Language shift, ethnolinguistic vitality and historical sociolinguistics: testing the models

Remco Knooihuizen

Early Modern Europe does not seem to have been a friendly place for minority languages. A quick glance at just four countries in North-Western Europe – Germany, France, the United Kingdom and the Netherlands – allows us to identify at least thirteen cases in which a language shift took place away from a minority, albeit at varying rates and with varying degrees of completion at the end of the Early Modern period (c. 1500–1800).

What were the social and sociolinguistic reasons for these language shifts? Are there common factors across these cases or did all these populations shift for different reasons? If there are common factors, are these specific to the Early Modern period, or are they generally comparable to the causal factors identified in language shift case studies in the 20th and 21st centuries?

In order to answer these questions, it is necessary to do a comparative study of several of these Early Modern language shifts, using a standard research method. A number of models is available (cf. below), but these are all based on case studies and theory from the second half of the 20th century. It needs to be established whether these modern models can be used for historical research. I will investigate the applicability to historical research of the 'typology of language endangerment' model by Edwards (1992) along two lines: firstly, does the model fit the historical data with regard to the causal factors it allows us to pinpoint; and secondly, does it fit with regard to the amount and quality of data from this period that is available to us (the so-called 'bad data problem', cf. below).

Language shift

Language shift has been a research topic within linguistics for approximately half a century, but a clear and universal definition seems to be lacking. However, there are three issues that frequently come up in discussions of language shift (Weinreich 1967; Mackey 1980: 35; Giacalone Ramat 1983: 495; Sasse 1992: 13; cf. also Clyne 2003: 20–21); these can function as a working definition for the purpose of this paper.

The first is 'changing patterns of language use'. This is based on the idea that there is a pattern of which language variety people use in what situations ('domains', cf. Fishman 1972b: 247–248). In a language shift, there is a change in the allocation of varieties to domains. In traditional accounts of language shift by

e.g. Joshua Fishman (1972a: 79–88), it is seen as a very neat domain-by-domain shift, so that the abandoned language is gradually replaced by the target language. In reality, there are issues like code-switching and code-mixing that mean that the shift is not as neat as described by Fishman, but it is still possible to use a catch-all phrase like 'changing patterns of language use'.

The second issue is the idea that language shift happens in a speech community. Language shift can of course be studied psycholinguistically at the level of the individual speaker, but for sociolinguistic studies it only becomes interesting once the shift happens community-wide, cf. the distinction between speaker innovation and language change by Labov (1972: 277).

The third and final issue is that language shift happens in a situation of language contact. If a community is to shift language, they needs to have a language available to them to shift to; in other words, there needs to be language contact.

The way language shift is to be understood in the context of this paper, then, is as gradually changing patterns of language use in a speech community in a language contact situation.

Language shift research

In broad terms, three types of research into language shift can be distinguished. The first type of research, of which Joshua Fishman is the main exponent, was very descriptive and was mainly concerned with the actual changes in the patterns of language use. Charting the changes in the answers to Fishman's famous question 'who speaks what language to whom, and when' over a period of (real or apparent) time allows us to see a language shift spread through the domains.

The explanatory framework in this type of research, however, is rather slim, and seems to be based on the assumption that the language used in high-prestige domains is naturally so attractive that will replace the language used in lesser-prestige domains, cf. the emphasis on the use of minority languages in high-prestige domains in Fishman's 'reversing language shift' work (1991: 87ff.).

Research was then done into the correlation of language choice and a wide range of social factors, such as the number and geographical spread of the language's speakers, whether languages were written or not, whether the dominant group was in any way suppressing the minority group, etc. This way of thinking was pioneered by Heinz Kloss (1966) and became dominant from the 1970s.

Einar Haugen's (1972) model of language ecology was the first model devised to look at language in its interaction with its speakers and their community. Giles *et al.* (1977) looked at a (minority) language's ethnolinguistic vitality on the axes of status, demography and institutional support. Giles *et al.*'s model is based on objective data, and they recognized the need for a subjective counterpart (1977: 318), which was worked out by Bourhis *et al.* (1981). A number of alternative models is available, of which I will mention the 'typology of language

endangerment' by Edwards (1992, revised by Grenoble & Whaley 1998, cf. below); see Edwards (1992) for a discussion of these alternative models.

From the late 1970s, commencing with Gal (1979), language shift research has also either explicitly or implicitly drawn on social network theory (see Govindasamy & Nambiar 2003: 26–28 for a concise but incomplete overview). Although some studies posit the existence of two separate social networks for immigrant individuals, one in their home country and one in their new country, which has not allowed for clear correlations to be identified (cf. Hulsen *et al.* 2002, Stoessel 2002), most studies use a one-network analysis. In these cases, the general tendencies from other social network studies apply (Milroy 1987: 170–171; Govindasamy & Nambiar 2003: 29): dense, multiplex networks of L1 contacts facilitate language maintenance, whereas sparse, pauciplex networks correlate with language change in the direction of the standard, in this case language shift in the direction of monoglot L2 use. (See Milroy 1987: 177–216 for a discussion of the role of social networks in language maintenance and change.)

These three methods look at language shift from different angles, so they can be used complementarily rather than substitutionally.

Historical sociolinguistics and the 'bad data problem'

When doing historical sociolinguistic research, we are faced with the so-called 'bad data problem' (Labov 1994: 10–11): we do not have access to the same amount and quality of data as we would have if we would do a case study today, for the simple reason that the people whose linguistic and societal behaviour we are studying are dead. Getting linguistic data is not merely a matter of getting a tape recorder and recording how people speak; we will have to make do with what is left of the written record, with all the implications this has for the representativeness of the data. Similarly, it is impossible to ask subjects about their societal behaviour, and we have to rely on information that is already written down, in correspondence and diaries, or more generally, in records and history books. In other words, we have to make the most of bad data (Labov 1994: 11, see also Nevalainen 1999).

This has been done quite successfully in variationist studies of Early Modern English (Nevalainen 1999, Nevalainen & Raumolin-Brunberg 2003, and others) and Early Modern German (Lippi-Green 1994, and others), which were based on a relatively large text corpus by known authors.

There are, however, some important differences between these variationist studies and the research into Early Modern European language shift currently intended. I do not expect the difference between a study into the spread of change in single linguistic tokens and one into the replacement of entire varieties to have many implications, apart from the fact that some minority languages were not used in writing. More importantly though, subordinate minority language

populations tended to be less literate than majority language groups (Houston 2003: 301–302). The amount of data would therefore be necessarily smaller than would allow a viable variationist study. Also, it means that for societal data about the minority language group, we are dependent on what the majority wrote down about them. Issues of accuracy and bias come into play here. It is therefore useful to address the implications of the bad data problem for the three currents in language shift research.

The first type of research, that focuses on diglossia and domains, is affected by the bad data problem in the sense that a sufficient amount of text on a wide range of topics should survive in order to be able to assess the allocation of languages to domains; this is often not the case. Additionally, surviving texts will only be able to show the languages used in writing; these need not necessarily be the same as those used in speech. It is however possible to find information about language use in historical comments. Although this does not give a full picture of language use, it can be used as a basis for research.

Similarly, research focusing on ethnolinguistic vitality will not have the full spectrum of data available to it, but surviving sources are generally enough to get an overview of the situation that is detailed enough to base conclusions upon. It goes without saying that this applies only to objective ethnolinguistic vitality in the terms of Giles *et al.*, not to subjective ethnolinguistic vitality – subjects are not available to fill in questionnaires. How the bad data problem affects research into ethnolinguistic vitality will be discussed in more detail in the second part of this paper.

The inavailability of live subjects also fosters major difficulties for historical social network research. Although rudimentary networks have be identified in some of the historical variationist work (cf. esp. Lippi-Green 1994: 25–26), but this was only possible because scholars could draw on a large corpus of correspondence, in the case of Lippi-Green's study by high-profile authors. For low-profile subordinate minority individuals who did not leave much of a written footprint, identifying their social networks will in my expectation prove almost impossible. I therefore expect the most viable results from an ethnolinguistic vitality approach.

Typology of language endangerment
The model I propose to use in this research is the 'typology of language endangerment' by John Edwards (1992). This model was specifically designed to research minority language situations, and in this respect it is preferable over more generally applicable models like Einar Haugen's language ecology. Although there are several other models that are also designed for minority language

situations (cf. above), Edwards' model seems to be the most extensive, i.e. it looks at the broadest range of possibly influencing factors.

Edwards set out eleven areas of factors that can influence a group's ethnolinguistic vitality at three different levels. The eleven areas are demography, sociology, linguistics, psychology, history, politics (including law and government), geography, education, religion, economics, and the media (labelled 'technology' in Grenoble & Whaley 1998). The three levels are 'speaker', which does not mean the individual speaker but the minority speech community; then 'language', which signifies the relationship betwen the minority language as an object and the area of influence in question; and finally 'setting', which the minority community within the majority community.

Crossing the eleven areas with the three levels gives thirty-three possible ways of influence. Edwards (1992: 50) provides a list of thirty-three sample research questions;[1] answering these questions should enable conclusions about a group's ethnolinguistic vitality and which factors render a group's language less viable.

The model was originally meant for a synchronic description, but I do not foresee any problems using it for a diachronic description. On the contrary, I think comparing diachronic answers to these questions to a diachronic domain analysis will allow us to see which linguistic changes correlate with social changes; it can then be investigated whether these correlations are causations.

∽

I will now briefly look at a case study in order to test the applicability of Edwards' model to historical situations. After a short introduction to the case study, I will investigate how the bad data problem influences research with this model, and whether the model allows us to find the reasons for the language shift in this case.

The test case: Shetland Norn
Until the 18th century, a Scandinavian language called Norn was spoken in the island groups of Orkney and Shetland, which were pawned to King James III of Scotland by the King Christian I of Denmark-Norway in 1468/1469 and never redeemed. By what we can tell from features of Shetland Scots (Donaldson 1983: 15) and also some comments from ministers and travellers from the period (cf. Knooihuizen 2005: 112–127), the language shift from Norn to Scots started no later than c. 1600, and there are no signs that Norn was in use as a community language after c. 1775.

---

[1] Edwards (1992: 50) warns about taking these questions too literally; they should really be seen as sample questions. Researchers should therefore be free to address other issues within the suitable matrix cell.

The case of Norn is one of extremely bad data. There is no reliable linguistic data for the period of shift. Scots was used in writing consistently from c. 1520 onwards (cf. Ballantyne & Smith 1994, 1999; many of the 16th- and 17th century documents in a variety of Scandinavian are of unclear provenance and were most likely not written in the islands themselves, cf. Smith 1990: 29), and a small Norn corpus dating from 1774 (Low 1879: 108–112) is almost certainly the language of a rememberer (Knooihuizen 2006: 106–107; cf. Dorian 1982: 32). There is, however, some anecdotal information about the language situation in the comments from ministers and travellers mentioned above.

The situation with regard to societal data is slightly better, although still by no means comparable to live ethnographic fieldwork. Local government and church officials have left a reasonable paper trace which can at least answer some of our questions, and detailed studies have been made of certain aspects of Shetland society, e.g. settlement and rental rights (Smith 2000) and education (Graham 1998). The nature of this surviving information, however, does have implications for how research is affected by the bad data problem; I will return to this in the next section.
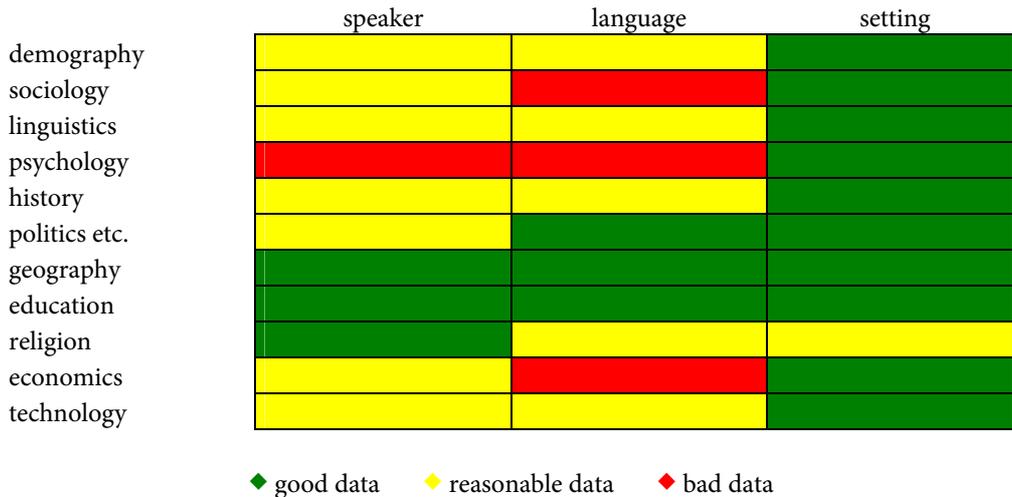
An additional caveat that needs to be made about the Shetland data is that the currently prevailing attitude in the islands is very pro-Scandinavian and at times anti-Scots (Smith 1990: 25; Melchers 1991: 463–464; Barnes 1998: 1), which could result in distorted data from a prejudiced reading of historical sources.

Earlier research into the Norn–Scots language shift (cf. Barnes 1984, 1998; Knooihuizen 2005, 2006; and others), which did not make use of Edwards' typology of language endangerment, has identified a number of possible contributing factors to the language shift.[2]

The first of these is the use of Scots in high-prestige domains as the church, administration and law, both as a written language and, more importantly in a probably largely illiterate society, also as a spoken language.

There was a growing exposure to the Scots language not just because of its use in high-prestige domains, but also through trade and an extensive immigration of Scots speakers to Shetland from the 16th century, accounting for approximately a third of the Shetland population by 1600 (Donaldson 1983: 13; for the implications of such a substantial immigration by a dominant-language group, see Thomason & Kaufman 1991: 122). Simultaneously, contacts with mainland Scandinavia, which would have functioned as a support base for Norn, declined. In other words, there was a significant change in language contact patterns in the islands.

---

[2] For a full discussion, see the references given.

	speaker	language	setting
demography	yellow	yellow	green
sociology	yellow	red	green
linguistics	yellow	yellow	green
psychology	red	red	green
history	yellow	yellow	green
politics etc.	yellow	green	green
geography	green	green	green
education	green	green	green
religion	green	yellow	yellow
economics	yellow	red	green
technology	yellow	yellow	green

◆ good data   ◆ reasonable data   ◆ bad data

Figure 1 'Bad data' analysis for Norn

A factor of unclear influence is the disappearance of a Norn-language ballad dancing tradition, as can still be found in the Faroe Islands: there is no convincing evidence for a causal relationship either way. It should also be mentioned that there is no record of Norn being actively combatted or of Scots or English being forcefully imposed on the Shetlanders like there were anti-Gaelic measures in the Highlands of Scotland in the same period.

Norn: bad data?

If we are to systematize the available data for the Norn–Scots language shift along the lines of Edwards' typology, we can have certain expectations. We can safely assume that general facts are more readily accessible than records of attitudes; as attitudes are generally asked for in the 'language' category of questions, we can expect these to be especially affected by the bad data problem. Both the 'speaker' and 'setting' categories are primarily concerned with facts, but we can also expect a difference between the available data for these categories: as most of the surviving data was written down by the Scots ('setting') population and with a Scots point of view, the 'setting' category should have better data than the 'speaker' category, for which we have little first-hand data.

An analysis of the available Norn data (cf. Figure 1) confirms these expectations.[3] It is clear that the best data is available in the 'setting' category, and

---

[3] Such an analysis is naturally highly subjective. It is therefore I have only divided the data into three categories, despite the fact that there is variation within categories. The categories should be read so that 'good' means the data is enough to base conclusions on without further ado; 'reasonable' means that the data gives us a fair

although there is not much difference between the 'speaker' and 'language' categories, the 'language' category seems to be the most affected by the bad data problem.

There is a number of deviances from the expected quality of data. These are generally due to inconsistencies in the model. The 'language' category, for example, deals mostly with attitudes, but in the areas of politics and education, the question relates to actual fact. Conversely, the 'setting' question for religion deals with attitude rather than the expected factual question.

It is also clear that the psychology questions are badly affected by the bad data problem, but as these deal with attitudes more than any other category, this is to be expected. The good data that is available for the questions on education is a result of the detailed research on Shetland education by John Graham (1998) mentioned earlier.

It is generally possible to fit the earlier research results about the Norn–Scots language shift into Edwards' model. The importance of the use of Scots in church, administration and law can be discussion in the 'religion/language' and 'politics/language' cells, respectively. The model allows for a discussion of the effects of immigration in the 'linguistics/setting' cell; the sample questions do not mention language contacts in general as Edwards only mentions contacts between the majority and the minority language groups, but given Edwards' emphasis on the preliminary state of the questions, I would argue that other language contacts, too, can be discussed in the 'linguistics/setting' cell. (Perhaps even more so than migration patterns, which would also fit in the 'demography/setting' cell.)

A discussion of language-specific cultural utterances could perhaps be placed in the 'psychology/language' cell, but this is unclear. The absence of prohibitive language legislation fits into the 'politics/language' cell.

Conclusions

The typology of language endangerment by John Edwards (1992) has been specifically designed to describe minority language situations, and investigates a large number of social factors that influence a language group's ethnolinguistic vitality. This would theoretically make the model ideal for use in historical socio-linguistic research into cases of language shift.

The bad data problem that plays a role in historical sociolinguistics also influences this research, perhaps even more so because the power dynamics in literacy, which decided what information was written down, were unbalanced in favour of the dominant language population. This makes first-hand data from the

---

idea, but we lack sufficient detail; and 'bad' means we do not have enough data to make defendable conclusions about the situation.

minority population extremely scarce. This majority bias is reflected in the way the bad data problem affects elements of Edwards' typology: 'setting' (majority) factors are generally less affected than 'speaker' (minority) factors. Data for the 'language' factors, which tend to deal with attitudes, is affected even more.

An attempt to fit earlier research results about the language shift from Norn to Scots in Shetland into Edwards' typology shows that the factors that were identified as playing a role in the language shift can be described using Edwards' model. A slight adaptation of the model is necessary to allow for a discussion of language contacts other than majority/minority interaction, but this is a flaw that is inherent to the model rather than a result of using a modern model for historical research.

In conclusion, the research in this paper would suggest it is possible to apply modern theory and modern models to non-modern situations. Further research into historical cases of language shift with the use of Edwards' typology can therefore be expected to foster viable results.

References

Ballantyne, John H. & Brian Smith (eds.) (1994). *Shetland documents 1580–1611*. Lerwick: Shetland Islands Council & Shetland Times.

Ballantyne, John H. & Brian Smith (eds.) (1999). *Shetland documents 1195–1579*. Lerwick: Shetland Islands Council & Shetland Times.

Barnes, Michael P. (1984). 'Orkney and Shetland Norn', in Peter Trudgill (ed.), *Language in the British Isles*. Cambridge: Cambridge University Press, 352–366.

Barnes, Michael P. (1998). *The Norn language of Orkney and Shetland*. Lerwick: Shetland Times.

Bourhis, R.Y., H. Giles & D. Rosenthal (1981). 'Notes on the construction of a "subjective vitality questionnaire" for ethnolinguistic groups', *Journal of Multilingual and Multicultural Development* **2**, 145–155.

Clyne, Michael (2003). *Dynamics of language contact: English and immigrant languages*. Cambridge: Cambridge University Press.

Donaldson, Gordon (1983). 'The Scots settlement in Shetland', in Donald J. Withrington (ed.), *Shetland and the outside world 1469–1969*. Oxford: Oxford University Press, 8–19.

Dorian, Nancy C. (1982). 'Linguistic models and language death evidence', in Loraine K. Obler & Lise Menn (eds.), *Exceptional language and linguistics*. New York, London: Academic Press, 31–48.

Edwards, John (1992). 'Sociopolitical aspects of language maintenance and loss: towards a typology of minority language situations', in Willem Fase et al. (eds.), *Maintenance and loss of minority languages*. Amsterdam: Benjamins, 37–54.

Fishman, Joshua A. (1972a). 'Language maintenance and language shift as a field of inquiry: revisited', in Anwar S. Dil (ed.), *Language in sociocultural change: essays by Joshua A. Fishman*. Stanford CA: Stanford University Press, 76–135.

Fishman, Joshua A. (1972b). 'The relationship between micro- and macro-sociolinguistics in the study of who speaks what language to whom and when', in Anwar S. Dil (ed.), *Language in sociocultural change: essays by Joshua A. Fishman*. Stanford CA: Stanford University Press, 244–267.

Fishman, Joshua A. (1991). *Reversing language shift: theoretical and empirical foundations of assistance to threatened languages*. Clevedon, Philadelphia, Adelaide: Multilingual Matters.

Gal, Susan (1979). *Language shift: social determinants of linguistic change in bilingual Austria*. New York: Academic Press.

Giacalone Ramat, Anna (1983). 'Language shift and language death: a review of Nancy C. Dorian, *Language death* and Susan Gal, *Language shift*', *Folia linguistica* **27**.3–4, 495–507.

Giles, H., R.Y. Bourhis & D.M. Taylor (1977). 'Towards a theory of language in ethnic group relations', in H. Giles (ed.), *Language, ethnicity and intergroup relations*. London: Academic Press, 307–348.

Govindasamy, Subra & Mohana Nambiar (2003). 'Social networks: applicability to minority communities in multilingual settings', *International Journal of the Sociology of Language* **161**, 25–45.

Graham, John J. (1998). '*A vehement thirst after knowledge': four centuries of education in Shetland*. Lerwick: Shetland Times.

Grenoble, Lenore & Lindsay J. Whaley (1998). 'Toward a typology of language endangerment', in Lenore Grenoble & Lindsay J. Whaley (eds.), *Endangered languages: language loss and community response*. Cambridge: Cambridge University Press, 22–54.

Haugen, Einar (1972). 'The ecology of language', in Anwar S. Dil (ed.), *The ecology of language: essays by Einar Haugen*. Stanford CA: Stanford University Press, 325–339.

Houston, R.A. (2003). '"Lesser-used" languages in historic Europe: models of change from the 16th to the 19th centuries', *European Review* **11**, 299–324.

Hulsen, Madeleine, Kees de Bot & Bert Weltens (2002). '"Between two worlds": social networks, language shift, and language processing in three generations of Dutch migrants in New Zealand', *International Journal of the Sociology of Language* **153**, 27–52.

Kloss, Heinz (1966). 'German-American language maintenance efforts', in Joshua A. Fishman, *Language loyalty in the United States: the maintenance and perpetuation of non-English mother tongues by American ethnic and religious groups*. Den Haag: Mouton, 206–252.

Knooihuizen, Remco (2005). 'The death of Norn: a study into the decline of the Scandinavian vernacular of Orkney and Shetland.' Unpublished M.Sc. dissertation, University of Edinburgh.

Knooihuizen, Remco (2006). 'The Norn-to-Scots language shift: another look at socio-historical evidence', *Northern Studies* **39**, 105–117.

Labov, William (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William (1994). *Principles of linguistic change. Volume 1: internal factors*. Oxford: Blackwell.

Lippi-Green, Rosina (1994). *Language ideology and language change in Early Modern German*. Amsterdam: Benjamins.

Low, George (1879). *A tour through the islands of Orkney and Schetland containing hints relative to their ancient, modern and natural history, collected in 1774*, ed. by Joseph Anderson. Kirkwall: Peace.

Mackey, William F. (1980). 'The ecology of language shift', in P.H. Nelde (ed.), *Sprachkontakt und Sprachkonflikt*. Wiesbaden: Steiner, 35–41.

Melchers, Gunnel (1991). 'Norn-Scots: a complicated language contact situation in Shetland', in P. Sture Ureland & George Broderick (eds.), *Language contact in the British Isles: proceedings from the Eighth International Symposium on Language Contact in Europe, Douglas, Isle of Man, 1988*. Tübingen: Niemeyer, 461–477.

Milroy, Lesley (1987). *Language and social networks*. Oxford: Blackwell.

Nevalainen, Terttu (1999). 'Making the best use of "bad" data: evidence for sociolinguistic variation in Early Modern English', *Neuphilologische Mitteilungen* **100**, 499–533.

Nevalainen, Terttu & Helena Raumolin-Brunberg (2003). *Historical socio-linguistics: language change in Tudor and Stuart England*. London: Longman.

Sasse, Hans-Jürgen (1992). 'Theory of language death', in Matthias Brenzinger (ed.), *Language death: factual and theoretical explorations with special reference to East Africa*. Berlin, New York: Mouton de Gruyter, 7–30.

Smith, Brian (1990). 'Shetland, Scandinavia, Scotland, 1300–1700: the changing nature of contact', in Grant G. Simpson (ed.), *Scotland and Scandinavia 800–1800*. Edinburgh: Donald, 25–37.

Smith, Brian (2000). *Toons and tenants: settlement and society in Shetland, 1299–1899*. Lerwick: Shetland Times.

Stoessel, Saskia (2002). 'Investigating the role of social networks in language maintenance and shift', *International Journal of the Sociology of Language* **153**, 93–131.

Thomason, Sarah Grey & Terence Kaufman (1991). *Language contact, creolization, and genetic linguistics*. Berkeley, Oxford: University of California Press.