tokens, there is a 72% chance that the low frequent variant is missed and 28% chance that its proportion will be overestimated, showing a presence in the material of \(\pm 33\%\).

\(^{31}\) The variance = the square of the standard deviation. In model I the values for “A” are 1 (= 100%) and 0. The standard deviation is 0.71 and the variance 0.5. For model II, the figures for “A” are 0.6 (\(\approx 60\%\)) and 0.4 (\(\approx 40\%\)). The standard deviation of those 2 figures is 0.14, the variance 0.02.
The mirror effect of the distortion is that a skewed ratio of two variants gives a great chance of finding the correct dominant form. This effect was already shown in the previous section on the reliability of geographical interpretations. In the case of ‘sons’, in the rather homogeneous south-west, the single token observations confirmed the general trend. Given a probability of the variant <e> in almost 100%, single token observations will most likely be 100% <e>. In the north-east, with a probability of about 50%, the chance of getting a strong deviation between low frequent observations and the real proportion is much greater: with one token 100% and with two tokens still 50%.

1.3.7.9 Variant mixture in practice
To be able to decide between the validity of model I or II, a series of words has been investigated for the level of mixture that appears in the charters.

The mixture of forms in one charter is only possible if two conditions are fulfilled:

• The two variants must be co-existing at that time;
• A charter must contain at least two tokens for the relevant form to be able to show different variants in one charter.

An example of seka/seken will be illustrated here in detail. Further information is given in the appendix 1. The question is: Are authors using the old and the modern form in one text? The old and modern endings of seka/seken coexisted between 1460 and 1500. From this period, there are 36 original charters with at least one relevant token, a nominative or accusative plural form of the noun seke. Twenty one tokens have the ending <a> and 30 the ending <en>. Therefore, based on the token count for this period, the mixture of both variants is 41% to 59%. When counting per charter, there are 18 charters with an <a> and <19> with an <en>. Subsequently, at the level of the charters, the mixture is 49 - 51%.

Eleven charters fulfill the condition of a minimum of two relevant tokens. These 11 charters contain 26 relevant tokens (average of 2.4 tokens per charter). In only one charter are both endings applied. This is OFO 1V-75, 1487 from Leeuwarden, with three tokens: Twice <a> and once <en>. This means that only one out of 11 charters (= 9%) that could exhibit a mixture of forms, does so. Counting the tokens, three out of 26 tokens (= 12%) appear in a so-called ‘mixed’ charter.

The results are summarised in table 1.11:
A.P. Versloot: *Mechanisms of Language Change*

<table>
<thead>
<tr>
<th></th>
<th>-a</th>
<th>-en</th>
<th>% mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>All tokens in this period</td>
<td>21</td>
<td>30</td>
<td>41%</td>
</tr>
<tr>
<td>All charters in this period with an attestation to...</td>
<td>18</td>
<td>19</td>
<td>49%</td>
</tr>
<tr>
<td>Tokens in charters with at least two examples</td>
<td>3</td>
<td>23</td>
<td>12%</td>
</tr>
<tr>
<td>Charters with...</td>
<td>1</td>
<td>10</td>
<td>9%</td>
</tr>
</tbody>
</table>

Table 1.11: Quantification of linguistic ‘mixture’

This leads to the following expressions of ‘mixture’:

- An overall ‘token mixture’ of 41% correlates with an actual level within the charters of 12%;
- Counting the charters, 9% of mixed charters (in fact only one example) correlates with an overall ‘mixture’ of 49%.

The following cases have been studied (details can be found in the appendix 1):

- *seke* ‘case’ singular -e or ø
- *seka* ‘cases’ plural -a or -en
- *bitalad* ‘paid’ ending with <a> or <e> / <i>
- *wesa* ‘to be’ root vowel <a> or <e>
- *kapad* ‘bought’ unstressed vowel <e> or <i>
- *degum* ‘days’ dative plural ending <Nm> or <en>
- *habbath* ‘have’ ending with <a> or <e> / <i>

Model II predicts that the level of mixture in individual charters is related to the overall mixture of the language at that time. The closer the ratio is to 50-50%, the greater the probability that mixture appears in individual charters with two or three relevant tokens. Graph 1.9 illustrates this relationship. On the X-axis the observed proportion of the less-frequent variant is used (either per charter or per token). On the Y-axis the actual level of mixture (again in tokens or in charters) is shown. The discussed example of *seka-seken* can be found at the coordinates (49,9) of the charter-count and (41,12) of the token-count in graph 1.9. The graph confirms the formulated expectation that the level of mixture in the charters reflects the proportion of alternative forms in the total corpus. The correlation with a linear
regression is slightly higher in the token count, but the trend is obvious in both methods. It shows that it is almost irrelevant which method is applied.

In the mixed charters, the ratio between the variants is usually between 1:2 and 1:1. There may be two reasons for this:

- Because most mixed charters have only two or three relevant tokens, the proportion of either variant is likely to be $\frac{1}{2}$ or $\frac{1}{3}$. However, in the example of sek(ē) (singular) with 29 tokens in mixed charters, the proportion is still 38% so it is not necessarily the data format that causes this effect.
- Even during ‘mixed’ periods, not every part of the language area was a ‘mixture zone’. While mixture of forms is most probable in actual mixture or transition zones, the proportion of mixtures in those mixture zones was not necessarily the same as the proportions of the two competing forms.

The implication of model II is that “the individual variance equals the population variance”. Graph 1.9 shows that individual variance reflects population variance. The mixture of alternative forms in single charters is not a matter of ‘errors’ but it mirrors the variation of alternatives that existed in the speakers’ community at that time and place.

Graph 1.9: Correlation between the overall presence of variants in the language (expressed in variant-probability, X-axis) and the level of variant mixture in individual charters (Y-axis).
1.3.7.10 Token count or charter count?

Van Reenen (1997) uses Middle Dutch charters to reconstruct dialectal variations in late mediaeval Dutch. Heeroma (1935, 4) uses charters from Holland and adjacent regions to reconstruct the base-level dialect (the oral forms, used by the middle and lower classes) from mainly the 15th century. Both scholars apply a form of charter count. Heeroma’s assumption is that local dialect features are regularly levelled out by supra-regional scribal practices and dialect mixture (Heeroma 1935, 7-8). Incidental appearances are in his approach a key to the ‘real, spoken dialect’. For example: /stein/, spelled <steen>, is the common Middle Dutch word for ‘stone’. In Holland, the form is spelled <stien> for example, in dialect texts from the 17th century from the region of Holland, and is /stin/ in contemporaneous archaic dialects from Holland. When <stien> occasionally appears in some charters from a city in Holland, Heeroma takes that exact feature as the map form. He considers the occasional cases of <stien> in the charters as the manifestation of the local dialect form.

Van Reenen counts every charter with at least one attestation to variant A or B. With this method, Van Reenen stresses the presence of minority forms. This is only attractive when you are rather sure about the location of the charters. It is a weakened effect of the Heeroma method.

Let us consider the outcome of their methods, by evaluating a fictional example, where variant A is a rather common form and variant B a dialectally limited form; both charters are from the same location:

<table>
<thead>
<tr>
<th>Variant A</th>
<th>Variant B</th>
<th>‘Heeroma’</th>
<th>‘Van Reenen’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charter 1</td>
<td>2</td>
<td>0</td>
<td>‘not relevant’ A</td>
</tr>
<tr>
<td>Charter 2</td>
<td>1</td>
<td>1</td>
<td>B A + B</td>
</tr>
</tbody>
</table>

Heeroma: A: supra regional
B = 100% dialectal

Van Reenen: two charters = 67%
one charter = 33%

Token count: three tokens = 75%
one token = 25%

Table 1.12: The impact of different counting techniques on the interpretation of variation.
Heeroma expected competition between more or less standardised supra-regional language and local dialect forms. Charter one, with two non-dialectal forms, is not considered to be very interesting, whereas the appearance of variant B in charter two brings Heeroma to the conclusion that variant B is the only autochthonous dialectal form of that location. For the type of data Heeroma assumes, this is a reasonable evaluation method.

The token count stresses high frequent forms. This only works when authors are not bound by any standardised form and when they are inclined to reflect mixture of forms in the spoken language in their writing. The absence of a Frisian standard spelling is discussed in § 1.3.2, the latter aspect is illustrated in § 1.3.7.9.

Therefore, even if in reality the differences between token count and charter count are not very great in the studied examples from the charters, a number of points can be made in favour of token count:

- The conclusion from graph 1.9 is that the actual mixture of forms in the charters at the level of tokens reflects general trends in variation in forms;
- Token count offers an elegant solution for the problem of how to count mixed charters;
- Token count increases the number of attestations and thus the statistical reliability of the observed trends in graphs and maps.

Despite these advantages, there is also a disadvantage:

- A few instances can be found in the material where relatively large sources with a consistent preference for one variant do in fact influence the overall picture: staette ‘estate’ in § 2.3.3.3, Hemma Odda zin’s hitalit ‘paid’ (§ 2.4.3.1) and mónendei ‘Monday’ (§ 2.3.4.2), for ‘for’ in the early 16th century in OFO III, 39, feet / feet / fotta ‘foot/feet’ (§ 2.3.4.1) from Dungeredael.

There are two more aspects of mixture that have yet to be mentioned:

- A form of mixture treated in the previous section: geographical mixture, for example, two charters from one location with homogeneous attestations, creating a geographical mixture. The geographical mixture appears to produce significant patterns (cf. § 1.3.7.7);
- Mixture of language forms of authors over different charters: for the items investigated and used in graph 1.9, between 0% and 50% of the identified authors shows variation in the (original) charters. Therefore, even when an author uses only one variant in one charter, he may very
well vary between forms over the totality of his writings and speech. This reinforces the trend in graph 1.9.

A study of variant mixtures in the charters shows that a mixture of variants is usual, consistent and produces statistically significant patterns. The intrinsic variation in the language resembles the data model II described in § 1.3.7.8. This model is supported by sociolinguistic observations and has technical advantages for the data reliability. Data model II is best served with a token count. Because there are few charters with several tokens for one item, the token count becomes quite balanced. Individual cases of distorted frequency distributions will be countered, for example, by geographical interpretations, where all attestations fall under one point, eliminating their impact on the overall image.

Section summary:

• There are sufficient reasons to assume that spelling variations in the original charters are a random sample of linguistic patterns of the time;

• Spatial interpolation techniques produce maps that are linguistically consistent and are confirmed by statistical testing;

• Variation within texts reflects linguistic variations in the speech community, making it unlikely that spelling variations are the result of erroneous or careless spelling;

• With a careful eye for temporal, geographical and individual distortions of the data, token counting is a sound and statistically beneficial way of evaluating data.
1.3.8 The language of Unia

The codex *Unia* contains the most archaic preserved Old West Frisian texts (cf. § 1.2.2). The texts in the codex are not linguistically consistent and include several linguistic ‘layers’. Sections of the text come from different eras and are not always from the same region. This implies that the linguistic variation found in the codex *Unia* can be caused either by diachronic or dialectal variation. The available text is a copy of a copy which perhaps originated from even older copies. The challenge of this analysis is to estimate at what time the different sections of the text received their linguistic shape.

For this analysis, the preliminary transcription made by Dirk Boutkan was used, counting 5799 lines. The text was split into sections (following Siebs 1895, 14-29). Each section contains independent content. Some texts that are available in the codex, are missing in the transcription, such as the *Tale of Charlemagne and Redbad*. An example of a selection from the text, with indication of sections, is presented in table 1.3.

<table>
<thead>
<tr>
<th>n = number of tokens</th>
<th>OFO I-1 (1329)</th>
<th>charters 1379 - 1405</th>
<th>charters 1405 - 1430</th>
<th>last regular recordings</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;and(e)&gt; ~ &lt;ende&gt;</td>
<td>100% (n=13)</td>
<td>42% (n=112)</td>
<td>4% (n=280)</td>
<td>1418</td>
</tr>
<tr>
<td>&lt;him/hine&gt; ~ &lt;him&gt;</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>last &lt;hine&gt; 1386; first acc. &lt;him&gt; 1418</td>
</tr>
<tr>
<td>&lt;thet/that&gt; ~ &lt;dat&gt;</td>
<td>100% (n=11)</td>
<td>84% (n=50)</td>
<td>43% (n=54)</td>
<td>± 1435</td>
</tr>
<tr>
<td>&lt;him&gt; ~ &lt;hem&gt;</td>
<td>100% (n=1)</td>
<td>75% (n=4)</td>
<td>67% (n=3)</td>
<td>after 1440: &lt; 10%</td>
</tr>
<tr>
<td>&lt;-um&gt; ~ &lt;-em/-im&gt;</td>
<td>82% (n=11)</td>
<td>20% (n=40)</td>
<td>24% (n=54)</td>
<td>± 1500</td>
</tr>
</tbody>
</table>

Table 1.13: Linguistic criteria for the dating of the *Unia* sections. The archaic form is the first form mentioned in column one, the more modern form follows the ‘~’. The percentages correspond to the archaic forms. ‘last regular recordings’ is the last year that the archaic form was attested as part of a more or less continuous temporal array. Incidental attestations in later years may occur.

To estimate the date of the different sections, the charter OFO I-1 from 1329 is a useful anchor point in the 14th century. Therefore linguistic criteria are needed
to show diachronic variations between the language of the charter OFO I-1 from 1329 and the old charters from the period 1379 to 1430. A section in Unia, complying with the language of charter OFO I-1 and differing from the late 14th century charters, can be dated from the early or middle 14th century. The following five criteria have been used. These criteria preferably appear in their archaic form in the charter from 1329 and are subject to a rapid modernisation in the charters from about 1400. The figures from table 1.13 are shown in graph 1.10.

The conjunction *ande* ‘and’ was replaced by Middle Dutch *ende* in the late-14th and early 15th century. All attestations to *and(e)* in the charters after 1405 are from the north-eastern region. After 1418, <ende> is the only form. In the sections of the Opstalsboom Statutes of 1323, Statute of the Dean of Wirdum and the ‘Dongra Bota’ on average 22% of the instances of the conjunction are archaic forms. These three sections constitute the youngest part of the codex Unia and are at least younger than 1380. They are further referred to as group C. In the rest of the codex the figure for <and(e)> lies over 90%.

The second criterion is that of a distinct form for the accusative of the masculine singular pronoun, Old Frisian *hine* (in contrast with the dative *him*). The form *hine* appears in the charter from 1329 and the last attestation in the charters is from 1386. The oldest instance of *hine* as an accusative form in an original charter is
from 1418. That makes the year 1400 an appropriate guess for the abolition of the form hine in Middle Frisian. The Statute of the Dean of Wirdum and the ‘Dongra Bota’ contain no instances of hine and do contain examples where him is used as an accusative form:

“... hit en se het hem sijn personna farra wrogia ..” (Statute of the Dean of Wirdum)
“... ende hem half dey ende nacht ...” (‘Dongra Bota’)

This criterion supports a post-1400 dating of the texts in group C. There are no relevant examples from the Opstalboom Statutes.

The Old Frisian form of the neuter article and conjunction <thet/that> is replaced by <dat> in the early 15th century. Between 1400 and 1435 the spelling with <th> gradually disappears. The archaic spelling is rare (<20%) in the three texts of group C and additionally in the ‘Leppa Wilkerran’ (50%, n = 4) and the ‘Leowerdera Bota’ (18%, n = 49). These last two mentioned texts constitute group B. The rest of the sections of Unia constitute group A.

The forms of the pronoun of the 3rd pers. sg. masc., Old Frisian him, hine are preferably written <hem> (<hene> only appears in Unia) in the charters after about 1420. The spelling with <e> is common in both the groups B and C.

The spelling of the dative plural ending of nouns, Old Frisian -um, shows a gradual decline from <um> to predominantly <em> and <im> in later periods. As long as the ending -um is used in the charters, until late in the 15th century, the spelling <um> still appears occasionally. In Unia group A, the average for the use of <um> is 83% against 61% in group B and C. However, there are substantial individual differences, for example in the B-text ‘Leowerdera Bota’, 77% (n = 29) of the dative plural forms take <um>, while this is only 38% (n = 13) in the A-text of the ‘Wilker thes nije londe’.

---

32 The form hine is attested as a pronoun for the 3rd person sg. in Modern West Frisian (WFT, volume 8, hin IV). The description of the word in the WFT dictionary suggests a direct continuation of the Old Frisian form hine. That, however, seems highly unlikely. There are few instances of <hen> in charters from 1452 and onwards, only one original (from 1515, OFO III-39, Littenseradiel). In this original example, the word is in the position where a dative singular could be expected: “[..] Doeke hat betelle ben [...]” = ‘D. has paid him’. This would appear to be a case of incidental phonetic reduction of [n] > [n], as in a few instances from early-Modern West Frisian, for example text 1626A: ‘du mochste wijte waer datst en brocht hast [...]’ ‘you should know where you put him’.

33 It has not yet been possible to assess the difference between that and het.
Group A as a whole keeps pace with the charter from 1329. This means that its language type at least represents the language of (parts of) Fryslân from the 14th century. In the following paragraphs some more details are given. The linguistic character of the texts in group B and C were thoroughly reshaped in the first decades of the 15th century. For the texts in group B, the period between 1410 and 1430 is not unlikely. The texts in group C were probably first (re)written after 1430. Note that the linguistic character of these texts can reflect aspects of older versions. In the previous section 1.3.6, it was shown that archaic features which were still optional at the time of copying / reshaping a text, were more likely to be retained than linguistic elements which were completely outdated. The high level of the dative plural ending <um> in B and C could be the result of copying from older sources with abundant presence of this archaic form. The archaic <um> could be retained in the copy because <um> was still a common option in the first half of the 15th century.

Group A is not a homogeneous group. The number of attestations per text of suitable examples for chronological discrimination is often very low. A relative order of age could however be established for six longer texts from group A:

<table>
<thead>
<tr>
<th>relative ranking</th>
<th>title</th>
<th>Gerbenzon-number</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-1</td>
<td>Older ‘Skeltenariucht’</td>
<td>855</td>
</tr>
<tr>
<td>A-2</td>
<td>The Twenty Four Land Laws</td>
<td>630</td>
</tr>
<tr>
<td>A-2</td>
<td>Synodal Law</td>
<td>857</td>
</tr>
<tr>
<td>A-2</td>
<td>The Seventeen Statutes</td>
<td>627</td>
</tr>
<tr>
<td>A-3</td>
<td>Voerdgung this riuchtis (Processus Judicii)</td>
<td>647</td>
</tr>
<tr>
<td>A-3</td>
<td>Autentica Rиеcht</td>
<td>876</td>
</tr>
</tbody>
</table>

Table 1.14: Relative dating of some of the older texts in the codex Unia (group A)

The two linguistic criteria used to define the relative ranking are:

- The use of <ande> alongside <and> ‘and’;
- The use of the full form <other> instead of <o(e)r> ‘other’.

In the text of the Older ‘Skeltenariucht’, the longer form of the conjunction is used in 11% of the tokens (n = 209). In group A-2, <ande> still appears, but only ≤ 5% of the tokens (n = 252). In group A-3, <ande> is fully absent. Detailed analysis of the forms of Old Frisian óther reveal that the transition from /ɔ:ðrf/ > /ɔːx/ was
A gradual process. At the oldest stage, the word is always written <other(-)>), both with and without case ending, so: <other, othera, othere, etc. The reduction to /ær/ first took place in inflected forms, for example othera, otherum, othere, othere, resulting in <ora>, <orum> etc. The long form was most persistent in the root form <other>. The texts from group A-1 and A-2 have only <other>, both in forms with and without a subsequent case ending. In the Seventeen Statutes, inflected forms are missing. The root appears only as <other>. In group A-3, <other> appears in 65% of the inflected forms and 86% of the root form.

The charter from 1329, OFO I-1, has only <and> and <orum> (2x) and <orne>. This implies that the texts from group A-1/2 are probably older than 1329. Geographical differences are probably not relevant for this comparison. The spelling of the word land ‘land’ as <land> and not <lond> (cf. following paragraph) suggests that the texts from group A-1/2 are from the western part of Fryslân, just as OFO I-1. The text of the Older ‘Skeltenaricht’ is (of the longer texts that provide extensive examples) by far the most archaic one. This for example, contains verbal forms without syncope, including <(une)deled> ‘(un)divided’ against <deld> in ‘Processus Judicii’, or: <kumith>, <havith> ‘(he) comes, has’ (also syncopated forms appear!), against only <comt> and <hat> in ‘Processus Judicii’. In the Older ‘Skeltenaricht’ (A-1) unsyncopated forms are significantly better represented than in the texts of group A-2. In the Older ‘Skeltenaricht’ words such as hand ‘hand’ and land ‘land’ are only spelled with <a>, just as in OFO I-1, stipulating a western provenance.

In seven texts of group A, including the two texts from group A-3 in table 1.14, the spelling <lond> appears for land ‘land’, in four texts in 100% of the tokens. The Compensation Tariff of Franekeradeel and Wonsradael (BFW) and the Compensation Tariff of Wybritteradeel (BF'B) also contain occasional instances of <hond> instead of the more common spelling <hand> ‘hand’. The spellings <lond> and <hond> appear in charters from the North-Eastern region (Leeuwarderadeel included) before 1452 only. Before 1452 the level of <o>-spelling remained rather constant. Charter OFO I-1 contains only <hand> and <land>. This suggests that the contrast between <and> and <ond> in these words were a fairly stable dialectal contrast in Fryslân for more than one century at least. The presence of the <o>-forms in parts of Unia implies that those sections (including group A-3 from table 1.14) were linguistically shaped in the north-east.

The texts in group A-3 show a high level of double consonant spelling in the words wesa ‘to be’ and seka ‘case (pl)’, for example <wassa, sacka>. This is a Middle Dutch spelling practice, that was applied since about 1400 (cf. § 1.3.3). This implies that these texts can be identified as texts from the north-east of Fryslân, from about 1400. Johnston (2001, 582) dates ‘Processus Judicii’ in the late 14th century and
'Autentica Riecht’ about 1400. The north-east appears to be a relatively conservative part of the language area in chapter two, for example in the retention of unstressed /a/ and the relatively late replacement of the feminine plural ending -a by the modern -en. Compare the following archaic linguistic features in the charters and in Unia group A-3:

- *And* ‘and’ (instead of ende) that survived in the north-east until ± 1420, about 15 years longer than in the other parts of Fryslân according to the charters;
- The spellings <hond> and <lond>, exclusively in charters from the north-east until ± 1450.

The linguistically conservative character of the north-east, where the texts in group A-3 must be located, may be responsible for their archaic character and hence their assignment to group A of the Unia texts, despite their relatively young age. The relative positioning of the linguistic character of group A-3 in the graphs 2.1 (the doubling of historical single consonants following a short vowel, as in <wessa> for wesa ‘to be’), graph 2.7 (the decline of the spelling <a> in the masculine plural ending -an) and graph 2.16 (apocope of unstressed /ə/) suggest that the texts in group A-3 postdate the year 1400 (± 1405-1410).

**Section summary:**

- The language of the majority of texts in Unia (referred to as group A) are from the 14th century and the first years of the 15th century;
- The Older ‘Skeltenaricht’ (group A-1) and some more sections (group A-2) are likely to be linguistically defined in the western part of Fryslân and are at least as old as OFO I-1, so ≤ 1329;
- ‘Processus Judicij’ and ‘Autentica Riecht’ and some more minor sections (group A-3) originate from the north-east and are from ± 1405-1410;
- The Older ‘Skeltenaricht’ is the oldest text in Unia, ± 1300.
1.4 The Power of Algorithms
§ 1.2.5 defines language as a “deterministic dynamic system, governed by self-organisation. Such systems can exhibit chaotic behaviour”. This is a reductionistic approach, different from the theory advocated by Chomsky e.a. which assumes humans have a specific inborn language ability. The following paragraphs give a working definition of reductionism, deterministic dynamic systems and self-organisation (cf. Ball 2004 for further illustrations of these concepts in non-physical contexts). This thesis does not assume to present a fully fledged model of language in a strictly defined theory. Rather this study is meant as a contribution to the establishment of such a theory. In § 5.1 and 5.2, this approach is applied to illustrate that it can work and that it provides elegant explanations for linguistic phenomena.

Reductionism
Reductionism means: Trying to understand a phenomenon by reducing it to its constituent elements. Dennett (2006, 75-78) brings up a contrast between ‘greedy’ reductionism and ‘good’ reductionism. ‘Greedy’ reductionists try to reduce everything to the lowest level immediately. ‘Good’ reductionists try to reduce features to one level lower, acknowledging that the sum can be more than the total of its component elements. Within this framework, language is the result of interaction between the following constituent elements:

---

34 Similar elements are for example mentioned by Bye (2004a, 2): “In recent years, a complementary paradigm has been developing [in theoretical linguistics], guided by the idea that cross-linguistic regularities are emergent, reflecting universal extra-grammatical constraints on domains such as articulation, perception and memory formation.”

There seems to be quite an overlap with recent positions taken by Noam Chomsky e.a., who formulate it thus: “[... we offered one potential cut through F[aculty of] L[anguage] B[road], explicitly distinguishing the sensory-motor (SM: phonetics/phonology) and conceptual-intentional (CI: semantics/pragmatics) systems from the computational components of language.” (Fitch, Hauser & Chomsky 2005, 182), where their SM-systems matches my field of ‘acoustics’, the CI-system matches ‘meaning’, while I seek the foundations for the working of their “computational components” in a more broadly defined ‘memory formation’.

This last assumption of the author of this thesis may be covered by a hypothesis formulated in the same article (idem, 206): “If it turned out that the capacity for recursion [= their main computational component of language] resulted from a phase transition in the pattern of neural connectivity that results automatically from increases in neocortex to sub-cortical tissue ratio, interacting with standard mammalian brain development, this would certainly be an interesting result.”
Meaning / semantics: The outside world and the human mental perception of it;
Articulation and acoustics: The entire field of biology and physics connected with the production and transmission of sounds;
The bio-chemical working of the mind resulting in human-specific ways of perceiving, storing and retrieving information. Linking language features directly to bio-chemical processes might be a case of ‘greedy’ reductionism; linking language features to general cognitive and psychological phenomena is the aim of a ‘good’ reductionistic approach.

‘Grammar’ is the total of structures emerging from the interaction of the aforementioned elements by self-organisation.

Deterministic Dynamic Systems
The constituent elements of many dynamic systems are subject to physical, chemical or biological laws or constraints. The field of acoustics is a good example of this. The interaction of all components, each behaving according to its own ‘laws’, produces dynamic systems with a high level of complexity. The atmosphere, including both daily weather and long-term climate change, is a familiar example of a dynamic system.

Physical laws are universal and eternal. The current state of the system is determined by history and the previously mentioned physical laws. Knowing the starting configuration and applying the ‘laws’ consistently, the outcome can be computed, assuming there is sufficient computation capacity. Making a weather forecast is in fact an effort to use the deterministic quality of the atmosphere.

There are several problems in this approach:
- The system contains so many elements that it is not feasible to determine the exact starting position of all of them;
- Even if the first problem could be tackled, the number of interactions and computations is so large that no man nor computer could handle them.

The concept of a deterministic dynamic system theoretically implies a high predictability. In the long run, it may result in a highly unpredictable system. Therefore, these systems are sometimes called chaos systems.\textsuperscript{35} Determinism is the central idea behind the models in § 5.1 and § 5.2.

\textsuperscript{35} Note that chaos systems are a well defined concept in mathematics, cf. Verhulst (2003, 24 ff.). This thesis does not use the chaos concept, nor does it go further into this here.
Self-Organisation

Patterns that result from self-organisation emerge as a consequence of repetitive and recursive actions. They are not the result of any intention and therefore lack any teleological explanation. Even when some of the revealing patterns seem to reinforce themselves by positive feedback, they do not represent any independent power. Whatever the degree of (temporal) stability of a structure, it is not the result of any teleological self-maintenance of the structure, but of the underlying repetitive actions. Any change in the environment can lead to a change or even collapse of the structure.36 The modelling of language change in Frisian in § 5.2 builds upon the idea of self-organisation.

Modelling by algorithms

The working of a dynamic system can be modelled with algorithms. An algorithm is recursive and repetitive. One or more elements interact with each other according to a predefined behaviour, and the result of this interaction is the input for the next step: \( a(x_0) \Rightarrow x_1, a(x_1) \Rightarrow x_2, \) etc.

In dynamic systems, sudden changes may take place. This is an important outcome of algorithmic processes. Their results are not necessarily linear. That is the power of a good algorithm. No actor changes plans or has on its own neither the force nor the intention to provoke a change, and still things are changing and sometimes even abruptly.

Another aspect of systems driven by an algorithm is that they may reach a (temporary) equilibrium state, where every new round returns the same result. But many systems have an engine. For instance, the sun is constantly adding new energy to the atmosphere. Many systems with some kind of energy input do not reach a final steady state, but are balancing at a non-equilibrium steady state all the time (Ball 2004, 294 ff). The growth of new generations in a mentally and physically ongoing changing society is an 'energy engine' for the dynamic system of language.

Wider perspective

Structuralist linguistics is trying to discover the rules / features / constraints / parameters or whatever they are called. They are responsible for observed linguistic patterns and structures and reflect the human language blueprint, the presumed

---

36It is precisely this approach of linguistic structures which provides the explanation for an observation made by the author in Versloot (1994, 93), a lexical-semantic study in a very strict structuralistic framework. The, at the time surprising, outcome of the study was that the behaviour of the structures is totally passive. There is no resistance whatsoever against loanwords which form a complete semantical mismatch with the existing lexical-semantic structures in the receiving languages.
inborn language ability. In a deterministic dynamic system, governed by self-organisation, the structures and patterns are secondary. They are the result of the way agents which want to communicate interact on a physical and biological substratum. To make a comparison with evolutionary biology, creationists look for the intention, the reason of an ‘intelligent designer’ to create an ant or a whale. Evolutionary biology shows that the interaction of spontaneously mutating genetic material with the environment, for instance the ongoing process of natural selection, can result in both an ant and a whale by the same forces. In a similar way the differences between languages are not the result of different rule orders / constraint rankings / parameter settings, etc. but different outcomes of the same deterministic dynamics, exhibiting chaotic properties. The Tower of Babel was not an accident. A complex, dynamic system such as human language, with its stochastic variation of individuals in production and perception, will always end up like that as the result of chaos theory.

This research has not systematically traced all the efforts that contribute to the theoretical concept outlined, but it may be worth mentioning some recent publications to create a wider perspective for this approach. Tracing the references in recent publications in this field, reveals that some of the aforementioned concepts were already formulated in the first half of the last century. This approach is gaining significantly from recent developments in computer techniques and the building of larger language corpora over the last few decades.

Zuraw (2003) and Ke (2004) provide interesting examples of algorithmic self-organisation in the emergence and dissemination of language change, resulting from the interaction of stochastic language variation and individual behaviour in social networks. Bart de Boer (2005) illustrates in his thesis The Origin of Vowel Systems (note the implicit reference in the title) how a system of contrasting vowels (phonemes) can emerge from random sounds by self-organisation. Oudeyer (2005) extends this direction of research. An interesting publication is also the thesis by Zach Solan (2006) called Unsupervised Learning of Natural Languages. Here a model is presented that creates syntactical and morphological grammars by the use of general pattern recognition algorithms. The concept of determinism is used in Pagel et al. (2007), showing how languages gradually renew their lexicon, with evolution rates controlled by the frequency of concepts.