The Role of the Individual in Language Change from the Point of View of Social Network Analysis

Alexander Bergs (University of Osnabrück, Germany)

Abstract

This paper makes four claims concerning linguistic change from the point of view of (historical) Social Network Analysis (SNA): first, that the linguistic behavior of individuals can play an important role in linguistic change, second, that the linguistic system of speakers remains flexible throughout their lifetime, third, that individual linguistic systems can and should be studied in relation to their corresponding social networks, and fourth, that historical network analyses often need to be supplemented by qualitative diachronic data due to the special nature of the network data available. These claims are illustrated on the basis of data derived from the Middle English Paston Letters (1421 to 1503).

1. Introduction

This paper is mainly concerned with two different, but related theoretical issues: the role of the individual in linguistic theory and language change, and the possibility of language change during adulthood. Both these problem complexes have received remarkably little attention in historical linguistics – both for methodological and theoretical reasons, it seems.

The role of the individual in language change and linguistic theory in general has always been problematic (cf. Johnstone 1996). While it is clear that languages are spoken by individuals, many approaches are mostly interested in the general linguistic capacity of speakers, not in individual features. 'Idiosyncracies' do not rank very highly in most linguistic theories: "We define language [...] as an instrument used by the members of the community to communicate with one another. Idiosyncratic habits are not a part of language so conceived, and idiosyncratic changes no more so" (Labov 1972, 277). Thus, most methodologies are geared towards investigating language in aggregates of speakers, or generalized, abstract and ideal speaker-hearers.

The possibility of language change in adulthood essentially goes against the grain of major linguistic theories such as generativism. It is often assumed that 'language' (i.e. 'grammar' or 'linguistic competence') is formed during first language acquisition, and that it cannot be substantially altered at later stages in life. Thus, language change (as change in the grammar) can only happen during first language acquisition. Adult speakers can only change the input to children's first language acquisition through their performance.

This paper argues that the linguistic behavior of individuals, their idiosyncracies, can play an important role in linguistic change, and that there are interesting and promising ways of investigating these factors. Furthermore, it will also argue that the linguistic system of speakers essentially remains flexible throughout their lifetime and that, therefore, language change can actually happen at any time, not only during first language acquisition. These claims will be made and illustrated on the basis of data culled from the Middle English Paston Letters, a collection of documents authored by members of the Paston family between 1421 and 1503. Contrary to what is commonly assumed about the poverty of (socio-) historical data, for most of these documents we know quite a lot about the context of their composition and their individual authors. The special nature of this database allows for longitudinal, i.e. real-time studies of the linguistic behavior of individual speakers for forty years and more. Furthermore, it also provides ample material for a more or less detailed reconstruction of the social networks of the authors and their individual biog-

raphies. Social network analysis (henceforth SNA) has become one of the mainstays in the methodological repertoire of present-day sociolinguistics (for an overview, see Schenk and Bergs 2004). It can offer a microscopic view on the 'anthill' of language use and perspectivize individual speakers in their concrete social environment(s). Thus we can gain the possibility of studying idiosyncratic language use and its repercussions in the corresponding network.

This paper is organized as follows. Section 2 will briefly present some of the basic facts on the two morphosyntactic variables which will be used in the present study: plural pronoun forms and relativizers. Section 3 will outline the basic ideas and methods of (historical) social network analysis and will give an overview on the background of the Paston Letters as a linguistic and sociohistorical database. Section 4, then, will discuss the role of the individual speaker in linguistic change, and linguistic change in adulthood from the viewpoint of social network analysis on the basis of quantitative analyses of the morphosyntactic variables presented in the first section. Section 5 draws conclusions and discusses implications for the theory of language change.

2. The morphosyntactic variables

This section presents the two nominal morphosyntactic variables for this study: plural forms of the personal pronouns and relativizers. These were chosen for a number of reasons. First, morphosyntactic variation in the authors' speech may not be affected in the same way by scribal influences as, for example, phonology/orthography (for evidence, see below). Second, morphosyntactic variation in general may not produce the same overall token frequency as phonological variation, for example. However, it can be expected that third person plural pronouns and relativizers produce a sufficient number of occurrences for quantificational analyses. Third, on the basis of handbooks and historical grammars (like the *Cambridge History of the English Language*, Volume II), it can be expected that these two variables are still very much in flux in the fifteenth century. Readers are referred to Bergs (2005) for more extensive theoretical discussions and analyses of two more variables: address pronouns and light verb constructions.

2.1 *Personal Pronouns*

In Old English, i.e. before c. 1000 C.E., third person plural pronouns were characterized by initial <h->: *hī(e) 'they'*, *hire/heora 'their'*, *him/heom* 'them', and *hī(e)* 'they'. In Middle English, we see a gradual shift from the Old English <h-> pronouns to the contemporary <th-> pronouns, which were borrowed from Proto Norse during the late Old English period. This development is summarized in a somewhat simplified form in table 1.

There is still some dispute about the language internal mechanics of this change. It is still unclear, e.g., if we see the borrowing of a paradigm or the borrowing of the nominative form only, followed by independent analogical leveling in the recipient language. Also, the geographical spread of the forms from north to south over more than four hundred years poses some interesting problems (cf. Thomason and Kaufman 1991). Without going into too much detail at this point, it can be said that phonological factors, such as the sign prominence of the dental fricative (Ritt 2001), may well have played a role in the initial phase of these changes, but that they did not factor in the later phase(s). However, some processing ease in connectionist terms may have facilitated the change, and may even account for the directionality and the spatio-temporal pattern and

	Old English	Middle English	Contemporary English
Nominativ	*Hī(e)*	*Þei*	*They*
Genitive	*Hire/heora*	*Her(e) ~ þeir*	*Their*
Dative	*Him/heom*	*Hem ~ þe(i)m*	*Them*
Accusative	*Hī(e)*	*Hem ~ þe(i)m*	*Them*

Table 1: Development of plural pronoun forms in English (simplified)

diffusion of the forms: once the interdental fricatives were introduced as Wickelphones/Wickelfeatures in the most frequent pronoun form (subject), a mixed system with initial dental and glottal fricatives was created. Frequent exposition to the dental fricative form may have strengthened this particular network node and may have ultimately pushed the system in the direction of regular, i.e. uniform forms – which require fewer active network nodes (cf. Bergs 2005, 83–103 for a full discussion). The important point for present purposes is that the paradigm apparently was not adopted in a catastrophic form throughout England, i.e. completely and at one point in time, but gradually. In the fifteenth century, which is the period of interest here, *they* was well established, while there was still considerable variation between innovative <th-> and conservative <h-> pronouns in the possessive and object forms.

2.2 *Relativizers*

The second nominal variable which will be investigated in the following are relativizers. The relative clause system which we see in contemporary standard English, i.e. a series of *wh*-relative pronouns (*which, who, whose, whom*), invariable *that*, and optional deletion in certain constructions, goes back to Middle English developments. Despite the fact that the *wh*-series was introduced into the relativization system in early Middle English, in the thirteenth and fourteenth century, *that* was more or less the only common relativizer (see Fischer et al. 2000, 91 and Morris 1895, 198). *Wh*-relativizers did not gain any noteworthy frequency until the late fourteenth century. Interestingly, they also did not appear instantaneously, but gradually trickled into the system, beginning with *whom, whose, which*. Finally, in the late fifteenth century we see the appearance of *who*. Needless to say, there is some complex interaction of different morphosyntactic factors in the English relativization system, now and then, including restrictiveness of the relative clause, syntactic function of the relativizer and of the antecedent, and ontological status of the referent. Romaine (1982) has already discussed the major syntactic factors and has pointed out that the introduction of the *wh*-series runs counter to what can be expected on the basis of Keenan and Comrie's accessibility hierarchy (1977). On the basis of Keenan and Comrie's study we would expect the subject forms to change first, since this is the most common and accessible position, followed by forms in the more complex syntactic functions, such as indirect object and possessive. For Romaine, the reversal of the hierarchy is indicative of a change from above, i.e. from formal, literate styles into more informal, oral ones. In addition, Bergs (2005) has also shown that the ontological status of the referent seems to have played a role in the early phase of this change. The *wh*-pronouns *who/whose/whom* were first used with 'Deity' antecedents, possibly in order to mirror the ontological markedness of the referent with a new and hence salient, marked linguistic form. Later on during the development, we see a bleaching and conventionalization process in which the new and marked forms gradually lose their marked status and are integrated as regular, productive forms into the linguistic system. The possible referents move from 'Deity' to 'Highly Respectable' to 'Friend' and finally 'Human' or even plain 'Animate'. One more relativizer needs to be mentioned before we move on: *the which*. This complex form can also be found in Middle and Early Modern English, arguably in "disorderly heterogeneity", i.e. more or less free variation with *which* (cf. Bergs 2005, 164–165; Raumolin-Brunberg 2000). It either goes back to a borrowing from Central French *liquel(s)/lequel(s)* (Einenkel 1916, Meier 1967) or to a reanalysis of fused Old English forms like *se e suahuelc* and *one suhulc* (*se* > *þe* and *huelc* > *which*), cf. Curme (1912). Interestingly, it has not been conclusively shown that *the which* as the obviously more complex and explicit form is significantly more common in complex morphosyntactic environments, such as distant relative clauses, as we would expect on the basis of Rohdenburg's (1998, 2000, 2003) and Hawkins's (1994, 2004) complexity principles (but also cf. Fischer 1992, 303–304 for the suggestion that *the which* is more common in non-restrictive relative clauses, especially when the antecedent is separated from the relative clause and/or lacks a demonstrative determiner). This issue seems to require further corpus-based research in the future.

After this brief discussion of some of the language internal aspects of the morphosyntactic variables to be investigated here, we will now turn to

the methodological and language external aspects, i.e. the aims and methods of (historical) social network analysis, and the Paston Letters as database.

3. Social Network Analysis and the Paston Letters

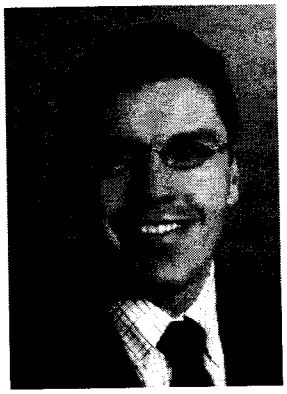
3.1 Social Network Analysis: now and then

Social Network Analysis (SNA) has been a prominent research tool in sociology and related disciplines like psychology and anthropology from the 1930s onwards. While it has been mentioned in passing in (socio-)linguistic studies from early on (e.g. Bloomfield 1933, 46f, Gumperz 1966, 34f), the first systematic implementation of SNA in (socio-)linguistics can be traced back at least to L. Milroy (1987, first published 1980), with pilot studies in the late 1970s.

One of the basic principles of (social) networks analysis is that certain entities (humans, computers, cities ...) do not exist in isolation, but that they are in some contact with each other. This contact may be stronger or weaker, and the resulting networks may be more or less dense. These very basic facts have already been described, in a refreshingly non-technical way, by Leonard Bloomfield in his 1933 classic *Language*:

> Imagine a huge chart with a dot for every speaker in the community, and imagine that every time any speaker uttered a sentence, an arrow were drawn into the chart pointing from his dot to the dot representing each one of his hearers. At the end of a given period of time, say seventy years, this chart would show us the density of communication within the community. Some speakers would turn out to have been in close communication: there would be many arrows from one to the other, and there would be many series of arrows connecting them by way of one, two, or three intermediate speakers. At the other extreme there would be widely separated speakers who had never heard each other speak and were connected only by long chains of arrows through many intermediate speakers. [...] The chart we have imagined is impossible of construction. [...] We believe that the differences in density of communication within a speech-community are not only personal and individual, but that the community is divided into various systems of sub-groups such that the persons within a sub-group speak much more to each other than to persons outside their sub-group. Viewing the system of arrows as a network, we may say that these sub-groups are separated by *lines of weakness* in this net of oral communication. The lines of weakness and, accordingly, the differences of speech within a speech community are *local* – due to mere geographic separation – and *non-local*, or as we usually say, *social*. (Bloomfield 1933, 46f, emphasis original)

Bloomfield hints at the idea that the networks that surround us as speakers may have something to do with the way we speak. But his account remains rather vague. It was only with the advent of extensive sociological and sociolinguistic research in the domain of social networks that this link was described and analyzed in greater detail. Numerous studies in sociology (for a recent overview, see Wasserman and Faust 1994) have shown that the attitude(s) and behavior of network participants depend at least partly in some complex way on the network structures they are involved in. In order to make these effects visible and measurable, network analysts have developed a



Alexander Bergs

Alexander Bergs received his Ph.D. at the Heinrich-Heine-Universität Düsseldorf in 2002 and is presently interim professor and deputy chair of English Language and Linguistics at the University of Osnabrück, Germany. He has worked on the (socio-)linguistic history of English, media and language change, morphosyntactic variation in present-day and historical Englishes, and constructional approaches to language. His publications include *Social Networks and Historical Sociolinguistics* (Mouton de Gruyter 2005) and *Modern Scots* (Lincom Europa, 2nd, revised and updated edition 2005). He is currently finishing a monograph on *Expressions of Futurity in Contemporary English* and editing two books (with Gabriele Diewald) on *Constructions and Language Change* and *Context and Construction Grammar*. His future plans include editing a multi-volume handbook on *English Historical Linguistics*, to be published with Mouton de Gruyter, and compiling a digital corpus of Late Modern English.

considerable armory of empirical research tools and measurements for social networks. The body of sociological research in this area today is extremely diverse and still growing at an amazing rate (well documented, *inter alia*, on the website of the International Network for Social Network Analysis – www.isna.org – and the journal *Social Networks* which sees its 27th annual volume in 2005 and ranks quite high on the citation index). It thus seems impossible and perhaps even unnecessary from a *linguistic* point of view to take into consideration each and every *sociological* and empirical detail of social network analysis. Leaving aside many of the details, social networks can generally be characterized by a few universal factors, which can be categorized into structural (quantity) and content (quality) components:

Structure/Quantity	Content/Quality
Density	Transactional Content
Clusters	Multiplexity
Centrality	Reciprocity

These components (in isolation and in combination) can have certain effects on the individual network members. But before we look at these effects, the components themselves should be briefly discussed. Density is defined as the number of actual ties in a network divided by the number of potential ties, which is (n(n-1)/2), *n* being the number of network participants. In a network like that in fig. 1 below we find eleven actual links and seven participants (nodes) in the 1st order zone. This gives us an overall network density *d* of 11/21=0.52. A density of 1.0 means that everybody in the network knows everybody else, 0.0 means that no one has any connections to anybody else. Thus, $d = 0.52$ is a medium value. Clusters are specific areas of high density in a given network, centrality is measurement of how central and connected a given participant is. Most network studies work on the basis of ego-focused networks with one particular participant (usually the most central one) as their anchor and starting point (marked by an asterisk in fig. 1).

On top of these structural/quantity criteria (which are usually easy to establish empirically and lead to a simple 'dots-and-lines-model' of networks), networks are also characterized by a number of content criteria. Most of these rely on the notion of transactional content, i.e. some sort of

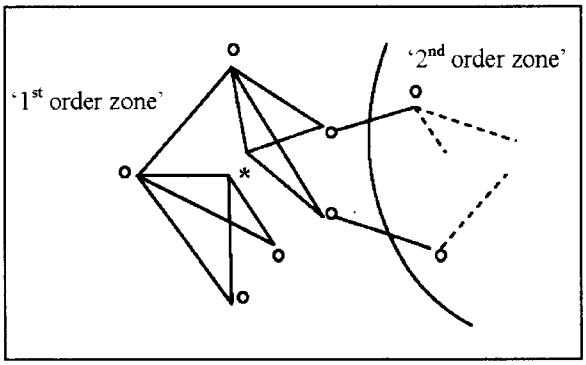


Figure 1: The 'dots-and-lines model' (adapted from Bergs 2005, 29)

transaction needs to take place between two participants in order to qualify this connection as a relevant link for the network. Transactional content can involve the transfer of material or immaterial goods, such as money, gifts, lending of books, or support, trust, and gossip. Multiplexity is a measurement for the complexity of a given link. Network participants may know each other in more than one capacity or social role, e.g. as family members, neighbors, friends, and workmates. If the link is characterized by only one such capacity it is described as uniplex; if the participants know each other in more than one role, their relationship is multiplex. Reciprocity, finally, looks at the directionality of the exchange, i.e. whether the two participants are on equal terms with regard to power and solidarity, or if there is any directionality in their exchange of goods. The final factor which needs to be introduced at this point is the very influential concept of 'tie strength', as it was developed mainly by Granovetter (1973, 1982). Tie strength is a single, combined measurement for a number of network factors: high frequency, high transactional content, a high degree of reciprocity, for example, foster strong ties between network participants. Low frequency, low transactional content, and a low degree of reciprocity foster weak ties. Vice versa: the stronger the tie the more willing network members are to exchange valuable goods, meet more often, and to reciprocate on presents, or gossip, for example. Granovetter's important contribution was his discussion of the role of weak and strong ties for the behavior and attitudes in individuals and networks. Following the line of argument developed in Mitchell, Ed. (1969) and Barnes (1969, 1972), he points out that dense,

multiplex networks with high transactional content, and therefore many strong ties, tend to generate and enforce uniform network norms inside the network, while loose, uniplex networks with low transactional content, and therefore weaker ties, are more tolerant towards non-conformity regarding network norms, i.e. "deviant" behavior. With this background, Granovetter's theory of strong and weak ties has been particularly fruitful in its application to language variation in social networks, as will be shown in the following section.

Although it is still not exactly clear whether networks and network structures can serve as predictors of behavior in the strict sense, or whether the "network concept" can merely be seen as "orientating statements" (Homans, cited in Barnes 1972, 2–3) and "a set of procedures rather than a fully-fledged theory" (L. Milroy 1987, 46), research by Granovetter and others has shown that network structures *can* have certain effects on the actual behavior and attitude(s) of network participants. Sociolinguistic research (e.g., Milroy 1987, Bortoni-Ricardo 1985, Barden and Grosskopf 1998, de Bot and Stoessel, Eds. 2002) has shown that, by extension, behavior also includes linguistic behavior. The rule-of-thumb is again that dense, multiplex networks often have norm-enforcing effects on the network members, while loosely-knit, uniplex structures are more volatile, i.e. susceptible for change and pressure from outside the network, for example the overt norms standard varieties. Thus, the term "norm" in this case must not be confused with its everyday use, where it usually refers only to "overt norms of correctness in standard varieties". The term "norm" in this approach has a broader extension. It refers to any kind of expectation or pressure to which members of a community or network are subjected. This of course comprises normative pressure to use standard or non-standard forms, depending on the community or network. This means that small, closely-knit communities, like well-established rural villages, often maintain their specific variety of language, while large, loose-knit communities as they can be found in major urban areas, are more prone to change towards standard varieties. Unfortunately, we still know very little about social networks and their effects on speakers in a historical perspective and in linguistic communities which do not have the ideological scaffolding of modern standard languages (some excellent historical studies like Imhoff 2000, Bax 2000, Lippi-Green 1994 mostly investigate networks as such without their embedding in the sociological macro-structure). But there is reason to believe that networks in principle should have the same norm-enforcing effect on speakers. This would mean that in a pre-standardized time like Middle English[1] dense, multiplex networks led to the preservation of certain regional and social varieties, while loose-knit, uniplex networks did not lead to standardization (as there was no 'modern standard' yet), but rather to greater linguistic freedom and diversity, since these individual network members were not subjected to the same normative pressures of their networks as members in dense networks. Thus, we could expect that members in loose-knit networks show greater linguistic flexibility and that they – just like in present-day studies – are the bridges across which innovations are transported from one network to another. This comes close to a reversal of the effect of networks which we see today: close-knit networks in Middle English may have led to greater uniformity (at least within the network), while loose-knit networks may have fostered greater flexibility, and hence diversity.[2]

1 Strictly speaking, there were certain linguistic 'standard' varieties in Middle English (cf. Samuels 1965; 1972, Smith 1996) and even in Old English (Gneuss 1972, Gretsch 2001, Lenker 2000). However, these clearly did not have the same status and did not play the same ideological role as those in present-day English. Note, however, that the network concept still perfectly ties in with what we know about the rise and function of the "Chancery Standard", for example.

2 One anonymous reviewer suggested that this reversal might "ruin the explanatory value of the network theory altogether". I strongly disagree. On the level of the individual, the socio-psychological mechanisms and effects (and hence the explanatory value) of networks remain the same – as a matter of fact, they may well be universal. But the macro-structures in which networks are formed and in which they operate obviously do change, and certain networks can lead to different effects, depending on their concrete (historical) context. Thus, if there is no linguistic 'standard' as we know it today, there can also be no standard norms which play a role for networks and their members. But this does not mean that networks did not play any role. On the contrary, the fact that the socio-psychological mechanisms and effects remain the same for all speakers and all times makes the concept so valuable, especially for historical research, in which macro-level concepts like "social class", for example, are not universally applicable.

One of the main problems with social network analysis, present and past, is of course data acquisition. Traditional SNA requires two different sets of data: linguistic data (preferably from individual speakers) on the one hand, and social data (from individual network participants and their ties) on the other. Although it is sometimes very surprising to see how much linguistic and social data can be unearthed even in very remote historical periods (see Tieken-Boon van Ostade, Nevalainen and Caon, Eds. 2000, Bergs 2005), it seems intuitively clear that data wealth correlates inversely with time depth. The further we go back in time, the more restricted our data becomes. Eventually, three data groups can be distinguished:

- Data that can be fully "individuated", i.e. where social and linguistic data of single speakers and their environments can be identified. One example are the members of Paston family and their letters, as they will be discussed in the following.
- Data that can only be defined on a macro-level, i.e. with regard to certain groups or locales. Individual speakers cannot be identified. One example is the language of the medieval Lollard community.
- Data that is completely "unsocial", i.e. anonymous. Speakers/scribes are unknown and their social context cannot really be established. One example here is the unknown author of the Old English epic Beowulf.
- (the fourth category – ample social, but no linguistic data – is of course also possible, but obviously does not concern us here).

So far, research using SNA in English linguistics has gone back as far as Old English. Lenker (2000) investigates the development and use of the so-called 'Winchester Vocabulary', a specific, standardized vocabulary that seems to have originated in the 'Winchester Circle', a group of monks from Winchester that can be characterized as "a closeknit, localised network cluster functioning as a mechanism of norm-enforcement" (Lenker 2000, 236). The Winchester Vocabulary probably developed in a process of cultural focusing within the monastic network. Lenker points out that social networks in this case cannot be used in a strictly quantitative sense: we do not know who the individual monks were and what characterized their individual ties. All we know is what monastic circles prototypically looked like and what kind of ties must have existed among monks and monasteries. Nevertheless, Lenker argues that the social network concept as a heuristic tool provides some very interesting and new insights for this historical sociolinguistic problem. Further English-based SNA studies mostly concentrate on Early and Late Modern English. Tieken-Boon van Ostade (2000), Bax (2000), and Fitzmaurice (2000), for example, discuss seventeenth- and eighteenth-century language and culture from a SNA perspective, and offer illuminating network studies on Sarah Fielding (Henry Fielding's sister), the Thrale family and its circles, and Joseph Addison and the Spectator Group, respectively.[3]

In the following, I will concentrate on the fifteenth century, i.e. (late) Middle English, one period of time between the Winchester Circle and Sarah Fielding. But first, the individual members of the Paston family and their networks will be described and discussed as a database for the following quantitative analyses of the two morphosyntactic variables described above.

3.2 The Pastons and their networks as a database

The Pastons are a family of the landed Norfolk gentry. Their history can be traced back at least to 1378, the year of birth of William Paston I. The period of interest for this study ends in 1504, a time when most of William's grandsons had died. Numerous historical studies have dealt with the family history. Davis's (1971) introduction to his comprehensive edition of the family documents is unsurpassed; similarly, Bennett (1995) and Richmond (1990, 1996) offer interesting and multifaceted accounts. Most of what follows is culled from their works. On the whole, we have 14 speakers in three generations, 422 documents and c. 245,000 words. The basic biographical and statistical data for each of the protagonists in this study are listed in table 2.

[3] There is a growing number of studies that deal with other periods and languages, like Imhoff (2000) on medieval Navarro-Aragonese and Lippi-Green (1994) on Early High German. For reasons of space, these cannot be discussed here; the reader is referred to Bergs (2005) for a somewhat broader coverage.

Biodata		No of texts	No of words (approx.)
Generation I			
William I	1378–1444	12	8,132
Agnes	?1400–1479	22	7,746
Generation II			
John I	1421–1466	44	33,092
Margaret	?1420–1484	107	67,305
Edmond I	1425–1449	2	569
Elizabeth	?1429–1488	3	3,958
William II	1436–1496	33	15,311
Clement	1442-?1479	7	3,254
Generation III			
John II	1442–1479	86	48,603
John III	1444–1504	78	43,490
Margery	?1455–1495	6	2,634
Edmond II	?1443-?1504	8	3,813
Walter	?1456–1479	4	1,305
William III	?1459-after 1504	9	4,508
TOTAL		422	243,847

Table 2: Biodata, number of texts, and number of words for the Paston family members

As can be seen in table 2, this is not a temporally balanced corpus; some speakers were much more productive than others. Nevertheless, even with this uneven distribution, a number of different perspectives and corresponding hypotheses may be tested. First, one may wish to consider changes in the database as a whole, i.e. between 1378 and 1504, without distinguishing between the different family members. This seems to be particularly attractive for a description of the macro developments, since this corpus comprises mostly one text type. The collection includes a few testaments, deeds, inventories, memoranda, and indentures, most of the documents (349 out of 422, i.e. c. 83%) are letters. This means, in other words, that we are *not* comparing the language of fourteenth century fabliaux with that of fifteenth century deeds of trust. Second, one may want to consider generational shifts. Here we have three generations of Pastons, and while the database may not allow for statistically perfect comparisons (one of the major differences between present-day and historical sociolinguistics!), the number of words still seems sufficiently large, at least for the second and third generation, so that there is a good chance of finding some generational patterns of language change, should there be any. Third, this corpus allows us to look at the language use of individual speakers. The two brothers, John II and John III, for example, have left us approximately the same number of words, and tracing differences in their individual language use may be quite illuminating, as Davis (1983) has already shown. We will return to this case study later on. Fourth and finally, this corpus also presents us with the chance to investigate some traditional, macro-sociological factors such as gender. The Paston Letters are one of the first data sources for gendered language use in English. Unfortunately, none of the female family members could read or write, and therefore had to dictate all of their writings. It could be objected then that the language of the letters mostly represents the language of the male scribes, not that of the female authors. This could be true, of course, but Bergs (2005, 79–80, 127–128) shows that we need to distinguish between the different language levels. Orthographic/phonological variables are perhaps affected by scribal practice, but morphosyntactic and lexical variables seem to remain mostly unaf-