

Quantifying linguistic changes

Experiments in Norwegian language history

Helge Sandøy
University of Bergen

There are reasons for assuming that different types of communities provide different social conditions for linguistic changes with consequences at least for the speed and for the type of grammatical changes. In order to explore this question, we need both a typology of communities and a model for measuring the extent of linguistic change. In this article I suggest and discuss a way of measuring by which it should be possible to systematically compare the degree of linguistic change. The tentative calculations will primarily be illustrated by data from dialects in Norway in the 20th century.

1. Aims

When we work on language history and on sociolinguistics for quite a long time, we come to realise very clearly that we cannot avoid quantifying linguistic changes and language differences. If we succeed in establishing some kind of scale that makes us less subjective when measuring, we will be able to:

1. improve our analyses and discussions of linguistic changes and the societal factors influencing the changes, or, more ambitiously, evaluate the strength of the various driving forces behind the changes; and
2. show how our concepts and awareness of linguistic changes and of the driving forces are an effect of ideology.

The latter point is of interest for the sociology of language, the former for sociolinguistics.

There are already claims or hypotheses about how societal factors influence linguistic changes. Peter Trudgill (1997), Henning Andersen (1988) and many others have focused on stability v. instability, endocentric v. exocentric, urban v. rural, denseness v. looseness of social networks, degree of contact v. isolation, etc. There are claims about the extension or degree of changes, e.g. that levelling tendencies

and changes increase in unstable societies. We therefore need a better grasp of both how to describe societies using independent criteria and how to quantify linguistic changes in order to check such claims.

My present intention is to present an explorative experiment, mainly from Norway, by using accessible sociolinguistic data from the last century. Before the experiment, I was curious as to whether quantification would yield any meaningful results at all – i.e. whether they would show some correspondences with what we already know and understand of language change, or whether the results would be totally random from one study to another. In the latter case, the experiment would fail; in the former, it could be fruitful and should be followed up.

2. Differences and changes

We should first of all emphasise that there is a difference between measuring a synchronic language distance (cf. the Levenshtein distances in Heeringa 2004) and measuring distances in language history (= changes). The point of departure is essential when performing such measurements. Thus, when we move between contemporary systems, the complexities we have to negotiate can turn out to be more difficult from system I to system II than when we move from system II to system I. When studying historical changes we can get around some of the problems because we have a diachronic point of departure, and we always move from I to II.

When we set out to measure changes, our two main problems are:

- a. to define what is the change, and
- b. to find a method/standard of weighing changes

3. Changes in grammar

If we want to count the number of grammatical changes, the first essential problem to be solved is how to define change, because what one grammatical model might define as two changes, another might define as only one. The simplest and wisest method is to stick to the surface level, which will normally invite us to use traditional grammar for the description. As a first experiment, I picked at random a passage from Snorri Sturluson's *Saga of Óláfr Haraldsson*, which is written in Old Icelandic, and transformed it into what is considered to be the Norwegian version of Old Norse from about 1200. Some of you certainly know that there were some tiny differences between Old Norwegian and Old Icelandic at that time. Afterwards

I transformed the same text into the stage where the Norwegian language was at about 1500, and here I followed standard books on language history faithfully. Finally, the text was transformed once more – now to the stage from 2000 – by taking into account the dominating features in Norwegian dialects today.

There is no space here to go into the details of the grammatical problems, but I will present the results:

For the period 1200 – 1500:	12 phonological changes 6 morphological changes 2 syntactic changes
Total	20 structural changes in 300 years.
For the period 1500 – 2000:	5 phonological changes 2 morphological changes 4 syntactic changes
Total	11 structural changes in 500 years.

This experiment underlines the traditional view that there were very many changes in the language during the late Middle Ages. Since then, we have had a more stable period. This makes me sceptical about the popular claim that our language is currently undergoing rapid change. In any case, such results invite us to discuss how this can be, and how these very big differences in the speed of change can be understood from what we know about changes in Norwegian society over these periods. My conclusion is that these results from our calculations represent a real challenge to language historians, although we have some ideas on how to explain them. However, such an explanation lies beyond this methodological paper.

What we do theoretically in this case is to count the changes that have to be carried out in our grammatical capacity in order for it to generate or produce the right language product at a new stage. This can be compared with the way of reasoning in traditional generative grammar, and in our quantification we have not taken into account whether a grammatical rule has been applied once or several times when producing a text. What is implied is that we calculate the “cost” of carrying through a grammatical change – or changing the software, to employ a metaphor. Whether we use a (changed) feature often or rarely fails to represent any difference in “cost” in our method so far.

4. Changes in frequency

In sociolinguistic studies we are accustomed to counting frequencies of the variants in the texts produced in interviews. Very often traditional dialect forms are

compared with new forms of the variables, and the results of a study are demonstrated in percentages of old and new forms, at the level of informants, groups or the whole community. Normally, each study has a whole set of variables and, in many cases, the tendency to use either traditional or new forms is quantified by calculating the average of the whole set of variables. Thus we are able to quantify the overall tendency to use the traditional dialect or the new dialect. And we might be tempted to say that we are able to give a precise measure of how much has changed from one point in time to another.

This method of quantification in sociolinguistic studies has proved to be fruitful within each study when comparing groups, styles, etc. However, there are some important objections to using the method at the overall level, and it cannot be used when we want to compare one community with another.

First of all, in such studies each variable has the same impact on the average. Insofar as the set of variables is the same for all groups and for all individuals in the study, this does not disturb the comparison too much. But when our studies from different dialect areas refer to different sets of variables, it goes totally wrong. A change in a variable instantiated once a week ought not to have the same impact on the overall percentage as another variable found in almost every sentence. An example of this is a study in which one variable was the past participle of a class of irregular verbs, which appears only once per 100,000 words in a normal text, and another variable was the phonological feature of retroflexion, which is found in every fortieth word in a normal text.

We can imagine that a study of frequency is a relevant approach when we want to understand changes, and I have therefore carried out an experiment in which the variables of all projects taken into account from the comparative perspective are assigned a coefficient representing their frequency in an average text. In this way, the percentage of retroflexion, for example, will have an impact on the overall percentage of changes that is 2,500 times greater than the impact of the rare irregular verb class mentioned above.

The point of departure for my so-called 'average text' is a dictionary of frequency in Norwegian (Vestbøstad 1989). In fact, this dictionary is about *Nynorsk*, the written language variant established on the basis of our dialects. Ideally, of course, the coefficients should have been based on the spoken language, but so far we have not had access to large databases on spoken Norwegian. I do not think this should cause great mistakes or distorted results for my explorative experiment, since I do not include lexical variables or variables concerning text composition. However, this will be improved as soon we have better conditions for a frequency dictionary of the spoken language. An advantage of such a dictionary is that it gives information and frequency of all variants and inflected forms of a lexeme.

On the basis of all these collected data it is possible to establish the frequencies of the many morphological and phonological categories of interest. These can be, for instance, an inflectional category of a specific declensional class of nouns or a phoneme in a specific phonotactic context. The result is a table of coefficients for all variables used in sociolinguistic studies of spoken Norwegian. In Table 1 and 2, I provide some examples. For practical reasons the coefficients used below refer to frequencies in an average text of 1,000 words.

Table 1. Coefficients of some morphological variables

<i>Morphological variables</i>	<i>Coefficients</i>
Mask. i-class PL.	0.414
Mask. a-class PL.	7.3947
Weak FEM. PL. (-ur)	1.416
Strong FEM. PL. (-ir)	2.438788
i-language (str. FEM. SG.DEF. + NEUTR. PL.DEF.):	4.092526
Bisyllabic infinitives	3.9087314
Inf. with vowel balance	11.067021
Pres. tense <i>hev</i> ('has')	11.119275
Irreg. verbs pres. tense	11.922791
Irreg. verbs 2nd CL. past tense (<i>au</i> vs. <i>øy</i>)	0.0450585

Table 2. Coefficients of some phonological variables

<i>Phonological variables</i>	<i>Coefficients</i>
Retroflexion	22.89
Palatalisation velars	16.46
Diphthong vs. monophthong (<i>ei</i> , <i>øy</i> vs. <i>e</i> , <i>ø</i>)	38.562

By using these coefficients I have recalculated the results from the various studies that can be used in a historical perspective and thereby have been able to compute an overall frequency for linguistic changes in each given dialect over a specific period. This can be illustrated in some detail by a study of a dialect from Vinje in the county of Telemark (Tvitekkja 1998), as in Table 3.

Table 3. Changes in the dialect of Vinje, Telemark

<i>Variable</i>	<i>Percentage of change in the Vinje study</i>	<i>Coefficient</i>	<i>Frequency of change</i>
1. retroflexion	89	22.89	20.3721
2. str. FEM. Umlaut-å	25	0.607	0.15175
3. nd : nn in final pos.	38	1.0304805	0.3915825
4. palat. velars	62	16.46	10.2052
<i>Total phonological variables 31.1206325</i>			
5. MASC. i-CL. PL.	75	0.414	0.3105
6. MASC a-CL. PL.	1	7.3947	0.073947
7. weak FEM. PL. (- <i>ur</i>)	25	1.416	0.354
8. str. FEM (- <i>ir</i>)	30	2.438788	0.7316364
9. i-language	42	4.092526	1.7188609
10. bisyllabic INF.	57	3.9087314	2.2279768
11. stem vowel INF. 2nd CL. str. verbs	82	0.0207774	0.0170374
12. INF. with vowel balance	10	11.067021	1.1067021
13. pres. tense hev 'has'	10	11.119275	1.1119275
14. pres. str. verbs	7	11.922791	0.8345953
15. past 2. CL. str. verbs (au vs. øy)	4	0.0450585	0.0018023
16. past 5th, 6th and 7th CL. str. verbs	24	1.2005394	0.2881294
17. stem vowel past pts. 2nd CL. str. verbs	15	0.0113679	0.0017051
<i>Total morphological variables 8.7788202</i>			

This study uses 17 variables to describe the linguistic changes of the dialect over the first eight decades of the 20th century, and it shows that the average young language user will apply a new linguistic variant almost 40 times when he or she produces a text of 1,000 words. The period of eight decades refers to the fact that the researcher, Tvitekkja, examined the dialect from about 1900 as the point of reference for the notion of 'traditional dialect' and that her youngest informants were born in 1980.

My intention here is for this figure of about 40 changes to be comparable with figures worked out in a corresponding way from other studies.

However, one aspect is troublesome. Two phonological variables make up more than half of the overall frequency, i.e. retroflexion with a coefficient of almost 23 and palatalisation of velars with over 10. In general, phonological variables tend to be more frequent than morphological ones, which raises the question whether results from the different grammatical levels of description can be compared directly or not. Can a frequency of 10 in phonology, for example, be considered

as being of the same proportion as 10 in morphology? Since we have not, so far, developed a theory concerning this, it is hard to find relevant arguments for a discussion, and I prefer, for the time being, to demonstrate the figures for each of the levels. In our case from Telemark they will be as in Table 4.

Table 4. Changes on two levels, Vinje

<i>VINJE</i>	<i>Frequency of change 1900 - 1980</i>
Phonology	31.1
Morphology	8.8
Total	39.9

In this experiment there is a more practical reason for being so precise; most studies concentrate on morphological variables, whereas phonological ones seem to be dealt with more accidentally. Furthermore, it may be that the morphology of a dialect is easier for the researcher to survey and to be aware of so that we can feel more confident that the morphological results in a study are more complete – given that our aim is to deduce an overall figure for changes. In the following I will therefore concentrate on morphology.

What, then, is being measured by frequency figures? We should assume that in a community there are restrictions on how much can be changed – not only because language is a means of communication, but also because of social restrictions, since a community has a fundamental claim on loyalty to its norms and there is a limit to what can be tolerated. We might imagine that the social restrictions on deviation from the norms apply to both *what* deviates and by *how much* it deviates. These social restrictions cannot be observed directly, but we may assume that the limits of tolerance are influenced by the type of society and by the degree of societal change. Furthermore societal factors can be observed independently. We therefore measure linguistic entities, and thereby linguistic changes, and compare them with these societal factors in order to look for correspondences.

5. Typology of society

Our next step is to find sociolinguistic studies from societies that can be characterised as different in terms of some given societal and social features. Table 5 shows the frequencies of morphological changes in seven societies (six Norwegian communities and one Swedish), for which I have recalculated the figures following the method already described in this paper. All of them take the period around 1900 as their point of departure when defining and registering changes.



Map of Norway and Sweden

First of all, we should notice that the frequency of change for Vinje is relatively low compared to what we see in other communities. The author of that study, however, considers the changes to be very dramatic – for the obvious reason that she had no exact standard with which to make a comparison.

We have, of course, made some assumptions about what a table like this will show us, and the hypotheses nearest at hand are that size and mobility will have an impact on the speed of linguistic change.

Table 5. Morphological changes in different types of society

<i>Type of society</i>	<i>Morphological changes in the first 6–8 decades of the 20th century</i>
Parish – stable: Vinje	8.8
City – stable community: Eskilstuna	13.1
Parish – incomers and out-movers: Eikesdalen	11.1
Parish – incomers: Nore	25.6
Industrial town – regional incomers: Sunndalsøra	37.9
Industrial town – regional incomers: Odde	5.4
Industrial town – national incomers: Tyssedal	39.1

The communities are quite different; *Eskilstuna* is a rather stable city (Nordberg 1972, Sundgren 2002), whereas *Vinje* is a stable rural parish. *Eikesdalen* is a very small unstable rural parish (Austigard 1995), with both incomers and emigrants, and the community has diminished from 250 people to 84 over the last 50 years. *Nore* is, on the other hand, an expanding rural parish characterised by a high proportion of incomers (Papazian 1999). The next three communities are new industrial towns and therefore linguistic melting-pots: *Sunndalsøra* (Jenstad 1983) and *Odda* have drawn incomers mainly from their own region, whereas *Tysseidal* was dominated by incomers from the whole country (Sandve 1976, Sandøy 1985). Table 5 sums up these findings by demonstrating the characteristics for the societies and the figures for changes in morphology for almost the same time span, i.e. the six to eight decades after 1900.

We see from these figures that unstable communities have a higher rate of change than stable ones, and new towns or industrial centres normally have more changes than communities with a social tradition and more historical identity. We can therefore conclude that this way of quantifying language change appears to be fruitful, and that the quantified results are not random. We see, furthermore, that we are facing new and interesting challenges. The community of *Odda* should be analysed more thoroughly because it returns a remarkably low figure despite its characterisation as a new industrial town. My hypothesis is that this is the case because *Odda* was an established community with a focussed local identity even before the great industrial expansion. From the starting point of industrialization in 1906 until 1920, *Odda* increased its population from 600 to 4,339, whereas the neighbouring *Tysseidal* increased from 39 to 1,124 people in the same period. (Sandøy 2004.)

6. Conclusion

I hope to have illustrated how these two quantification methods can in fact give us a tool for comparing changes in different societies and even at different periods. I have also checked the two ways of quantifying on the same set of data, i.e. both counting grammatical changes and counting frequency. On the whole they seem to give the same ranking order, but the relative proportions are different. The tentative results of this experiment demonstrate interesting correspondences between our quantification results and societal factors that can help us to understand language change, and we therefore venture to conclude that it will be worthwhile to elaborate further on the method in order to improve it. Given that this appears to be a fruitful approach, perhaps an even more important challenge is to develop further a more precise model for societal factors, so that we can establish reliable criteria for a comparison of different communities. This would hopefully enable

us to develop a method that could improve our insight into the societal conditions for tolerance of linguistic change.

References

- Andersen, Henning. 1988. "Center and Periphery: Adoption, diffusion, and spread". *Historical Dialectology. Regional and Social* ed. by J. Fisiak, 39–83. Berlin: Mouton de Gruyter.
- Austigard, Liv Dagrún Gjelsvik. 1995. *Eikesdalen – eit marginalt språksamfunn*. M.A. thesis. Trondheim: Nordisk institutt.
- Heeringa, Wilbert. 2004. *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. Groningen: University of Groningen.
- Jenstad, Tor Erik. 1983. *Eit nytt sentrumsmål veks fra. Drag frå talemålet til ungdom på sunndalsora*. Trondheim: Tapir.
- Nordberg, Bengt. 1972. "Morfologiska variationsmönster i ett centralsvenskt stadsspråk". *Språk och samhälls 1* ed. by B. Nordberg, 14–44. Lund: LiberLäromedel.
- Papazian, Eric. 1999. "Dialektød i Numedal? Om språkutviklinga i Nore og Uvdal". *Austlandsmål i endring. Dialektar, nynorsk, og språkhendingar på indre Austlandet* ed. by Turid Kleiva, 49–61. Oslo: Det Norske Samlaget.
- Sandve, Bjørn Harald. 1976. *Om talemålet i industristadene Odda og Tyssedal. Generasjons-skilnad og tilnærming mellom dei to målføra*. M.A. thesis. Bergen: Nordisk institutt.
- Sandøy, Helge. 1985. *Norsk dialektkunnskap*. Oslo: Novus.
- . 2004. "Types of Society and Language Change in the Nordic Countries". *Language Variation in Europe. Papers from the Second International Conference on Language Variation in Europe, ICLaVE 2, Uppsala University, Sweden, June 12–14, 2003* ed. by Britt-Louise Gunnarsson et al., 53–76. Uppsala: Uppsala universitet.
- Sundgren, Eva. 2002. *Återbesök i Eskilstuna. En undersökning av morfologisk variation och förändring i nutida talspråk*. Uppsala: Institutionen för nordiska språk.
- Trudgill, Peter. 1997. "Typology and Sociolinguistics: Linguistic structure, social structure and explanatory comparative dialectology". *Folia Linguistica. Acta Societatis Linguisticae Europaea* 31:3–4.349–360.
- Tvitekkja, Sigfrid. 1998. *Talemålsendring og sosial variasjon*. M.A. thesis. Oslo: Nordisk, Universitetet i Oslo.
- Vestbøstad, Per. 1989. *Nynorsk frekvensordbok*. Bergen: Alma Mater

HISTORICAL LINGUISTICS 2007

SELECTED PAPERS FROM
THE 18TH INTERNATIONAL CONFERENCE
ON HISTORICAL LINGUISTICS,
MONTREAL, 6 - 11 AUGUST 2007

Edited by

MONIQUE DUFRESNE

Queen's University

FERNANDE DUPUIS

Université du Québec à Montréal

ETLEVA VOCAJ

Université du Québec à Montréal

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

2009