Volume 30:2017 **sky**

SKY Journal of Linguistics

Editors:

Maija Hirvonen, Johanna Isosävi, Milla Luodonpää-Manni,
Kaarina Mononen, Mirka Rauniomaa

SKY Journal of Linguistics is published by the Linguistic Association of Finland (one issue per year).

Notes for Contributors

Policy: *SKY Journal of Linguistics* welcomes unpublished original works from authors of all nationalities and theoretical persuasions. Every manuscript is reviewed by at least two anonymous referees. In addition to full-length articles, the journal also accepts short (3–9 pages) 'squibs' as well as book reviews.

Language of Publication: Contributions should be written in English, French, or German. If the article is not written in the native language of the author, the language should be checked by a professional in that language.

Style Sheet: *SKY Journal of Linguistics* follows the Generic Style Rules for Linguistics, with complementary house rules, which can be found at http://www.linguistics.fi/skystyle.shtml.

Abstracts: Abstracts of the published papers are included in *Linguistics Abstracts* and *Cambridge Scientific Abstracts*. The published papers are included in *EBSCO Communication & Mass Media Complete*. *SKY JoL* is also indexed in the *MLA Bibliography*.

Editors' Addresses (2017):

Maija Hirvonen, Translation Studies and Terminology Research Community, P.O. Box 24 (Unioninkatu 40), FI-00014 University of Helsinki, Finland

Johanna Isosävi, Department of Modern Languages, P.O. Box 24 (Unioninkatu 40), FI-00014 University of Helsinki, Finland

Milla Luodonpää-Manni, French Department, FI-20014 University of Turku, Finland

Kaarina Mononen, Department of Finnish, Finno-Ugrian and Scandinavian Studies, P.O. Box 3 (Fabianinkatu 33), FI-00014 University of Helsinki, Finland

Mirka Rauniomaa, English Philology, Faculty of Humanities, P.O. Box 1000, 90014 University of Oulu, Finland

Editors' E-mail: sky-journal(at)helsinki(dot)fi

Publisher:

The Linguistic Association of Finland
c/o General Linguistics
P.O. Box 24 (Unioninkatu 40)
FI-00014 University of Helsinki
Finland
http://www.linguistics.fi, http://www.linguistics.fi/skyjol.shtml

The Linguistic Association of Finland was founded in 1977 to promote linguistic research in Finland by offering a forum for the discussion and dissemination of research in linguistics, both in Finland and abroad. Membership is open to anyone interested in linguistics. The membership fee in 2017 was EUR 25 (EUR 15 for students and unemployed members). Members receive *SKY Journal of Linguistics* gratis.

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

Cover design: Timo Hämäläinen 1999

SKY Journal of Linguistics
30

Suomen kielitieteellisen yhdistyksen aikakauskirja
Tidskrift för den Språkvetenskapliga föreningen i Finland
Journal of the Linguistic Association of Finland

2017

ISSN-L: 1456-8438
ISSN: 1456-8438 (Print)
ISSN: 1796-279X (Online)

Painosalama, Turku 2017

Contents

Reviewers of SKY JoL 30 (2017)

Terhi Ainiala (University of Helsinki), Marios Andreou (Heinrich Heine University Düsseldorf), Arto Anttila (Stanford University), Östen Dahl (Stockholm University), Kimmo Granqvist (Södertörn University, Stockholm), Matti Kilpiö (University of Helsinki), Vadim Kimmelman (University of Amsterdam), Niina Kunnas (University of Oulu), Aki-Juhani Kyröläinen (University of Turku), Bahar Köymen (University of Manchester), Heini Lehtonen (University of Helsinki), Ilkka Lindstedt (University of Helsinki), Tapani Möttönen (University of Helsinki), Tommi Nieminen (University of Eastern Finland), Stéphane Pagès (Aix-Marseille University), Przemyslaw Pawelec (University of Wroclaw), Saila Poutiainen (University of Helsinki), Päivi Rainò (Humak University of Applied Sciences), Jari Sivonen (University of Oulu), Krista Teeri-Niknammohgadam (University of Turku), Reetta Toivanen (University of Helsinki), Maria Vilkuna (University of Helsinki), Mikko Virtanen (University of Helsinki), Eva Thue Vold (University of Oslo), Mari Wiklund (Helsinki Collegium for Advanced Studies)

The editors are most grateful to all the scholars who have acted as reviewers for SKY Journal of Linguistics in 2017, including those who wish to remain anonymous and are therefore not listed here.

Types and trends of name signs in the Swedish Sign Language community

Carl Börstell
Stockholm University

Abstract

This paper investigates the domain of name signs (i.e., signs used as personal names) in the Swedish Sign Language (SSL) community. The data are based on responses from an online questionnaire, in which Deaf, hard of hearing, and hearing participants answered questions about the nature of their name signs. The collected questionnaire data comprise 737 name signs, distributed across five main types and 24 subtypes of name signs, following the categorization of previous work on SSL. Signs are grouped according to sociolinguistic variables such as age, gender, and identity (e.g., Deaf or hearing), as well as the relationship between name giver and named (e.g., family or friends). The results show that name signs are assigned at different ages between the groups, such that children of Deaf parents are named earlier than other groups, and that Deaf and hard of hearing individuals are normally named during their school years. It is found that the distribution of name sign types is significantly different between females and males, with females more often having signs denoting physical appearance, whereas males have signs related to personality/behavior. Furthermore, it is shown that the distribution of sign types has changed over time, with appearance signs losing ground to personality/behavior signs – most clearly for Deaf females. Finally, there is a marginally significant difference in the distribution of sign types based on whether or not the name giver was Deaf. The study is the first to investigate name signs and naming customs in the SSL community statistically – synchronically and diachronically – and one of the few to do so for any sign language.

Keywords: name sign, onomastics, anthroponyms, variation, sign language, Swedish Sign Language, naming customs

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

1 Introduction

One phenomenon reported from many sign language communities around the world is the use of so-called *name signs* (or, alternatively, *sign names*). A name sign is a sign that is used with reference to an individual[1] and is normally used alongside a spoken language community name. Name signs are important as identity markers in the community because it is often a symbol of inclusion, that is, individuals with name signs are mostly those that are members of the community, although name signs are also given to individuals often talked about within the community without themselves being members, such as famous people or others with some relevance to the community (McKee 2016: 805).[2] Name signs are not merely nicknames/bynames or hypocorisms, but the default way of denoting an individual in the community. However, name signs are not used vocatively, since calling somebody's attention is done by physical contact, unless one is already visible to the addressee (Deaf Studies 1997: Chap. 4).

The formation of name signs is based on different types of motivations. Many name signs are descriptive in some way, such that they iconically portray the visual appearance or physical mannerisms of an individual. Other name signs are motivated by the individual's spoken name (Supalla 1990). Figures 1 and 2 show Swedish Sign Language (SSL) name signs of famous people that are motivated based on the physical appearance of their referents: in Figure 1, Donald Trump's hair moving in the wind is depicted; in Figure 2, Elvis's sideburns are depicted.

[1] A name sign is, in a wider definition, any sign used as a proper noun, such as toponyms (see Nonaka 2015; Nyström 2002), company names, etc. In this paper, the term *name sign* is used specifically to mean signs that denote individuals – i.e., anthroponyms.

[2] Ingela Holmström (personal communication) notes that also non-humans, such as pets, receive name signs of their own, at least in the SSL community.

Figure 1. The SSL sign for Donald Trump depicting hair blowing up and down in the wind (Svenskt teckenspråkslexikon 2017 sign #16543)

Figure 2. The SSL sign for Elvis Presley depicting thick sideburns (Svenskt teckenspråkslexikon 2017 sign #07430)

As for name signs motivated by a person's behavior/mannerisms or dynamic appearance, we have the SSL sign for Charlie Chaplin depicting his characteristic twirling of a cane, as shown in Figure 3.

Figure 3. The SSL sign for Charlie Chaplin depicting the twirling of a cane (Svenskt teckenspråkslexikon 2017 sign #14983)

Some name signs may be a combination of these types of depiction motivations, such as the SSL sign for Adolf Hitler, sequentially showing his mustache (i.e., static) followed by a reduced depiction of a Nazi salute (i.e., dynamic) – see Figure 4.

Figure 4. The SSL sign for Adolf Hitler depicting a mustache and a Nazi salute (Svenskt teckenspråkslexikon 2017 sign #10086)

Previous research on name signs in SSL has shown that physical appearance motivated signs are common (Hedberg 1989; 1994). However, the exact distribution of name signs across types and subtypes of motivations has not been systematically investigated for SSL, nor the possibly change in distribution over time, and whether or not social and group identification factors have an influence on the distribution. Thus, the main aim of this study is to investigate sociolinguistic properties of the assignment of name signs in the SSL community (e.g., identity of name bearer/giver, naming customs changing over time), in order to evaluate whether the findings from the previous study have changed over time, and add to the results by including more variables (see §3).

2 Background

Name signs have been documented in many of the world's sign language communities. However, a systematic investigation of the types and distribution of name signs of these communities is quite limited, one exception being Rainò's (2004) dissertation on name signs in Finnish Sign Language. Descriptions of name sign customs are found for American Sign Language (Meadow 1977; Mindess 1990; Supalla 1990), British Sign Language (Day & Sutton-Spence 2010), Chinese Sign Language (Yau & He 1989), Estonian Sign Language (Paales 2010; 2011), Finnish Sign

Language (Rainò 2004), French Sign Language (Delaporte 1998), Greek Sign Language (Kourbetis & Hoffmeister 2002), Japanese Sign Language (Nonaka et al 2015), New Zealand Sign Language (McKee & McKee 2000), Quebec Sign Language (Dubuisson & Desrosiers 1994), Russian Sign Language (Esipova 2013), Swedish Sign Language (Hedberg 1989; 1994; 2009), and the sign languages of Uganda, Mali, The Netherlands, and the Adamorobe village (Nyst & Baker 2003).

Although some Deaf[3] children receive their name signs early, a trend found in several sign language communities seems to be that a name sign is often acquired during school years (Hedberg 1989: 8; Yau & He 1989: 306; McKee & McKee 2000: 22–23). This means that unlike spoken language community names, name signs are often assigned to individuals by their own peers rather than the parents/family (McKee & McKee 2000: 23). However, if the name sign is given by Deaf parents to their child, the naming may take place immediately after birth in some communities (Esipova 2013: 2). Also, it is known in several Deaf communities that an individual may be bestowed with more than one name sign during one's lifetime, and one's name sign may also change (Hedberg 1989: 46; Day & Sutton-Spence 2010: 47; McKee 2016: 805), although some sign language communities seem to use a single sign for an individual throughout that person's life (Kourbetis & Hoffmeister 2002: 42).

Among the sign languages for which name signs have been studied, American Sign Language (ASL) and Japanese Sign Language (JSL) stand out in that non-descriptive name signs are very common. In ASL, many name signs are simply initialized (i.e., using the manual alphabet handshape of the name's initial letter), and these arbitrary names are argued to form the native naming system, as this is what Deaf parents tend to use when naming their children (Supalla 1990: 121). Interestingly, in the British Sign Language (BSL) community it also seems as though Deaf parents are more likely to resort to fingerspelling when naming their children, as compared to hearing parents (Day & Sutton-Spence 2010: 46–47). However, in Greek Sign Language, initialized signs tend to be used mostly as a result of hearing people entering the Deaf community (Kourbetis & Hoffmeister 2002: 35). In JSL, the biggest category of name signs is translating the Japanese name – that is, the meaning of the individual morphemes of a name, if corresponding to regular words – into a

[3] Capital D "Deaf" is used to refer to the cultural and linguistic community identity centered around deaf individuals, distinguishing it from a clinical label (Schembri & Lucas 2015).

sign-by-sign corresponding form in JSL (Nonaka et al 2015: 78). Apart from these languages, other sign languages studied seem to prefer descriptive name signs (McKee 2016: 807). More specifically, name signs depicting the appearance (e.g., body shape, facial features, hair, clothes) constitute the biggest group in New Zealand Sign Language (McKee & McKee 2000: 13), Finnish Sign Language (Rainò 2004: 102), and Greek Sign Language (Kourbetis & Hoffmeister 2002), and it is one of the biggest in BSL (Day & Sutton-Spence 2010: 48) and Swedish Sign Language (Hedberg 1994: 419).

In ASL, there are some name signs that are associated with gender by being articulated high (male) or low (female) on the face, following a general gender-based high–low distinction in the lexicon, and similar cases of location or handshapes have also been found for Asian sign languages, but not for BSL (Day & Sutton-Spence 2010: 27; McKee 2016: 808). There are no obvious indications that descriptive name signs in SSL are overtly gendered by default, seeing as there are several name sign forms used to denote both men and women.[4]

For SSL, the use of name signs is documented already in the first dictionary of the language, written by Deaf Swede Oskar Österberg (1916), in which they are referred to as "characterization signs" or "sign language names". Österberg writes that such signs are assigned to students when entering the Deaf schools, and gives examples such as 'upturned nose', 'gray coat', 'round eye', 'red behind ear', and 'striped pants' (Österberg 1916: 17). However, the most – and in reality only – systematic study of name signs in SSL was conducted by Hedberg (1989; 1994), following a documentation project of the history of and experiences within the Deaf community conducted in the late 1980s (Hedberg 2009). Hedberg collected 3,114 name signs, of which he had etymological descriptions – mostly provided by the name bearers themselves – for just over a third (n = 1,108). When documenting these signs and their etymology, Hedberg describes that he could discern a number of different types and subtypes of name signs. These types have formed the basis of this study. The categories that Hedberg (1994) lists are as follows: *appearance*, based on the (static) visual attributes of the referent (e.g., body shape, hairstyle, or clothes); *mannerisms*, denoting behavior (e.g., signs, gestures, or movements used often); *social group*, referring to an individual by their occupation or place

[4] I am grateful to Bernhard Wälchli for raising the question of overt vs. covert gender in name signs.

of origin (e.g., the sign DANMARK[5] 'Denmark' for a Danish person); *name*, using the spoken language community name as the basis of the name sign (e.g., a sign that translates into the same or a similar word, or a initialized sign); and finally *numbers*, for name signs that are based on an identification number given to the individual in the Deaf school.[6] Concerning the *name* type category, there is an interesting group of signs in SSL that consists of fixed name signs, that is, for some spoken language community names, there is a fixed sign that can be used to refer to a specific name (although not all bearers of that name need to have the fixed form). For instance, the name *Anders* has a fixed sign in SSL identical to the sign for 'prankish, playful' (Swe. *busig*) (see Svenskt teckenspråkslexikon 2017 sign #02337).[7]

The types of name signs investigated in this study are mainly those described by Hedberg (1994), although with a slightly altered categorization. The categories are presented in §3 below.

3 Data and methodology

Since previous studies of name signs in the SSL community are becoming outdated, the purpose of this study was to update the research in terms of which types of signs are currently found in the community, having present-day community members answer questions about their name signs rather than also documenting signs from previous – possibly no longer living – generations. Also, previous research has not presented quantitative data on the distribution of name signs across the different sign types, or how the preferences for certain categories may be influenced by the identity of the name giver/bearer or change over time, specifically taking into account the time of naming. Thus, these points were addressed in the current study. In

[5] The convention in sign language research is to refer to signs by the use of word labels in small caps.

[6] The category *number* is not included in the current study, as it was very marginal and archaic already in Hedberg's (1994) data, and is not productive in forming name signs today.

[7] The specific case of *Anders* is, according to several consulted Deaf signers, based on a specific individual named Anders who, during his years in the Deaf school, often came up with pranks and (harmless) mischiefs, hence receiving this name sign himself. Later, this association has been made a fixed one between the written name and the sign, rather than between the individual and the sign.

the end, three main research questions were formulated, which were stated as follows.

1. When and by whom is a name sign given, and does it differ between identity groups?[8]
2. What is the distribution of types of name signs, and does it differ based on the identity and gender of the name giver and bearer of the name sign?
3. Has the distribution of types of name signs changed over time?

In order to investigate these questions quantitatively, a large sample of name signs was needed. An online questionnaire was set up and distributed via Deaf groups on social media (mainly Facebook). Anyone could access and answer the questionnaire, and multiple responses from a single individual were allowed since one person may have (had) more than one name sign. The questionnaire consisted of four steps and a total of twelve questions: four concerning the type/subtype/origin/rating of the name sign and eight concerning the participant/name giver and the circumstances of the naming. In the first step, the study and the questionnaire were described. In the second step, the participant was asked to categorize their name sign as belonging to one of five main types in order to specify the sign language from which the sign was taken (SSL or other)[9] and to rate their perception of their own name sign from 1 to 5 (negative to positive). In the third step, the participant was asked for the subtype of the name sign, dependent on the main type chosen earlier (i.e., the specific categories to choose from in the third step was conditioned by the choice made in the second step). Finally, in the fourth and last step, the participant was asked to fill out information about their own and the name giver's identity – i.e., if they identify as Deaf, hard of hearing, hearing, or so-called "Coda" (*Child of Deaf Adult*, referring to a hearing person with a Deaf parent(s)) – in addition to the time and place of naming, and the present age of the name bearer and their age at the time of naming. No personal information that could identify a participant was requested by the questionnaire, leaving the participants anonymous.

[8] The term 'identity group' is used in this paper to refer to an individual's membership with regard to the groups Coda, Deaf, hard of hearing, and hearing, i.e., what the individual identifies as (as opposed to a clinical definition of hearing).

[9] The vast majority of signs collected come from SSL (683 vs. 44). In the following, all name signs are considered regardless of origin, since they are all found in the language community.

The main types and subtypes of name signs included in the questionnaire are described in the following.

Appearance: Signs denoting the visual appearance of the name bearer. This category contains five subtypes: *hair* (e.g., 'curl', 'bangs'); *physical feature* (e.g., 'dimple', 'blue-eyed'); *clothes* (e.g., 'striped shirt', 'glasses', 'ribbon'); *metaphor* (e.g., 'lighthouse' to refer to being tall); and *other* (if none of the above applied).

Person: Signs denoting the character or doings of the bearer's person. This contains seven subtypes: *personality* (e.g., 'laugh', 'happy'); *hobby* (e.g., 'ice-skates' for liking to skate); *profession* (e.g., 'carpenter'); *gesture* (a characteristic movement, e.g., scratching one's chin); *metaphor* (e.g., 'monkey' for being good at climbing trees); *sign* (a sign associated with the individual because of frequent use, e.g., often using a sign with the meaning 'oh really?' in conversation); and *other* (if none of the above applied).

Name: Signs relating to the spoken language community name of the bearer. This category consists of eight subtypes: *fixed* (using one of the fixed name signs, see §1); *whole name*, using a sign corresponding to the exact form of the name (e.g., BJÖRN 'bear' for *Björn* – name meaning 'bear'); *partial name*, using a sign corresponding to the form of a part of the name (e.g., BJÖRK 'birch' for *Björkstrand* – literally 'birch beach'); (*partial*) *homonym*, a sign for a word that sounds/looks like the name (e.g., KATT 'cat' for *Katarina*); *extension*, a sign for a related concept (e.g., BLOMMA 'flower' for *Florén*); *initialized*, the first letter of the name using the handshape from the manual alphabet; *fingerspelled*, fingerspelling the name using the manual alphabet; and *other*, if none of the above applied.

Background: Signs denoting some aspect of the referent's background. This consists of three subtypes: *country* (the sign for the native country of the bearer); *city* (the sign for the hometown of the bearer); *inherited* (if the name sign is inherited from a relative).

Other: Signs that do not fall into any of the above categories. This category did not contain any subtypes.

The main types are motivated by the bearer's being (*Appearance*), doing (*Person*), or history (*Background*), with the type *Name* being motivated simply on the form or meaning of the written name.[10] This is a distinction different from the simple distinction of descriptive vs. arbitrary

[10] As noted by an anonymous reviewer, some *Person* name signs are also visually salient (i.e., related to appearance). The distinction made here is based on static vs. dynamic properties.

that has been used in some previous studies (e.g., Supalla 1990), but also more informative by following a more detailed categorization based on previous work on SSL (i.e., Hedberg 1989). Since Hedberg's work had provided a detailed description of the possible name sign types found in SSL, this study aimed to let participants match their name signs to these established categories, rather than document the exact form of the name signs. The reason for this is twofold: first, this avoids the problem of not being able to classify a poorly described name sign motivation into categories; second, it allows for swift and anonymous collection of distributional data. Thus, whereas Hedberg's aim was to document forms and types, this study aims to describe the distribution of Hedberg's types.

The questionnaire was open January 5–28, 2017. After the responses were collected, the data were exported and compiled using a custom Python script, and then analyzed using the statistical programming language R and plotted with the R package ggplot2 (Wickham 2009). In total, 737 name signs were collected, distributed across a range of ages (median birth year: 1980; see Figure 5), and gender and identity groups (Table 1). As can be seen in Figure 5, the majority of the name signs in the data come from participants born in 1960–2000.

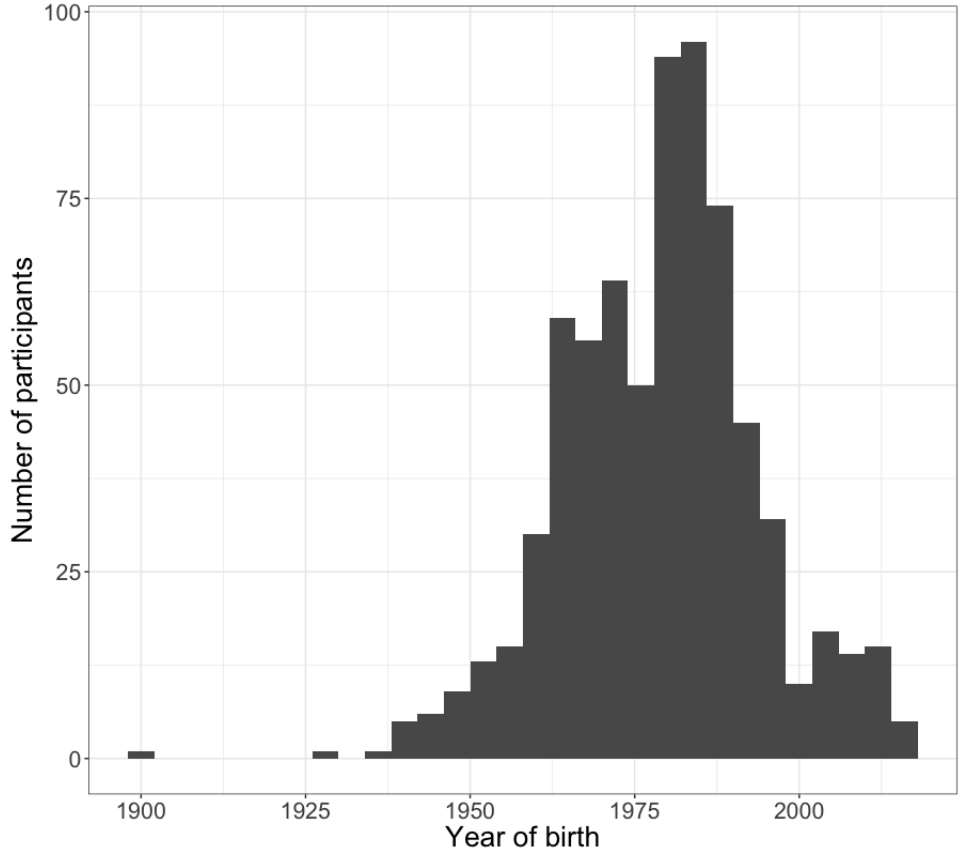

Figure 5. Distribution of birth years among participants[11]

Table 1 shows that the distribution across genders is not balanced, with the female group being in clear majority across all identity groups (especially skewed for the hearing group). Also, the Deaf and hearing groups are by far the biggest, with the hard of hearing and Coda groups falling behind.

Table 1. Distribution of name signs across genders and identity groups

	Female	**Male**	**Other**	***Total***
Deaf	194	108	3	305 (41%)
Hard of hearing	55	17	3	75 (10%)
Coda	60	25	0	85 (12%)
Hearing	221	48	3	272 (40%)
Total	530 (72%)	198 (27%)	9 (1%)	737

[11] It is possible that the single data point at 1900 as the year of birth is erroneous. This should not, however, affect the data to any significant extent. Some participants explained that they filled out the questionnaire for their children, which is why there are data from names given as late as 2016.

4 Results

4.1 When and by whom a name sign is given

A first thing to look at in the data is when the name signs were given, to see if there are differences across the identity groups. Figure 6 shows the distribution of ages at which a name sign was given, across the four identity groups. As can be seen from this figure, there is a clear but unsurprising pattern for the different groups. Whereas the Coda group generally receive their name signs early (median 1 year old), the Deaf group participants often receive their name signs when they start school (median = 7), and the hard of hearing group towards the end of the school years (median = 15). While some in the Deaf and hard of hearing groups may have Deaf parents themselves, most individuals in these groups have hearing parents (McKee & McKee 2000: 23), which accounts for the later age of naming in these groups – that is, only Coda group individuals are assured to have early contact with the Deaf community. Turning to the hearing group, they are the only group mostly getting their name signs in adulthood (median = 22.5). This is unsurprising, seeing as hearing individuals who are not born into a Deaf/signing family would normally not encounter sign language or the Deaf community until after graduating from high school and, for instance, taking a sign language class.

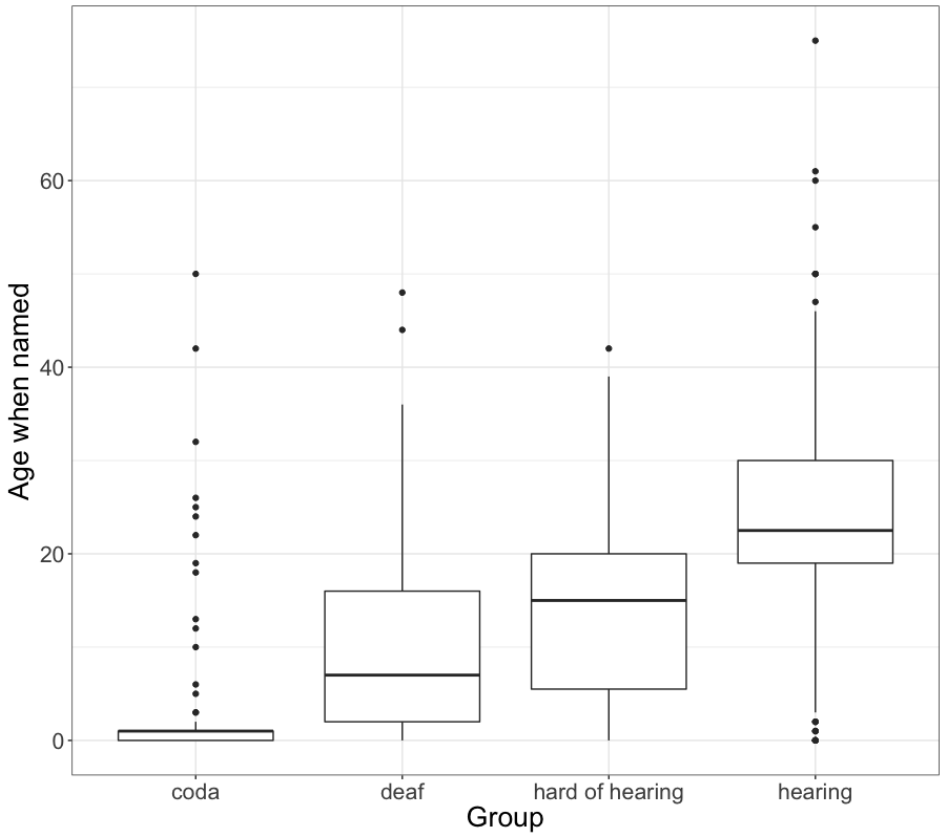

Figure 6. Age when given a name sign across the four identity groups

Looking at the identity of the name giver, rather than the name bearer, we also see some clear patterns across the groups. Figures 7 and 8 show the identity of the name giver across the identity groups of the bearer. First, in Figure 7 we see the distribution of name givers based on their relationship with the name bearer. We here see what we anticipated from Figure 6, namely that the Coda group is mostly named by their family – unsurprising as they are born into a Deaf family – whereas all other groups have a minority of family given name signs. The most common name giver for both the Deaf and hard of hearing groups is a friend, which is a pattern also found among Deaf individuals in, for instance, the British Sign Language (McNamara 2003 cited in Day & Sutton-Spence 2010: 44) and American Sign Language (Mindess 1990: 5) communities. Although the school years constitute the time of naming for the majority of participants in the Deaf and hard of hearing groups, the number of name signs given by teachers is fairly low, especially for the Deaf group, showing that the peer assignment

of names is stronger than the intergenerational assignment of names for the participants of this study.[12]

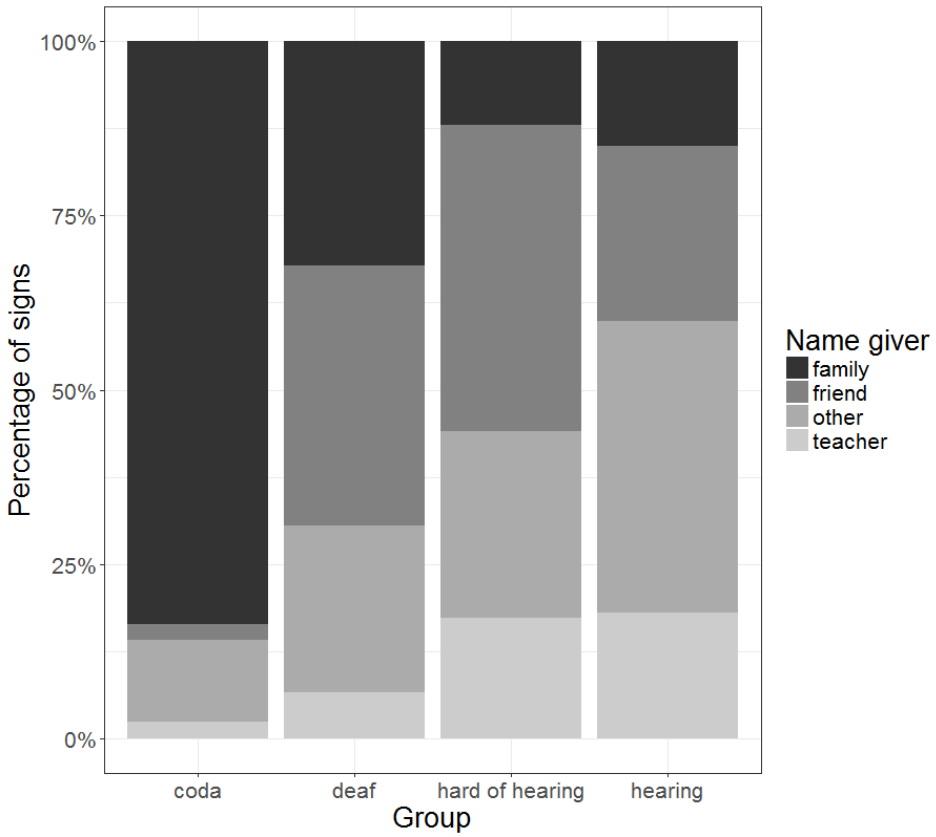

Figure 7. Relationship to name giver across identity groups

The quite high proportion of *other* categorized name givers is possibly due to the name bearer not knowing exactly who assigned the name sign. This is especially reasonable considering that the *other* group is the largest among the hearing participants, for which it is likely that the name bearers are not full-fledged members of the community and might therefore have been named in their absence.[13]

[12] As pointed out by an anonymous reviewer, it is possible that there will be noticeable changes in these patterns with the advent of cochlear implants (CI) and the growing number of children born deaf/hard of hearing who do not attend Deaf schools (Holmström & Schönström 2017). By not going to a Deaf school with signing peers, it is possible that name sign assignment will take place under different circumstances. However, due to the limited data for those born since the year 2000, and a lack of detailed metadata about each participant, I leave this issue to future research.

[13] I thank Östen Dahl for raising the issue of known name givers. However, according to Johanna Mesch (personal communication), most Deaf people can identify the exact individual who gave them their name sign.

Turning to Figure 8, we see again that the most common name giver for the Coda group is Deaf, indicating that the parents are the default name givers, although it is also possible that, for instance, these individuals also have Deaf siblings.[14] Interestingly, the name giver is most often Deaf across all identity groups of the bearers. For the Deaf bearers, they have received their name sign from a Deaf individual in about three quarters of the cases. For the hearing group, this is a little less common, and for the hard of hearing group, the name giver is Deaf in just over half of the cases. The interesting thing here is that the second most common name giver identity group is the same as the bearer's identity group for the hearing and hard of hearing groups, suggesting patterns of in-group interaction.

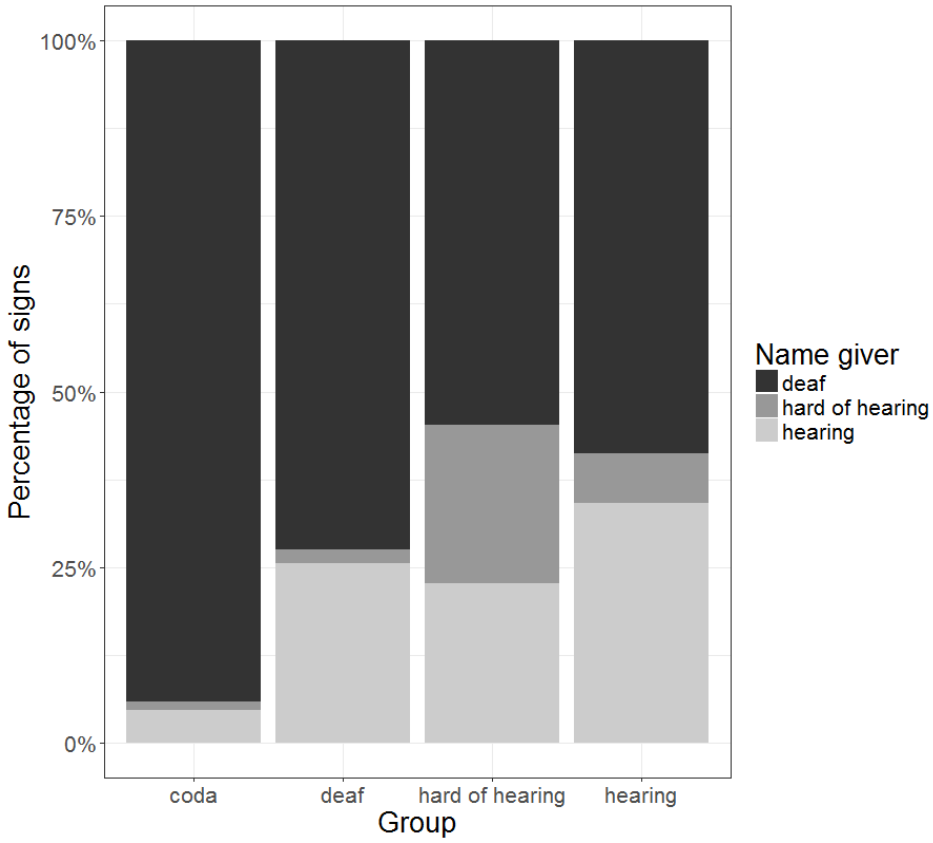

Figure 8. Identity of name givers across identity groups

4.2 Distribution of types of name signs

Moving on to the types and subtypes of name signs in the data, we find that all the possible types included in the questionnaire (based on the types in

[14] I thank an anonymous reviewer for this last point.

Hedberg 1994) are found, but they differ in terms of frequency. Table 2 shows the distribution of name signs in the data, across types and subtypes. As can be seen here, the majority of name signs are either based on the bearer's appearance or person. With almost 16% of the total number, name signs that relate to one's spoken language community name are also fairly common. On the other hand, the signs denoting one's background (geographical affiliation or inherited family sign) are very infrequent. Less than a tenth of the total number of signs could not be categorized into any of the four basic types. Among those *other* classified name signs, participants reported various motivations, such as a specific incident, for instance being named after a sign once used incorrectly by accident, or something the bearer dislikes (as opposed to liking – e.g., hobbies, etc.).[15]

Within the appearance type signs, we see that physical features and hair are the most common subtypes by far. Of the person type signs, the preference is less clear, but favors signs denoting the personality of the bearer, followed by signs depicting a characteristic gesture/movement or a hobby. Among the name-derived signs, most subtypes are based on the initial(s) of the bearer's name. One noteworthy thing here is that the number of fingerspelled name signs is very low ($n = 4$). It is likely that this is to some extent a consequence of the terminology, since *persontecken* 'name sign' (lit. 'person sign') in Swedish is usually a term specifically to regular signs used with reference to a person, in opposition to fingerspelling that person's name. However, just as there are lexical signs that are always fingerspelled, some members of the SSL community (Deaf or not) only have a fingerspelled name sign. Possibly, such fingerspelled names are more common for shorter names as they fit the phonological template of signs better, just as shorter words are preferred as fingerspellings in general (Börstell et al 2016: 167–168), as has also been discussed for American Sign Language (Battison 1978; Supalla 1990: 101).

[15] For participants giving *other* as the type, they had the choice of writing the specific motivation. Most participants who answered this question did not provide a description detailed enough for these signs to be re-categorized into any of the existing types or clustered as a new type. Thus, the *other* responses were left without further analysis. This is also true for the subtypes *other*, used if the participant could attribute their name sign to a main type but were unable to match it to any of the listed subtypes – in these cases, participants did not describe their specific subtype further.

Table 2. Distribution of types and subtypes of name signs

Type	**No. of signs**	**Subtype**	**No. of signs**
Appearance	301 (40.8%)	hair	114
		physical	117
		clothes	47
		metaphor	3
		other	20
Person	236 (32.0%)	personality	85
		hobby	44
		profession	11
		gesture	43
		sign	9
		metaphor	10
		other	34
Name	117 (15.9%)	fixed	9
		whole name	10
		partial name	18
		homonym	6
		extension	2
		initialized	48
		fingerspelled	4
		other	20
Background	16 (2.1%)	country	8
		city	6
		inherited	2
Other	67 (9.1%)	–	67
TOTAL	**737 (100%)**	**TOTAL**	**737**

Since we have data on the identity of both name giver and bearer, it is interesting to see whether this correlates with the distribution of name sign types – that is, if the distribution differs based on the identity of name giver or bearer. Figure 9 shows the distribution of the main types of name signs across the identity groups of the bearer. We can see here that the distributions are very similar across the groups, and in fact a chi-squared test – intended to evaluate possible non-random skewedness in distributions across groups – shows no significant difference in the distribution across all groups ($\chi^2(12) = 11.046$, $p = 0.525$) or between the largest and most distinct groups, Deaf vs. hearing ($\chi^2(4) = 4.038$, $p = 0.4009$).

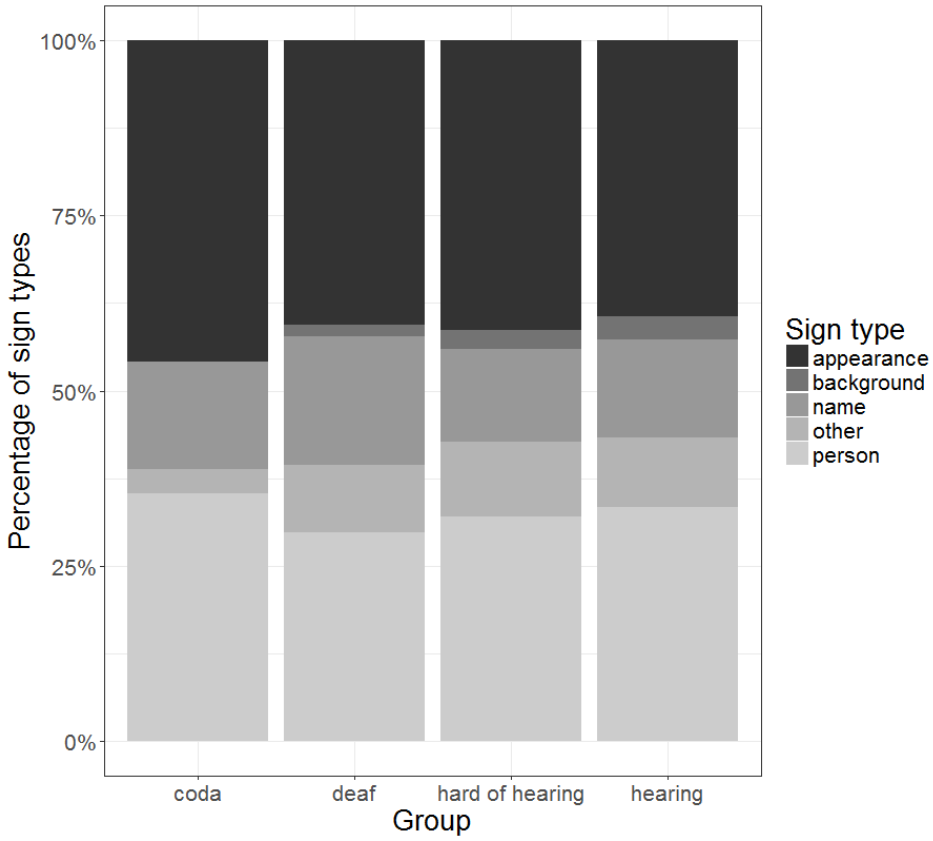

Figure 9. The distribution of sign types by identity group

Because there are no differences in the identity of the bearer, the following will not make a distinction between the groups, but rather investigate all name signs as a sample from the same community. However, when looking at the identity group of the name giver, we see a marginally significant difference between name signs given by Deaf vs. not Deaf[16] ($\chi^2(4) = 11.009$, $p = 0.02647$). The main differences in the distribution are that appearance-based signs are more dominant when Deaf name givers give the name, whereas name-related signs are relatively more common with non-Deaf name givers. An interesting thing to speculate about here is, of course, whether the primary language of the name giver (spoken vs. signed) is the reason for name-related signs being more or less common, that is, whether one is more focused on the spoken language community name and its form if that language community is one's primary community.

[16] Because of the low numbers for hard of hearing name givers, these were collapsed with the hearing name givers.

Table 3. Distribution of name sign types based on name giver identity

Name giver	Appearance	Background	Name	Other	Person
Deaf	219 (44%)	14 (3%)	69 (14%)	43 (9%)	157 (31%)
Not Deaf	82 (35%)	2 (1%)	48 (20%)	24 (10%)	79 (34%)

Turning to the gender identity of the name bearers, we see some striking differences. Comparing the female and male groups in Figure 10, we see a clear difference in the preference for either appearance or person sign types.[17] Among female bearers, the most common type of sign denotes the appearance (e.g., hair, physical features, clothes), whereas the most common type among males is person-based signs (e.g., behavior, mannerisms, hobbies). This distribution even exhibits a significant difference between genders (female vs. male; $\chi^2(4) = 25.269$, df = 4, $p < 0.0001$).

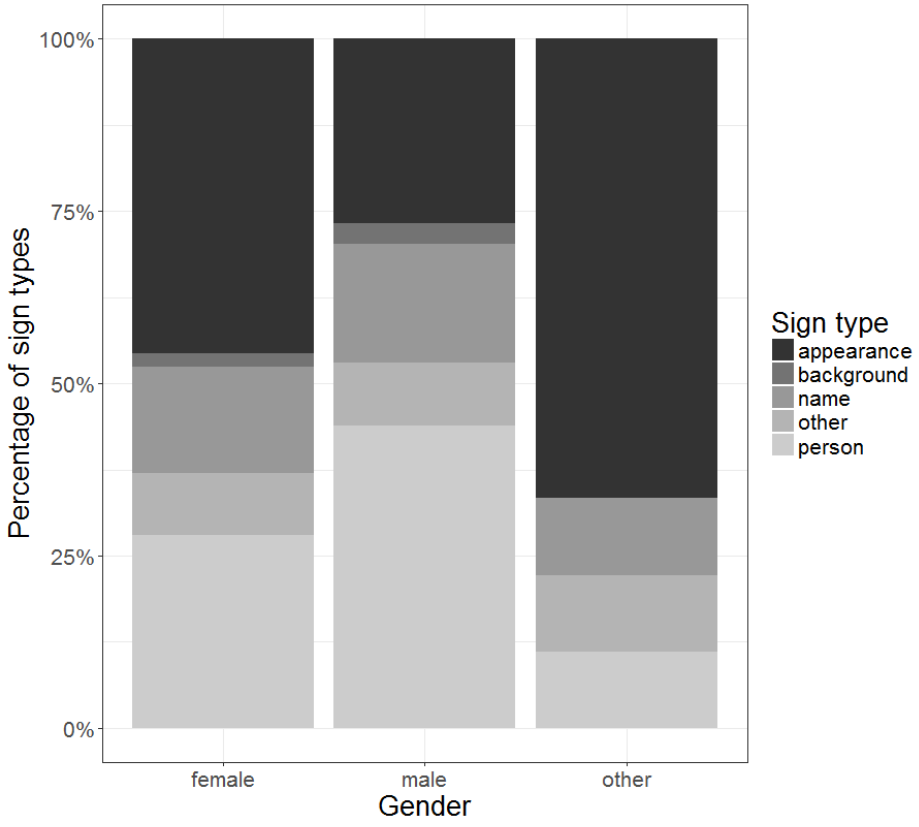

Figure 10. The distribution of sign types by gender

We see a difference in the distribution of signs according to the gender of the name bearer, but this only takes into account the whole sample without

[17] The number of participants identifying as neither female nor male are so few (n = 9) that this group cannot be used for statistical testing.

factoring in possible differences over time. As was shown in Figure 1, the sample consists of a diverse set of individuals in terms of their respective age. Thus, we want to see if the dominance of certain name sign types changes over time. In order to do this, we take the time of naming (i.e., the name bearer's birth year + age of naming) to find out if there are diachronic changes in the naming preferences within the language community.[18] The number of name signs per decade is not balanced, with most name signs in the data having been assigned in the decades 1980–2010, as seen in Table 4. This of course means that the confidence about the distribution of types is low for the decades 1950–1970.

Table 4. Distribution of name signs assigned per decade and gender

	1950	**1960**	**1970**	**1980**	**1990**	**2000**	**2010**
Female	6	11	37	138	144	93	82
Male	2	2	18	42	58	29	39
Total	8	13	55	180	202	122	121

In Figures 11 and 12, we see the distribution of name sign types grouped by the decade of naming (1950–2010) for the female and the male participants, respectively.

In Figure 11, showing the changes in the distribution of name sign types for female bearers over time, we see a very clear pattern of appearance type signs moving from dominant to constituting about a quarter of the sign types today. Parallel to this, the person-based sign types are increasing, moving from marginal to the most common type of sign today.

Looking at the same distribution over time for the male participants, we do not find as clear a pattern (see Figure 12). We find that appearance signs are becoming less common over time, though not as distinct as among the females, and also less clearly because of an increase in person-based sign types. However, since there are fewer males than females in the data, a possible pattern may be less visible because of variation in data size for the decades included (cf. Table 4).

[18] Out of the 737 name signs in the data, 712 had information about the birth year of the bearer. Thus, 25 signs in the data are excluded from the diachronic investigation concerning the time of naming.

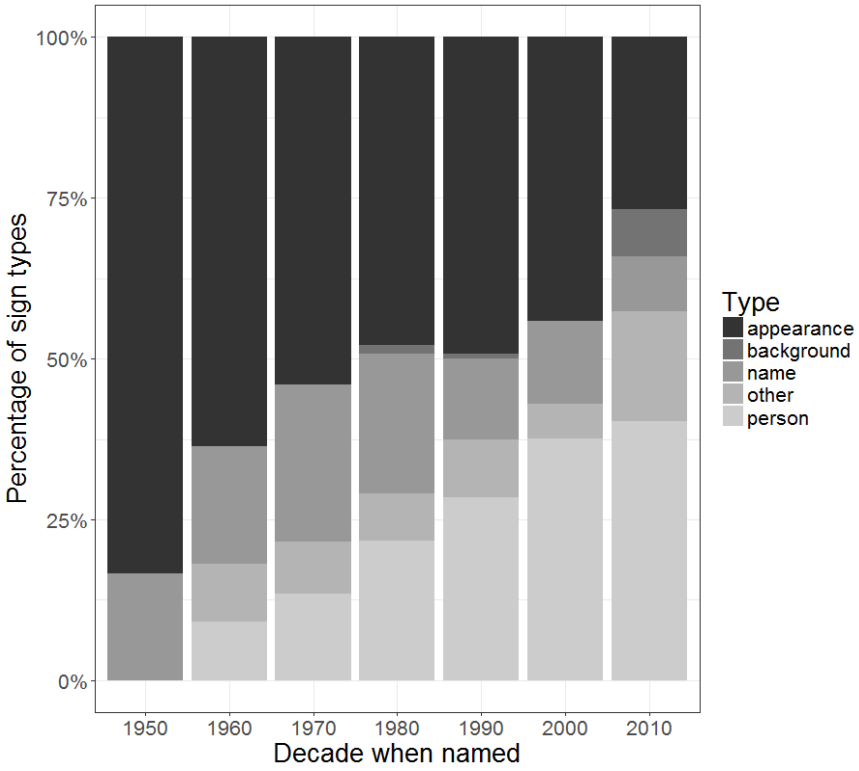

Figure 11. Distribution of sign types by decade of naming (female)

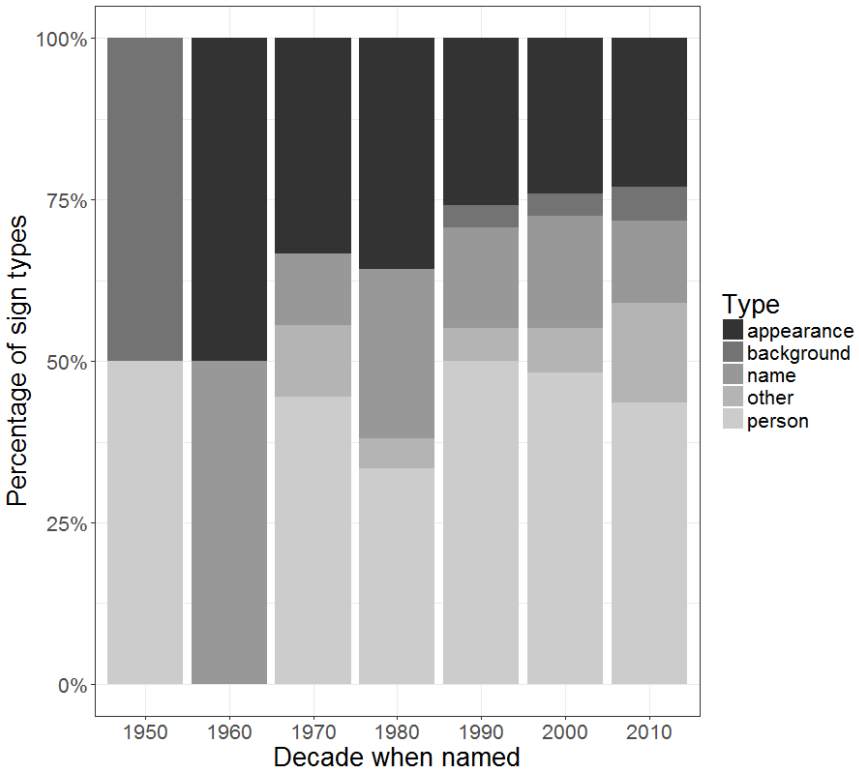

Figure 12. Distribution of sign types by decade of naming (male)

Overall, looking at possible changes over time relative to previous studies of name signs in the SSL community, we can divide the data into signs being assigned before or after Hedberg's (1989) investigation, that is, we divide the names into those assigned before 1990 (n = 262) vs. those assigned in 1990 or later (n = 450). Comparing these groups, we see the distribution across name sign types as given in Table 5. Here, we see the general pattern of the appearance-based signs losing their dominance over time and becoming balanced with the person-based signs. There is a significant difference between the distribution of types based on the year of naming (before 1990 vs. from 1990; $\chi^2(4) = 25.112$, $p < 0.0001$).

Table 5. Distribution of name sign types based on the year of naming (before vs. from 1990)

	Appearance	**Background**	**Name**	**Other**	**Person**
Before 1990	123 (47%)	3 (1%)	56 (21%)	19 (7%)	61 (23%)
From 1990	168 (37%)	12 (3%)	57 (13%)	44 (10%)	169 (38%)

Looking at the data, it is visible that hearing participants are overrepresented in the later decades, which could possible point to a skewing of the results. However, when extracting the distribution of name signs across the types for the Deaf and hearing groups, respectively, we still see that the Deaf group exhibits a significant difference between the types based on year of naming (before 1990 vs. from 1990), using here a Fisher's Exact Test due to the lower token counts for each cell ($p < 0.01$).[19] Table 6 shows the distribution of types for the before vs. after 1990 groupings for the Deaf participants, and Table 7 shows the same for the hearing participants.

Table 6. Distribution of name sign types based on the year of naming (before vs. from 1990) – Deaf only

	Appearance	**Background**	**Name**	**Other**	**Person**
Before 1990	79 (46%)	2 (1%)	38 (22%)	15 (9%)	36 (21%)
From 1990	42 (34%)	3 (2%)	17 (14%)	10 (8%)	51 (41%)

From Table 7, it is clear that the distribution of names across types is quite similar for the before vs. after 1990 groups among the hearing participants. However, there are very few tokens for the before 1990 group. Thus, again using a Fisher's Exact Test due to the low token counts, we find that there

[19] A Fisher's Exact Test also evaluates distributional skewedness, but is especially used for cases in which certain categories display low token counts. In such cases, a chi-squared test may be inaccurate, whereas a Fisher's Exact Test is less sensitive to fewer tokens. Thus, the Fisher's Exact Test is more reliable here.

is no significant difference between the before and after 1990 groups for the hearing participants (p = 0.7636). Note that this comparison should be analyzed with caution, due to the low total amount of tokens in the before 1990 group. It is nonetheless noteworthy that it is the Deaf group that shows a change over time (or, specifically before vs. after Hedberg's previous studies) in terms of naming customs, based on these data.

Table 7. Distribution of name sign types based on the year of naming (before vs. from 1990) – hearing only

	Appearance	**Background**	**Name**	**Other**	**Person**
Before 1990	15 (41%)	0 (0%)	7 (19%)	3 (8%)	12 (32%)
From 1990	85 (38%)	8 (4%)	29 (13%)	24 (11%)	78 (35%)

This basic categorization of naming before vs. after 1990 does not, of course, take into account the age of the bearer at the time of naming. One could speculate that this has an effect, that is, whether or not the name bearer is younger or older at the time of naming could be relevant. However, we see no significant difference between age of naming, using age 7 as the cut-off point, since it corresponds to the distinction of before or after school age (age of naming <7 vs. ≥7; $\chi^2(4) = 4.9031$, $p = 0.2974$).

5 Conclusions

This paper set out to investigate the customs and trends of name signs in the Swedish Sign Language (SSL) community. Because this topic had been studied previously (Hedberg 1989; 1994), the purpose was to give an updated description of the name signs in the community, and also look at various aspects that had not been resolved previously. For instance, the previous research on name signs in the SSL community had not taken into account the distribution of signs across different identity groups, the influence of name giver on the types of names assigned, or the possible changes in the preferences for certain name sign types over time.

In this study, the aim was not to document types of name signs found, but rather look at the distribution of name signs according to the types identified in Hedberg's work. This, of course, makes the current study quite different from Hedberg's, in that it allows participants classify their own name signs according to pre-defined categories, rather than identify categories and patterns based on detailed documentation. In general, we have seen that the categories proposed by Hedberg (1989; 1994; 2009) are still relevant today. The types of name signs described by Hedberg are

found in the community (with the exception of the no longer productive number signs), and the general distributional preferences indicated in passing seem to still be relevant. Thus, this study complements Hedberg's work by confirming that the identified types are still found in the community, but adds detailed information about the distribution of the types, and whether the identity of and relationship between the name bearer and giver has an effect on naming. The current study is also diachronic in two respects: first, it generally evaluates changes over time concerning naming in the SSL community; second, it provides a specific second point of reference (to which Hedberg's work constitutes the first point of reference) in collecting name sign data from the community, which in itself is used to evaluate diachronic change.

Furthermore, we see that several aspects of the naming customs are mirrored in previous findings from other sign language communities, in that, for instance, name signs are often assigned by peers rather than family (Mindess 1990; McNamara 2003 cited in Day & Sutton-Spence 2010), and that entering the school years is often associated with receiving a name sign among Deaf children (Yau & He 1989; Day & Sutton-Spence 2010; Esipova 2013; McKee 2016). These naming customs point to differences between identity groups within the community (Deaf, Coda, hard of hearing, and hearing) and that social interaction and connection to the group community affect the circumstances of naming. Although similar results have previously been found in other sign language communities, and have been implied for the SSL community, this paper provides the first quantitatively based description of the specifics concerning naming customs within the SSL community. However, as noted above, possible changes in the community in terms of language and education (e.g., the introduction of cochlear implants, CI) may alter the patterns found in this study, seeing as the data mostly cover the pre-CI community. This would be an issue for future research.

What has also been implied for the SSL community, and also found in several other sign language communities (see McKee 2016), is that descriptive signs are the most common in the SSL community, specifically signs depicting the appearance of the name bearer, for instance hair, physical features, or clothes. One noteworthy finding of this study is that there has been within the SSL community an ongoing change in the preference trends for name sign types over time. The clearest example of this is the increase of person-based signs (personality, behavior, mannerisms) at the expense of the appearance-based signs. This trend is

particularly clear among the Deaf female participants in this study, but overall there is a statistically significant difference in the distribution of name sign types assigned before and after Hedberg's documentation of name signs (concluded in 1989). What the driving force of this change is cannot easily be resolved, but one could theorize that it has to do with social factors in society at large, as well as the interaction between the Deaf and hearing communities. For instance, it is sometimes said that a distinct feature of Deaf culture is a direct and honest way of speaking, which may entail commenting on visually salient physical traits (or changes of such, e.g., gaining weight) in people, and that this sometimes causes clashes in social interaction between Deaf and hearing (cf. Hoza 2007). Perhaps the Deaf community is moving towards naming customs that avoid reference to physical features as a consequence of influence and conforming to general social pressures in society. However, it should be noted that an overwhelming majority of participants, across all identity groups, saw their name sign as very positive (median for all groups being 4 or 5). For hearing individuals, receiving a name sign is often a symbol of being included in the sign language community, at least to some extent (Mindess 1990: 13–14; Paales 2011).

In this study, we have seen statistics of the naming customs and distribution of name sign types in the SSL community. The findings show that several patterns corroborate the claims of previous studies based on the SSL community and other sign language communities. The changes over time suggest certain trends in the preferences of name sign assignment in the community, and it is thus important that future studies follow up on these results, in order to see how these trends develop in a longer time span, particularly considering current changes in the community.

Acknowledgments

I wish to extend my gratitude towards all the individuals who participated in this study by filling out the online questionnaire. I also want to thank Ingela Holmström, Johanna Mesch, Krister Schönström, and two anonymous reviewers for comments and suggestions on an earlier version of this paper, as well as Natalia Perkova, Bernhard Wälchli, and other participants at the HY–SU text workshop in Saltsjöbaden in March 2017 for general comments on the project and the study. I am also grateful to Ezra Alexander for proofreading the text.

References

Battison, Robbin. 1978. *Lexical borrowing in American Sign Language*. Silver Spring, MD: Linstok Press.

Börstell, Carl & Hörberg, Thomas & Östling, Robert. 2016. Distribution and duration of signs and parts of speech in Swedish Sign Language. *Sign Language & Linguistics* 19(2). 143–196. doi:10.1075/sll.19.2.01bor.

Day, Linda & Sutton-Spence, Rachel. 2010. British sign name customs. *Sign Language Studies* 11(1). 22–54. doi:10.1353/sls.2010.0005.

Deaf Studies. 1997. Deaf Studies. Centre for Deaf Studies, University of Bristol. (http://www.bris.ac.uk/Depts/DeafStudiesTeaching/dhcwww/contents.htm) (Accessed 2017-02-15.)

Delaporte, Yves. 1998. Des noms silencieux. Le système anthroponymique des sourds français. [Silent names. The anthroponymic system of the French Deaf]. *L'Homme*. 7–45.

Dubuisson, Colette & Desrosiers, Jules. 1994. Names in Quebec Sign Language and what they tell us about Quebec Deaf culture. In Ahlgren, Inger & Bergman, Brita & Brennan, Mary (eds.), *Perspectives on Sign Language Usage: Papers from the Fifth International Symposium on Sign Language Research, vol.* 2, 249–260. Durham: The International Sign Linguistics Association.

Esipova, Maria. 2013. Proiskhozhdeniye i funktsionirovaniye zhestovogo imeni v lingvokul'ture glukhikh Rossii (Происхождение и функционирование жестового имени в лингвокультуре глухих России) [Sign names in the Russian Deaf culture: Origin and functioning]. *Moscow University Young Researchers' Journal* 1. 1–5. (http://youngresearchersjournal.org/wpcontent/uploads/2013/02/esipova-sign-language.pdf)

Hedberg, Tomas. 1989. Persontecken: Deras härkomst, bildningssätt och användning [Name signs: Their origin, formation, and use]. *Forskning om teckenspråk* (*FoT*) XVI, Department of Linguistics, Stockholm University.

—— 1994. Name signs in Swedish Sign Language: Their formation and use. In Erting, Carol J. & Johnson, Robert C. & Smith, Dorothy L. & Snider, Bruce D. (eds.), *The Deaf Way: Perspectives from the International Conference on Deaf Culture*, 416–424. Washington, DC: Gallaudet University Press.

—— 2009. Persontecken i teckenspråket [Name signs in sign language]. *Studia anthroponymica Scandinavica* 27. 97–114.

Holmström, Ingela. February 17, 2017. (Personal communication.)

Holmström, Ingela & Schönström, Krister. 2017. Resources for deaf and hard-of-hearing students in mainstream schools in Sweden. A survey. *Deafness & Education International* 19(1). 1–11. doi:10.1080/14643154.2017.1292670.

Hoza, Jack. 2007. *It's not what you sign, it's how you sign it: Politeness in American Sign Language*. Washington, DC: Gallaudet University Press.

Kourbetis, Vassilis & Hoffmeister, Robert J. 2002. Name signs in Greek Sign Language. *American Annals of the Deaf* 147(3). 35–43.

McKee, Rachel Locker. 2016. Sign language: Name signs. In Gertz, Genie & Boudreault, Patrick (eds.), *The SAGE Deaf Studies Encyclopedia*, 805–808. Thousand Oaks, CA: SAGE. doi:10.4135/9781483346489.n256.

McKee, Rachel Locker & McKee, David. 2000. Name signs and identity in New Zealand sign language. In Metzger, Melanie (ed.), *Bilingualism and identity in Deaf communities*, 3–40. Washington, DC: Gallaudet University Press.

McNamara, Geraldine. 2003. What drives the allocation of name signs in British Sign Language? Centre for Deaf Studies, University of Bristol. (Undergraduate dissertation.)

Meadow, Kathryn P. 1977. Name signs as identity symbols in the Deaf community. *Sign Language Studies* 16 (Fall). 237–246. doi:10.1353/sls.1977.0015.

Mesch, Johanna. March 17, 2017. (Personal communication.)

Mindess, Anna. 1990. What name signs can tell us about Deaf culture. *Sign Language Studies* 6(16). 1–24. doi:10.1353/sls.1990.0009.

Nonaka, Angela. 2015. Toponyms in Ban Khor Sign Language. Learning Communities: *International Journal of Learning in Social Contexts* 16. 66–91.

Nonaka, Angela & Mesh, Kate & Keiko Sagara. 2015. Signed names in Japanese Sign Language: Linguistic and cultural analyses. *Sign Language Studies* 16(1). 57–85.

Nyst, Victoria & Baker, Anne. 2003. The phonology of name signs: A comparison between the Sign Languages of Uganda, Mali, Adamorobe and The Netherlands. In Baker, Anne & van den Bogaerde, Beppie & Crasborn, Onno (eds.), *Cross-linguistic perspectives in sign language research. Selected papers from TISLR 2000*, 71–80. Hamburg: Signum.

Nyström, Staffan. 2002. Place-names in the Swedish Sign Language of the Deaf. In Boullón Agrelo Ana Isabel (ed.), *Actas do XX Congreso Internacional de Ciencias Onomásticas. Santiago de Compostela, 20–25 de setembro de 1999*, 873–876. A Coruña: Biblioteca Filolóxica Galega. Instituto da Lingua Galega.

Paales, Liina. 2010. On the system of person-denoting signs in Estonian Sign Language: Estonian personal name signs. *Sign Language Studies* 10(3). 317–335. doi:10.1353/sls.0.0048.

—— 2011. Name signs for hearing people. *Folklore* 47. 43–76. doi:10.7592/FEJF2011.47.paales.

Rainò, Päivi. 2004. *Henkilöviittomien synty ja kehitys suomalaisessa viittomakieliyhteisössä* [The emergence and development of personal name signs among sign language users in Finland]. Helsinki: Kuurojen Liitto ry.

Schembri, Adam & Lucas, Ceil (eds.). 2015. *Sociolinguistics and Deaf communities*. Cambridge: Cambridge University Press.

Supalla, Samuel J. 1990. The arbitrary name sign system in American Sign Language. *Sign Language Studies* 67. 99–126. doi:10.1353/sls.1990.0006.

Svenskt teckenspråkslexikon. 2017. Swedish Sign Language Dictionary online. Stockholm: Sign Language Section, Department of Linguistics, Stockholm University. (http://teckensprakslexikon.su.se)

Wickham, Hadley. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. (http://ggplot2.org)

Yau, Shun-chiu & He, Jingxian. 1989. How Deaf children in a Chinese school get their name signs. *Sign Language Studies* 18(65). 305–322. doi:10.1353/sls.1989.0020.
Österberg, Oskar. 1916. *Teckenspråket*. Uppsala: P. Alfr. Persons förlag.

Contact information:

Carl Börstell
Department of Linguistics
Stockholm University
106 91 Stockholm
Sweden
e-mail: calle(at)ling(dot)su(dot)se

(Re)defining linguistic diversity: What is being protected in European language policy?

Dave Sayers
Sheffield Hallam University & Cardiff University
Petra Lea Láncos
German Research Institute for Public Administration

Abstract

Linguistic diversity is complicated. It involves two main elements: a headcount of "languages", plus variation and variability within and between them. In this article we show how language policy in Europe claims to protect diversity but falls short on these two measures. Our legal analysis examines the institutional politics of the European Union, details of accession, and institutionalisation of multilingualism. We describe the manifestation of a multilevel language hierarchy: working languages are topmost, then official languages, then non-official languages. This largely privileges national languages, principally English. Meanwhile allochthonous ('immigrant') languages are discounted, despite outnumbering autochthonous ('indigenous') languages. Our legal analysis therefore suggests an early stumble for linguistic diversity: even limited to a headcount of "languages", most are neglected. Next, our sociolinguistic analysis examines the Council of Europe's approach to protecting minority languages. We show how diversity can decline even among protected languages, using two case studies: Cornish, a young revival; and Welsh, an older, more established revival. The Cornish revival could only proceed after agreement on singular standardisation. Meanwhile the internal diversity of Welsh has declined significantly, fuelled by the normative reproduction of its standard form in education, and by sharpened social pressures against local dialects. Moreover, by comparing the EU and the Council of Europe, we aim for an overarching argument about "European language policy". We conclude that linguistic diversity is neglected, through exclusion of most of the languages spoken in Europe, and pressures on language-internal diversity within protected languages. Linguistic diversity is something richer and more complex than the limited goals of existing policies; it transcends language boundaries, and may be damaged by planned intervention.

Keywords: linguistic diversity, European Union, Council of Europe, indigenous and immigrant languages, Welsh, Cornish

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

1 Introduction

> For last year's words belong to last year's language
> And next year's words await another voice. – T. S. Eliot

There are lots of languages in the world. Languages also have different dialects. Languages and dialects both change over time. They are difficult to pin down. They are messy, and promiscuous. They sprawl, mix together, split apart, morph into new shapes. People with different languages or dialects at their disposal are forever combining pieces of them in conversation, creatively and dynamically. This is true of lesser used languages just as it is of more widely used languages. To understand this enigmatic worldwide complexity, in all its ceaseless, cacophonous, joyous tumult, is to understand linguistic diversity.

In this paper, we scrutinise policies which set out to "protect linguistic diversity", but which fundamentally misunderstand just what an enormous job that would really be. Using legal and sociolinguistic insights, we show how language policy in Europe creates structural inequalities between languages, and exacerbates inequalities within them, with the effect of driving down diversity in unseen ways.

Our legal analysis scrutinises the European Union: the legislative framework that guides its inner workings, and the policies it has published with regard to language learning across Europe. We examine the politics of the EU, details of accession, and the institutionalisation of multilingualism. We show how the workings of the EU create a multi-level language hierarchy: working languages are topmost, followed by official languages, while non-official languages have transitional recognition. This hierarchy largely privileges national languages, principally English. Meanwhile *allochthonous* ('immigrant') (FUEN n.d.: 14) languages are discounted, despite outnumbering *autochthonous* ('long-established, indigenous') (FUEN n.d.: 14) languages around four to one (cf. Sayers 2015 for a critique of the 'immigrant'/'indigenous' distinction). Our legal analysis demonstrates an early stumble for linguistic diversity: even constrained to a headcount of distinct "languages", most are neglected.

Next, our sociolinguistic analysis examines the Council of Europe, and its method of promoting minority languages. We show how this ultimately puts pressure on diversity within protected languages, exacerbating homogenisation of their spoken forms. We consider two case studies: Cornish, a somewhat nascent revival; and Welsh, an older and

more established revival. Although limited to the UK, nevertheless these two give a useful breadth of perspective given their very different stages of evolution. Holding up these case studies to our definition of linguistic diversity serves to extend our overarching critique. And, as a side note, since these case studies relate to the Council of Europe, not the EU, they will remain relevant regardless of future UK-EU relations.

As Blommaert & Verschueren (1998: 205) point out, "within minorities there are always minorities" (cf. Blommaert 2001). Woolard & Schieffelin (1994: 60–61) further argue that "movements to save minority languages ironically are often structured around the same received notions of language that have led to their oppression and/or suppression". We conclude that linguistic diversity in practice is putatively celebrated but palpably neglected in European language policy. This occurs through hierarchical privileging of official languages, exclusion of most of the languages actually spoken in Europe, and neglect of language-internal diversity within the relatively few languages under protection.

2 (Re)defining linguistic diversity

What is linguistic diversity anyway? Researchers of language policy and minority languages mention it frequently, but actually define it only vaguely. Nettle's 1999 volume *Linguistic Diversity* is a case in point, defining linguistic diversity as "the total number of languages" (Nettle 1999: 3). Nettle then discusses how languages borrow from each other and that their boundaries are uncertain; but this is part of a philological procedure to establish etymological relationships between languages, to trace their emergence as distinct, separable entities. Nettle (1999: 10) then lists three types of linguistic diversity: *language diversity* (total number of mutually unintelligible languages); *phylogenetic diversity* (different lineages of languages, i.e. number of branches on language trees); and *structural diversity* (range of permutations in linguistic structure, e.g. word order). Though these categories may be related, they are nevertheless posited as discrete. Indeed, his overarching aim is to explain "[t]he way in which the languages of the world have diverged" (Nettle 1999: 12). A co-authored follow-up to the volume (Nettle & Romaine 2000), which regularly mentions linguistic diversity, is still quite candid about this reductionism:

> [I]t is difficult to say precisely how many languages there are in the world. In addition to languages, there are also varieties or dialects of languages, many of which are also at risk. We confine ourselves here, however, to the topic of language endangerment. (Nettle & Romaine 2000: 27)

Despite this caveat, their frequent use of the term "linguistic diversity", without qualification, suggests that *all* diversity is under discussion.

But our critique here is not entirely new. Mobilising poststructuralist theory, Wright (2007a) distinguishes "language-as-practice" from "language-as-system". She flags up a "constant tension […] between the acceptance of the heterogeneity of practice and the necessity of fixing a set of forms that will remain invariant across all domains" (Wright 2007a: 221) – to the detriment of "creativity and evaluation of meaning" (Wright 2007a: 208):

> The trade-off seems clear. Where a language becomes a language of power of any kind (the language used in […] democratic institutions and in bureaucracies and the language spread through the state-run education system), the cost is acceptance of that language as system – a codified, stable written standard that may not entirely reflect the practice of those designated as its speakers. (Wright 2007b: 96)

We build on Wright's account, offering a sociolinguistically informed definition of linguistic diversity, and using this to scrutinise the claims of European language policy.

We begin with Marcellesi (2003; cited in Jaffe 2007: 71)[1] who makes a similar distinction to Wright, between *variation* and *variability* in language. The former is the total of the existing differences in all language – the differences we notice in language around us, at the present moment. For example, when we think of "different accents" that we remember hearing, that is variation. Variability, meanwhile, is harder to grasp. It is the capacity for language to change in new and unforeseen ways. Variability is the unknown and the unknowable. What differences will there be in language in five years, or ten years, or a hundred years? This is the enigma of variability.

Put another way, variation is three-dimensional while variability is four-dimensional. This in turn gives us a working definition of linguistic diversity: all the existing *synchronic* differences in language (at one point

[1] Marcellesi does not actually use the terms "variation" and "variability"; rather, Jaffe derives this distinction from his work (pers. comm.), hence our citing both of them here.

in time) plus all the ongoing *diachronic* changes (across time), including future changes which we cannot know.

"Languages" are contested categories, whose boundaries are often challenged in sociolinguistic research (e.g. Blommaert 2005; Dufva et al. 2011). Whatever their denotation, languages are superordinate to "dialects", so dialects are more useful in understanding diversity in all its fullness. But ultimately, languages and dialects are both just essentialised ontological crutches – flags in the sand, delimiting imaginary frontiers. Indeed, Erker (2017) suggests replacing "dialect" with "dialectal" for this reason. But in variationist sociolinguistics – the study of language variation and change – there is a long tradition of mindful cognitive dissonance conceding this ontological fragility but maintaining dialects as necessary heuristics. Dialects hold our hand as we grapple with the bewildering blur of our myriad language differences. Trudgill (1999: 7) is refreshingly candid here: "We realise that dialects form a continuum, but for the sake of clarity and brevity, we divide this continuum up into areas at points where it is least continuum-like". So, there is an understanding of fluidity within and across these imaginary boundaries. Variationism is also particularly well equipped to depict *variability* – the creative capacity for language to diverge in new and unpredictable ways. With all this in mind, tracking dialects can be a more useful gauge of linguistic diversity than just counting languages.

Still, despite being equipped to define linguistic diversity, variationist sociolinguists seldom actually use that term. "Because sociolinguists' treatment of language focuses on its heterogeneity, they seek a unit of analysis at a level of social aggregation at which it can be said that the heterogeneity is organised" (Eckert 2000: 30). Understanding this aggregative focus requires a brief scholarly history. A prime mover in the origin of sociolinguistics was an urge to counter prescriptive notions of "languages" as existing in some perfect form, and to argue that non-standard vernacular varieties are not aberrations or inchoate gibberish but structured and orderly forms, with their own rules of grammar and phonology (see e.g. Murray 2006: 2432–2433). Ironically then, as variationism developed as a field, it tended to reify some of the very ontological structures it became best equipped to challenge: "Over the past few decades, sociolinguistic research has concentrated on the *structured heterogeneity* inherent in all speech" (Smith et al 2007: 63, emphasis added).

Studies of variation and change have also said little about minority languages, the prime subject of language policy. Stanford & Preston (2003: 3) provide a rare collection in this vein – as they note, "such languages have received comparatively little attention". But their volume overall is explicitly limited to analysing variation within these languages from a technical linguistic point of view, not relating this to language policy. Kasstan (2017) highlights the same lacuna, and provides highly useful theoretical insights on how minority language contexts can inform variationist paradigms, though again without aiming to make explicit arguments about policy as such. With all this in mind, there remains scope for our current discussion.

Perhaps most useful to our task are the variationist concepts of *divergence* and *convergence*. The former describes dialects splitting apart into new varieties, usually because their speakers become physically separated – for example in diasporic migrations or coercive relocations. Convergence, meanwhile, arises when "two or more varieties become more alike", involving "the loss of geographically and demographically restricted, or 'marked', [linguistic] variants" (Torgersen & Kerswill 2004: 24). All this further disturbs the idea of "languages" as a particularly useful measure of diversity. There is so much diversity echoing around inside languages, variationism evidently offers a fuller understanding.

Putting all this together, linguistic diversity overall can be represented by all the dialects of all the languages in the world, plus the potential for dialects to change in new ways. The total number of languages does not encapsulate this; but nor does the total number of dialects. Ongoing change and new differences are essential too (cf. Mac Giolla Chríost 2007: 104). If dialects are diverging, diversity is going up. If dialects are converging, diversity is going down. This will be our benchmark. First though, we begin with the simpler measure of separate languages, to see if European language policy encourages diversity even on that limited basis.

3 Hierarchy and lingua franca in the European language regime

This section examines the policies of the EU, including the European Parliament and European Commission, and the influence of these policies on linguistic diversity in Europe.

3.1 The EU as a multilingual political community

European law has a significant impact on the rights and obligations of all EU citizens (Toggenburg 2005: §2.1). The freedom of EU citizens to access European law, and to contact EU institutions in their native language, are therefore decisively important for EU decision-making as well as legal certainty (Marí & Strubell 2002). It is self-evidently important that citizens identify with the law enacted on their behalf. Meanwhile, legal certainty guarantees that the subjects of supranational power have veritable access to, and understand, the law binding upon them.

Article 2 of the Treaty on European Union (TEU) stipulates respect for fundamental rights – including the rights of persons belonging to minorities. Article 3 extends to protection and promotion of the cultural and linguistic diversity of the EU. Member States are also signatories to documents such as the European Convention on Human Rights (Henrard 2004) as well as the International Convention on Civil and Political Rights, which affect EU language policy and legislation.

3.2 The development and goals of EU language policy: The Action Plan and the Framework Strategy

While Member States are not officially required to have a language policy as such, nevertheless they are effectively compelled to determine one, so as to efficiently manage European political, social and economic processes (van Els 2003: 45). Multilingualism projects gained impetus around the turn of the millennium. As a result of the resolution of the European Parliament (2003) regarding regional and lesser-used languages, the Commission issued an Action Plan (EC 2003; see also Nic Shuibhne 2008: 127) entitled *Promoting Language Learning and Linguistic Diversity (2004–2006)*.

Although the Action Plan asserts that "linguistic diversity is one of the European Union's defining features", and that "[r]espect for the diversity of the Union's languages is a founding principle of the European Union" (EC 2003: 12), its details are notably circumscribed. It focuses on formal language learning (EC 2003: 7–9) with a view to acquiring "the skills to communicate with one another effectively and to understand one another better" (EC 2003: 3). Furthermore, it declares that "regional and minority language communities do not seek support for the teaching of their languages as foreign languages" (EC 2003: 12). Here then we have the first

clear constraint: promoting the learning of other languages but restricted to official languages of Member States. Furthermore, education in regional and minority languages is only supported for the speakers of such languages, regardless of whether those languages are in decline and the pool of speakers is shrinking.

In 2005, the Commission issued its first communication dealing expressly with languages and multilingualism as a policy (EC 2005a). The *Framework Strategy for Multilingualism* states three goals: i) promoting language learning among EU citizens to contribute to maintaining linguistic diversity; ii) promoting a competitive, multilingual economy; iii) securing access to EU legislation and information for EU citizens in their native languages. EU language policy therefore encompasses protection and promotion of cultural identity, competitiveness and the respect for fundamental rights. The Strategy recalls the 2002 Barcelona goals of the European Council urging every EU citizen to learn at least two foreign languages (EC 2002: 19), financed through EU funds (EC 2005b).

It is important to recognise that the protection of linguistic diversity and the promotion of multilingualism are only compatible at first sight. Multilingualism policies, underlain by probability-sensitive language learning, tacitly favour "popular" or "big" European languages. This increases the imminence of language loss overall (van Parijs 2008: 21). In particular, while EU-sponsored official languages gain status and significance, regional and minority languages become less appealing and speakers may be indirectly induced to language shift.

In assessing the Framework Strategy, we may conclude that although linguistic diversity is putatively prioritised, similarly to the case of the Action Plan described above, the Commission sees language learning as the main workhorse of this endeavour. But this strategy, to protect linguistic diversity by promoting language learning, is based on a slippery premise, and may ultimately contribute to reducing linguistic diversity.

Naturally, language choice is not a zero-sum game. The nascent literature on *translanguaging* (e.g. García & Lin 2017) urges focus on the way multilingual speakers blur languages together dynamically in normal interaction, and advises that education should be modelled on this. But our point is more fundamental: that European language learning was putatively premised on exposing citizens to languages they otherwise would not encounter – yet the actual scope for that exposure is highly constrained.

All this chimes with Kraus (2008: 10) who concludes that, although the EU seems to glorify diversity on an abstract level, the concrete

measures of the institutions are vague and indeterminate. Similarly, analysing the term *linguistic diversity* in European political discourse, Strubell (2007: 159) notes:

> [I]t would seem reasonable to argue that 'safeguard' and 'preserve' refer to the maintenance of an existing state of affairs that may be under threat […]. Clear support for this view can be gleaned from the many Calls for proposals published up until 2000 by the European Commission to provide measures to *promote and safeguard regional or minority languages*. The object is much less abstract, and therefore much clearer: it is not 'diversity' […] being addressed, but rather minority languages and cultures. (Strubell 2007: 159, emphasis in original)

3.3 Regulation 1/58/EEC on language use in the institutions

The legislative act governing language use in EU institutions effectively reinforces secondary status for regional and minority languages. The central role of the language rules of European integration is highlighted by their prominent place in European Economic Community legislation, as early as the EEC Council's "Regulation No 1 determining the languages to be used by the European Economic Community" (EEC Council 1958). That Regulation has been amended several times during accession of new Member States, but has remained the basis of the language regime of the institutions for sixty years.

The preamble of the Regulation reads: "Whereas each of the four languages in which the Treaty is drafted is recognised as an official language in one or more of the Member States of the Community […]". Although this paragraph may only serve as a tool of interpretation, it went on to decisively influence the preambles of further legislative acts of the Community, thereby determining the scope of languages eligible for funding (Marí & Strubell 2002: 4; cf. Ó Riagáin 2002: 3–4). Furthermore, the preamble is even deemed the basis for affording official status to a Member State language at the EU level (Marí & Strubell 2002: 3). Marí & Strubell note that the above-mentioned passage has been interpreted unduly restrictively, since it does not foresee that only those languages of the Member States which are official in the *entire* territory of a Member State may be afforded official status. Milian-Massana (2008: 96) points out that the very wording of the passage implies that regional official languages may also be eligible, by the wording "an official language in", not "an official language of" (emphasis in original).

The restrictive interpretation of the preamble is underlined by the new Article 55 paragraph 2 TEU which came into force following the Lisbon amendment of 2007: accordingly, the "treaty may also be translated into any other languages as determined by Member States among those which […] enjoy official status in all or part of their territory." As such, the restrictive conditions (cf. Marí & Strubell 2002: 5) for affording official status to Member State languages become the rule employed to constrain the scope of languages eligible for this special status (cf. Király 2007: 36).

On their accession, Member States determine in their Act of Accession the languages they wish to use as official languages in the EU (Láncos 2009: 123; cf. Fidrmuc 2011). This rule does not expressly prohibit the Member State determining more than one official language – moreover, the status of a Member State language may be modified even after accession. According to Article 8 of the Regulation: "If a Member State has more than one official language, the language to be used shall, at the request of such State, be governed by the general rules of its law." So, the Regulation expressly foresees the possibility of elevating regional and minority languages to EU official languages. But the list of existing EU official languages demonstrates that Member States have a restrictive interpretation of the Regulation, and do not add further official languages other than those dominant in their respective territories. The secondary status of minority languages is therefore reproduced and reinforced at the EU level. Overall then, as minority languages are shunted into secondary status at both Member State level and in the EU, the practice of EU institutions leads to a multi-level language regime (Láncos 2012: 100).

3.4 The multilevel language regime

Based on Regulation 1/58/EEC on language use, and on the jurisprudence of the European Court of Justice, as well as the practice of the institutions, bodies and agencies of the EU, we see the emergence of a hierarchy of the languages spoken across the EU (cf. Marí & Strubell 2002: 4). As a result, different languages may be used in different spheres of official communication within the EU, while EU funding varies for the protection and promotion of languages. Some remedial work here is done by Articles 21–22 of the Charter of Fundamental Rights ("Non-discrimination" and "Cultural, religious and linguistic diversity"); but even within the status of official languages there is differentiation between working languages and other official languages. Due to legal, logistical or technical reasons (cf.

Marí & Strubell 2002: 11), the institutions, bodies and agencies of the EU give preference to certain languages (deemed working languages) over other official languages (Lanstyák 2004: 47), while regional and minority languages are largely neglected.

3.4.1 The hierarchy of the official languages

Article 55 paragraph 1 TEU begins: "This Treaty, drawn up in a single original in the Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovenian, Spanish and Swedish languages, the texts in each of these languages being equally authentic […]". Neither the Treaty nor the Regulation distinguishes Treaty languages from official languages. Certain scholars have concluded that the "principle of the equality of languages" forms part of the European constitutional order (de Witte 2004: 221). This seemed substantiated by the jurisprudence of the European Court of Justice, according to which, when interpreting European law, all authentic language versions of the Treaties must be taken into account (Mayer 2006). However, in the *Kik* judgement (Case C-361/01 P *Kik* v *OHIM* [2003] ECR I-8283), the Court expressly denied the existence of a principle of equality of languages, confirming the use of working languages in EU institutions (Nic Shuibne 2004). Namely, according to Article 6 of the Regulation on language use in the institutions, the institutions may determine *working languages*. This reinforces the assumption that differentiating between the official languages is legitimate. As such, Article 6 of the Regulation may serve as the basis for restricting the scope of official languages for institutions in internal and inter-institutional communication (Arzoz 2008: 178; de Witte 2008: 179). In practice, the Commission instituted English, French and German as working languages. In a recent judgement, however, the General Court found that the recruitment practice of the European Personnel Selection Office – which foresaw the mandatory command of either English, French or German – amounted to a discrimination based on language (T-124/13 and T-191/13 joined cases). This is because, in fact, no institution had made use of the opportunity contained in Article 6 of determining working languages in their rules of procedure; as such, the EU has no de jure working languages that would justify the existing hierarchy between the official languages of the EU.

3.4.2 Non-official languages spoken in the EU

The official languages of the EU are all official and majority languages of Member States, with the exception of Irish (co-official and minority). In general, therefore, it seems as though minority languages of the EU are non-official; but look more closely and there are nuances. According to Felföldi (2011: 3), minority languages may be classified as follows:

i) officially recognized language which is not an official language of the EU (e.g. Letzeburgesch);
ii) minority language spoken in only one Member State or a region thereof (e.g. Sorbian in Germany);
iii) minority languages spoken in various Member States (e.g. Catalan);
iv) minority languages with a kin-state (e.g. Hungarian); and
v) deterritorialized languages (e.g. Romani, Yiddish).

As regards the fourth of these categories, although for example Hungarian is a minority language in certain Member States, it is also an official language of the EU and the majority language in Hungary.[2] Other minority languages have a kin-state, such as Turkish and Russian; however, Turkey and Russia are not Member States. We may conclude that the non-official languages of the EU constitute a complex, nebulous category, which comprises autochtonous ('long-established, indigenous') minority languages, regional and deterritorialised languages, and allochthonous ('immigrant') languages (FUEN n.d.: 14). The EU has a varied approach towards these: some may achieve a sort of semi-official status; others may be eligible for funding; while the rest are neglected, most especially allochthonous languages (see Skutnabb-Kangas 2002: 10).

3.4.3 Privileged non-official languages

The first category of non-official languages is "privileged non-official languages", comprising all non-official languages of the EU which are co-official or regional official languages in a Member State (Lanstyák 2004).[3] Following the Lisbon amendment, Article 55 paragraph 2 TEU, the "treaty may also be translated into any other languages as determined by Member

[2] See Opinion of AG Maduro in *Spain v Eurojust* (16.12.2004), paras 48–49.
[3] This situation was changed by the new guarantees introduced under Articles 21–22 of the Charter of Fundamental Rights.

States among those which […] enjoy official status in all or part of their territory." Based on the Administrative Agreement concluded between the Spanish government and the EU institutions (Marí & Strubell 2002: 11), certain acts adopted in ordinary legislative procedure shall be translated to Catalan, Galician, and Basque. Speeches may be held in these languages in certain institutions; and to facilitate communication between speakers of these languages and EU institutions, Spain shall appoint intermediary bodies. Paragraph 11 of the Agreement states that all costs incurred as a result of the "semi-official status" of these languages shall be borne by Spain. Meanwhile Paragraph 1(c) exempts the Council from liability for the precision of translations.

Importantly, only Member States may afford "semi-official status"; that is, the EU-level recognition is dependent on Member State action, and entails Member State expense. Meanwhile the eligible language groups are also restricted to autochthonous languages. Allochthonous languages are excluded altogether. As regards the concrete administrative agreement already in force, this is only applicable to Spanish citizens, not for example French citizens with Basque as their mother tongue (Lanstyák 2004: 47). Translations made in these languages are not deemed authentic; the EU has no relationship to these languages, and all related costs are borne by Spain (de Witte 2004: 221).

Milian-Massana (2008: 203) points out that this form of recognition may therefore not be deemed a novel institutional status (cf. Mayer 2006: 372); however, it may enable future delimitation of such autochthonous languages from other non-official languages and institutional recognition of the same on the EU level (Milian-Massana 2008: 219). There is also increasing pressure on the EU to recognise such languages at the EU level, since they are spoken by more than 10 percent of EU citizens (Bradean-Ebinger 2011: 4).

3.4.4 Other, lesser used languages

The category of other, lesser used languages includes those not deemed official by either the EU or Members States. The protection afforded to these languages varies greatly. There are three categories of such languages: protected by law; not recognized by law; and prohibited (Lanstyák 2004). The first of these are afforded some level of protection, potentially supported institutionally or through education (de Witte 2008: 179; Lanstyák 2004: 51). Languages not recognized by law have no status,

but are at least passively tolerated. Prohibited languages are sanctioned in order to discriminate against or assimilate their speakers (Felföldi 2011: 3). Bradean-Ebinger (2011: 4) cites the Vlach and Macedonian language communities in Greece.

3.4.5 Allochthonous "immigrant" languages

In the second half of the 20th century, the Member States of the European Community became progressively more focal destinations for in-migration (Bradean-Ebinger 2011: 4). As noted above, immigrants whose languages are not considered *autochthonous* European languages are disadvantaged; their languages are not recognized either at EU or Member State level (Fidrmuc et al 2006: 9; Skutnabb-Kangas 2002: 10). This is no small matter; on 1 January 2015 (the most recent data available) the proportion of non-EU immigrants in Member States averaged 7.5 percent – in five states it exceeded 11 percent (Eurostat 2015). And the meaning of "indigenous" is highly contestable; languages such as Hindi and Arabic have been widely spoken in Europe for several generations, yet their "immigrant", "allochthonous" status appears to have been elliptically granted in perpetuity (see Sayers 2015). Extra & Verhoeven (1993: 10–11) note that while most Western European states supported immigrant language use in the media and education in the 1980s, by the 1990s many Member States introduced assimilationist policies, believing this to be in immigrants' best interests.

The emerging immigration policy of the EU seems to fall in line with this trend. The Third EU Ministerial Conference on Integration in 2008 led to the so-called Vichy Declaration (Carrera 2014: 174), which urged comprehensive integration strategies including language programmes and courses on the history, institutions and values of the EU (EC 2008). This means, firstly, that the cultural and linguistic heritage of immigrants are neglected in the interests of integration; and secondly, that immigrants who become EU citizens but speak a non-official language are disadvantaged relative to other EU citizens in Union level political participation. Finally, allochthonous languages are only supported from EU funds to aid the competitiveness of the European market – that is, widely spoken non-European languages whose kin-states are important commercial partners (Milian-Massana 2008: 218–219).

3.5 A language regime based on restricted multilingualism

The language regime of the EU engenders a hierarchy of languages similar to an inverted pyramid, topped by the (non-stipulated) "working languages" used in internal communication of the institutions. Below these few select languages are the other official languages of the EU used by institutions in communication with EU citizens and Member States, and into which certain documents are also translated.

While privileged non-official languages may acquire a semi-official status at the EU level, this status is highly contingent on political and fiscal fair winds in the Member State(s) in which they are spoken. Privileged languages and lesser used languages are potentially eligible for EU funding and are protected through the horizontal principle enshrined in Article 3 paragraph 3 TEU. Lesser used languages and allochthonous languages are excluded from external communication of the EU, and funding for them is scarce.

Since only a fraction of the languages spoken in the territory of the EU is represented in the internal and external communication of the institutions, bodies and agencies of the EU, we may conclude that the multi-level regime of the EU is based on "restricted multilingualism" (Bradean-Ebinger 2011: 4; see also Derlén 2011: 156–157 on "limited multilingualism").

From this institutional legal critique, we now move on to a sociolinguistic investigation of the European approach to protecting and promoting minority languages, and how linguistic diversity fares here.

4 Declining linguistic diversity within protected languages

The previous section set out how the institutional machinery of the European Union serves to limit the recognition and learning of minority languages, therefore placing unrecognised constraints on linguistic diversity. This was principally a legal analysis, taking languages as separate entities and considering how these are – or are not – effectively and equally facilitated. The current section moves down into diversity within minority languages being protected in Europe, using sociolinguistic insights. This moves us away from the European Union and towards the Council of Europe, proprietor of the European Charter for the Protection of Regional or Minority Languages. By relating the EU and the Council of

Europe in this way, we offer a picture of an overarching "European language policy".

In a wide-ranging critique, Perley (2012) takes aim at linguists who record and analyse dying languages without also attempting to revitalise their use. He lambasts the "disembodiment of language from speakers" (Perley 2012: 134), "a ghoulish process where linguists go out to find the last speakers of dying languages and record their last words. That is not saving the language. It is mortuary linguistics" (Perley 2012: 140). Perley endorses community-led efforts to teach and thereby revive declining languages. But if we heed Perley's counsel and revive minority languages, does that also support linguistic diversity? The answer feels obvious; but that feeling should always send a question mark shivering up the scholarly spine. Let's take a closer look.

Published in 1992 under the auspices of the Council of Europe, and eight years in the making, the European Charter sets out for both revival and diversity. It goes beyond passive tolerance of prior international law, as "the only international legal instrument whose primary aim is the protection and promotion of regional or minority languages" (Grin 2003: 67). It posits "linguistic diversity" as an explicit priority:

> Linguistic diversity is one of the most precious elements of the European cultural heritage. The cultural identity of Europe cannot be constructed on the basis of linguistic standardisation. On the contrary, the protection and strengthening of its traditional regional and minority languages represents a contribution to the building of Europe, which […] can be founded only on pluralist principles. (CoE 1992a: §26)

Recalling Perley's critique above, the Charter does at least require more than just documentation. But Perley may take issue with another aspect of the Charter. In prioritising languages, "the Charter does not establish any individual or collective rights for the speakers of regional or minority languages. In this, the Charter is in some ways a step backward from the Framework Convention [for the Protection of National Minorities]" (Dunbar 2000: 49; cf. CoE 1992a: §11). So Perley's prescription is problematic. Attempts to revive languages can still explicitly involve their "disembodiment" from their speakers. The Charter priorities languages themselves, not people. Of most relevance to our discussion is the plan to identify and protect particular languages, on the premise that this encourages linguistic diversity.

In our introduction, we quoted Wright's (2007a) distinction between "language-as-practice" and "language-as-system". The Charter seems geared towards the latter, given its focus on languages as discrete, defensible entities. How, for example, could "the number and percentage of oral interactions […] between civil servants and the public […] in the regional or minority language" (Grin 2003: 108) be recorded? Such checks, based on categorising interactions according to language, make sense for taking the pulse of languages as discrete systems, but less straightforwardly for intra-linguistic variation and variability.

Moreover, the Charter is worded in a purposively and diplomatically versatile way, to attend to "the specific conditions and historical traditions in the different regions of the European States" (CoE 1992b: Preamble). This acknowledges differences between linguistic minorities (Grin 2003: 76), but not within them. The actual requirements of the Charter are all binary, to provide services in "the regional or minority language" and "the dominant language" (Art.VI.15.1). There is an underlying presumption that these languages can be readily applied in a measurable manner, to achieve quotas of use.

Having said that, nowhere in the Charter is there any explicit call for standardisation. The Charter simply notes that there are these "languages", and that they should be protected. Their existence is presupposed, in the technical sense of a non-cancellable proposition (Levinson 1983: 207). Decisions over what constitutes "the regional or minority language" are left to the unspecified "authorities" in Member States. As we will see, it is in the subsequent planning process – downstream from the initial policy – that pressure upon diversity materialises.

Part III of the Charter contains six Articles "to promote the use of regional or minority languages in public life": "Education", "Judicial authorities", "Administrative authorities and public services", "Media", "Cultural activities and facilities", "Economic and social life" and "Transfrontier exchanges". Of these, the judicial, administrative, economic and transfrontier requirements are mostly reactive, limited to providing translations on request. The media provisions are hedged to apply only where "the public authorities […] play a role […], and respecting the […] autonomy of the media" (CoE 1992b: XI.1). The cultural provisions meanwhile are somewhat highbrow, "especially libraries, video libraries, cultural centres, museums, archives, academies, theatres and cinemas, as well as literary work and film production, vernacular forms of cultural expression, festivals and the culture industries" (CoE 1992b: XII.1). But

these are all rather minor cogs in the machine. The main muscle of the Charter is the first Article of Part III, Education, specifically to make primary and secondary education available in regional or minority languages.

To reprise the distinction between a headcount of "languages" and overall diversity, recall Strubell's (2007: 159) remarks quoted earlier about "the maintenance of an existing state of affairs that may be under threat". One can read different rationales into support that is constrained to languages as discrete countable systems. Jaffe takes a wry stance, hinting at a staunchly rationalist, perhaps neoliberal underpinning (cf. Petrovic 2005):

> [G]iven the long tentacles of the dominant ideologies of language and identity, the celebration of multiplicity, hybridity and ambivalence is not a powerful discursive position. You do not get money, or books, or official recognition by claiming ambiguous relationships with several identities, and shifting and contingent forms of identification with multiple linguistic codes. (Jaffe 2004: 278)

We offer a more mundane explanation, focusing on the spread of "New Public Management" as a form of public governance across the world from the 1980s into the twenty-first century (Broadbent & Laughlin 2002: 102; Schedler & Proeller 2002: 163). NPM endorses interventionist state action like minority language promotion, but requires strictly quantifiable measures of performance: a government framework designed to substantially alter social behaviour, but with close attention to productivity.

But both these explanations – neoliberal conniving or bureaucratic box-ticking – favour an understanding of languages as distinct, countable entities, aligned with groupings of citizens within the purview of the governing body, in this case the Council of Europe. The above review is a brief window into how "diversity is rhetorically turned into a problem that needs to be 'managed'" (Muehlmann 2007: 16).

Fulfilment of Charter commitments is monitored by an appointed Committee of Experts (CoE n.d.); but it is the States themselves that actually create and execute the necessary language policies. The question for us is how the prioritisation of diversity filters down into modern language revivals, and whether Charter-based measures can really encourage such a thing. This is the uniting theme of our two case studies below. For Welsh, our focus is the current sociolinguistic profile of the language, and how its diversity is faring in the context of this relatively mature revival. For Cornish, a more germinal revival, we look at recent efforts to promote the language, and emergent pressures to agree on a

single standard form for use in education – the principal field of activity, per the Charter.

For Welsh, we are reviewing dialectological data – recalling our earlier discussion about the value of dialects for gauging diversity. Cornish, however, died out in the eighteenth century and has been manually reconstructed for its revival (Sayers & Renkó-Michelsén 2015) so it has no "dialects" in any conventional sociolinguistic sense. For Cornish we therefore focus on policy data, asking whether the conditions are being established for diversity to grow in future.

4.1 Welsh

Welsh-medium education began as a private endeavour in the 1930s, only receiving state-funding in 1951 (May 2000: 125), considerably ramped up after the Education Reform Act 1988 (Dunbar 2000: 57). The Census of 1991 showed self-reporting of Welsh at an all-time low of 18.7%, with significant geographical variation. The 2001 and 2011 Censuses suggested that this decline had been at least stemmed (Higgs et al. 2004; Statistics for Wales 2012), a change routinely attributed to Welsh-medium education (Aitchison & Carter 2000: 141; Farrell et al. 1997; ONS 2004; Williams 2008: 254). The 2011 Census suggested either a slight drop or a plateauing from 1991, depending on the significance of a change to the relevant Census question (ONS 2012: §10), but this challenging result has not spurred a radical rethink of policy. In the first post-2011 Welsh Government language policy document, *Cymraeg 2050* (Welsh Government 2017), education remains the bulwark. And education is increasingly the primary point of exposure to Welsh for a largely non-Welsh-speaking population. According to the Welsh Language Commissioner: "Four out of five 5–15 year olds now mainly learn to speak Welsh at school" (Huws 2016a; for further data see Huws 2016b).

In 2001, the UK Government ratified the European Charter in respect of Welsh (McLeod 2008). They also decided that "the existing range of measures in place to support Welsh meant that the requirements of the Charter were already more than being met in Wales" (Dunbar 2000: 65). This is important. If the requirements of the Charter were already being met, then we can extend our remarks about Charter provision further back in time. To that end, we review Welsh variationist sociolinguistic data from the 1980s, as well as more recent evidence.

The success of the Welsh language revival is normally defined using Census figures and other surveys: "Demography – the numbers and distribution of people reporting themselves to have ability in Welsh, based on census data – is the usual focus of debate on the current 'health' of the language" (Coupland et al. 2005: 2). If education is the main reason for stemming the decline of Welsh use, then the *kind* of Welsh being used is more likely to be influenced by education. This is thrown into sharper relief by "the continuing shrinkage of the "heartland" zones for intergenerational Welsh language transmission" (Coupland et al. 2006: 353), further foregrounding education as the main life support for the language.

The significance for us of the Welsh revival is partly its relative maturity, partly its palpable influence on other language revivals around Europe, "a rare and celebrated exception to […] minority languages suffering language shift and decline" (Coupland 2011: 79–80) – "regarded with envy" (Huws 2006: 147; cf. May 2003: 218; Sallabank 2005: 59; McLeod 2008). Indeed, there is evidence of influence well beyond Europe, for example a task force from Inuit Tapiriit Kanatami in Canada making a five-day tour of Wales in December 2016 for guidance on standardising their writing system (FEL Canada 2017: 9). Analysing the Welsh case therefore allows a degree of generalisation to European language policy, and to an extent further afield.

We review below two variationist reports of spoken Welsh. And, as discussed earlier, we are looking at dialects as heuristics, indicators of diversity, not diversity itself.

We begin with Thomas (1987: 99), who explains that the "spoken standard" for Welsh education is in fact a relatively modern phenomenon, "the result of language-planning policy during the 1960s and 1970s". "The model thus devised is […] a dialectal hybrid […]. It is a purely prescriptive model which relates to no reality outside the classroom" (Thomas 1987: 104). Thomas aims to gauge the influence of this supralocal standard on local dialects. He brings variationist insights to the task, analysing change in spoken Welsh across age cohorts. He compares those who learnt Welsh at home and at school – respectively "primary" and "secondary" bilinguals. He conducts his fieldwork in "Aberdaron [extreme Northwest Wales], in an area of high-density Welsh incidence, and Merthyr [South Wales], in one of low-density incidence" (Thomas 1987: 108).

Thomas (1987: 110) analyses three variables indicative of local dialect resilience: the initial consonant mutation system; the pronunciation of final-syllable orthographic diphthongs; and occurrence of the possessive

pronoun. Consonant mutation, for example, is a typical irregular feature, requiring "a great deal of unstructured knowledge […] which cannot be quickly assimilated" (Thomas 1987: 110). Learning the nuances of this feature takes long-term exposure in spontaneous conversation, above and beyond the structured acquisition of classroom skills. On the basis that younger speakers tend to be exposed to more standard Welsh, Thomas compares three age groups in both locations: 5–19, 20–49, and 50+. His quantitative data for both locations combined are summarised in Figure 1.

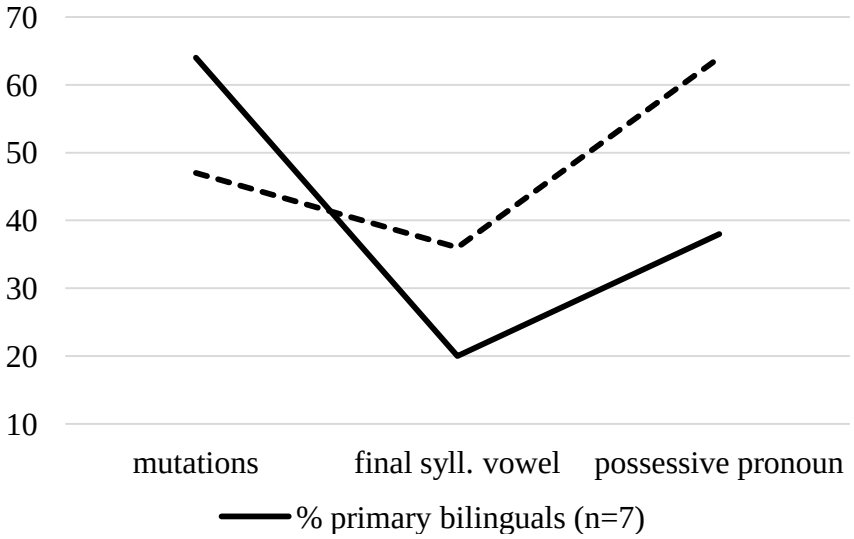

Figure 1. Adapted summary of data from A. R. Thomas (1987: 110)

Thomas (1987: 110) concludes that primary bilinguals showed "a conflict between dialectal and standard usage", while for secondary bilinguals "a major determinant of usage is the […] perceived standard". Primary bilinguals showed greater evidence of dialectal features, secondary bilinguals were affected by scholastic acquisition of Welsh, yet both were apparently influenced increasingly by the standard. These findings represent disparate pressures on diversity within Welsh: some pre-existing, some apparently introduced or sharpened by Welsh-medium education. However, Thomas (1987: 108) stresses the limited size of his sample, and that his results represent only "trends in usage, and the kind of data which it would be useful to investigate in a fuller enquiry". To that task rises Jones (1994; 1998).

With a broadly similar research design, Jones (1998: 45) compares two communities: low Welsh-density Rhymney in South Wales, with 6.7% of residents aged 3+ Welsh-speaking; and high Welsh-density Rhosllannerchrugog, in Northeast Wales, with 38.1% (Jones 1998: 158). In

Rhymney, she notes that Welsh is primarily acquired at school, so she only records secondary bilinguals. In Rhosllannerchrugog, owing to higher levels of home use, she also records primary bilinguals. The comparisons are especially instructive for our discussion.

Signs of declining diversity are detected in a range of features, for example soft mutations; that is, the replacement of voiceless with voiced consonants in certain environments, as in [k], [p] and [t] becoming [g], [b] and [d]. "[W]hile still used in a historically appropriate way by two-thirds or more of the adult informants", it "was far more unstable amongst the younger generation who, in most cases, omitted it altogether" (Jones 1998: 59). Similarly: "Adjective lenition after a feminine noun was not well preserved", and "the 'tip' had obviously occurred with the younger generation" (Jones 1998: 66). Of education, Jones (1998: 71) notes: "The high instance of soft mutation made in feminine nouns after the numeral *un* ('one') and the relatively high maintenance of gender-marked numerals also suggests that these are grammar points which may have been emphasized in the classroom".

Examining other variables, Jones (1998: 72–74) notes simplification in the distribution of possessive pronouns in all age groups, with the oldest speakers simplifying these dialectal distinctions the least. Jones (1998: 81) is careful to caution that similar loss of dialect features is found in most other "healthy" languages. What is unique to the Welsh case is "the quantity of changes" and "the accelerated rate at which they are taking place" (Jones 1998: 81).

Dialect loss turned to dialect disappearance in other cases, for example post-tonic devoicing (provection), "eliminated from the speech of the younger generation" (Jones 1998: 93). Similarly, Jones (1998: 95) found the local dialect feature third-person singular preterite ending *-ws* "has almost totally disappeared […] completely replaced by Standard *-odd* in the speech of all but the oldest informants". There was overall "a large degree of standardization of the speech of the under forties, this drops dramatically in the speech of informants aged between 40 and 74, while informants aged 75 and over show no evidence of standardization" (Jones 1998: 101). Overall, certain local features of the Rhymney dialect "had to all intents and purposes been eliminated from the speech of the schoolchildren" (Jones 1998: 109).

In the second community, Rhosllannerchrugog, similar trends obtain in secondary bilinguals. Perhaps the most interesting comparisons for our discussion are between, on the one hand, secondary bilinguals in Welsh-

medium education and primary bilinguals in English-medium education. The latter speak Welsh at home but are not taught it in school, and so are not exposed to normative pressures from the supralocal prescribed standard. Dialect loss in Rhosllannerchrugog was apparent, as in Rhymney. What stands out is that secondary bilinguals in Welsh-medium education used significantly fewer dialectal forms than primary bilinguals in English-medium education. The Rhosllannerchrugog feature of inserting an epenthetic vowel in certain word-final clusters appeared on the wane, but English-medium educated Welsh speakers "were retaining this dialect feature to a greater extent" (Jones 1998: 189).

In most cases, declining diversity was more advanced in Rhosllannerchrugog. The English-medium educated Welsh speakers showed greater declines in their use of Welsh overall, but significantly less assimilation to Standard Welsh (and never showing greater dialect loss):

> The speech of these children, who learn Welsh at home and do not receive Welsh-medium education, is still heavily coloured by local features. This is irrefutable evidence of the influence of Welsh-medium education on the local dialect in Rhosllannerchrugog. Most significant of all was the fact that a correlation was found between the results obtained in Rhosllannerchrugog – a relatively strong Welsh-speaking community – and those obtained in Rhymney – a relatively Anglicized community. […] [T]he Standard is gaining substantial ground in the speech of these informants with each successive age group. (Jones 1998: 204)

There is also the matter of peer pressure among young people, reported in Rhosllannerchrugog as creating "stigma" in local dialect features, which "provoked […] a conscious attempt to conform to a more standardised variety" (Jones 1998: 196). These younger respondents saw local dialects as irrelevant, even divisive (Jones 1998: 227). (See also Robert 2009: 95, on secondary bilinguals "drowning out" primary bilinguals in Welsh schools.)

The main dialectological analyses in Rhymney and Rhosllannerchrugog were followed up by matched guise perceptual tests, in which young people struggled to recognise their own local Welsh dialect. These findings can be instructively compared to a separate study conducted more recently, investigating 15-year-olds' recognition of dialects of English in the same areas (Garrett et al. 2003: 200). The two studies can be compared as follows:

> Rhymney: 21% recognition of local dialects of Welsh (Jones 1998: 117), set against 27.6% recognition of local dialects of English (Garrett et al. 2003: 200);

> Rhosllannerchrugog, 32% recognition of local dialects of Welsh (Jones 1998: 209–210), set against 48.3% recognition of local dialects of English (Garrett et al. 2003: 200).

In Rhosllannerchrugog, Jones (1998: 209–210) notes that "many of the [Welsh local dialect] words were so unfamiliar […] that they identified them as coming from the opposite end of the country". It should also be noted that the Garrett et al. (2003) study actually made accurate recognition of local dialect features less likely for English than Jones did for Welsh, in three ways. First, Garrett et al. asked open questions, while Jones gave multiple-choice selections. Second, Garrett et al. required greater precision, by splitting Wales into six zones (to Jones' four). Third, Garrett et al. conducted their study several years later, so if these trends are ongoing, then the later the study, the less chance of accurate recognition.

Overall then, both production and recognition of local dialects of Welsh are declining in these two quite different areas of Wales. Jones (1998: 101) concludes that "there has been a 62 per cent increase in dialect loss over the past sixty years". She attributes this to the predominant exposure of secondary bilinguals to "Standard Oral Welsh" in the classroom, "a nationwide, non-localized variety of the national language" (Jones 1998: 116). "Their Welsh is becoming a non-locatable amalgam of elements drawn from all over Wales" (Jones 1998: 117).

Though cautious not to over-generalise, Jones (1998: 229) mentions that her chosen research sites are "typical of their kind". She further ventures that "in Rhymney and Rhosllannerchrugog […] we are witnessing an instance of language suicide [in] which […] the dialects of Wales are becoming progressively divested of some of their phonetic regional features and idiosyncratic lexical items" (Jones 1994: 256). Figures 2 and 3 represent the combined data for dialect loss in both locations, showing a general decline in structural distinctions.

Whether or not Jones' (1998: 208) future vision of "a variety of Welsh […] devoid of all regional features" is realised, the point remains that modern spoken Welsh appears to be declining in diversity. Reprising our benchmark of linguistic diversity, dialects appear to be converging; variation and variability are decreasing. As Jones (1998: 137) concludes: "while the status of the Welsh language as a whole may be improving, the fate of its dialects is more pessimistic".

The fate of dialects in language revitalisation is a live topic of debate in relation to linguistic justice: distinguishing *inter-linguistic* and *intra-linguistic justice* (De Schutter 2017); and noting that a raised profile for

languages may exacerbate or even create injustices among speakers of their dialects (Blommaert 2001). We are making a parallel critique here: to repeat a point made earlier, even if dialects were somehow recognised, even protected, this would just nudge the reductionism down a level, on to dialects as a smaller unit of language. Teaching a standard language does not automatically cause mass linguistic conformity, nor does it crush innovation. However, it does seem to introduce or exacerbate pressures that reduce variation, and inhibit variability.

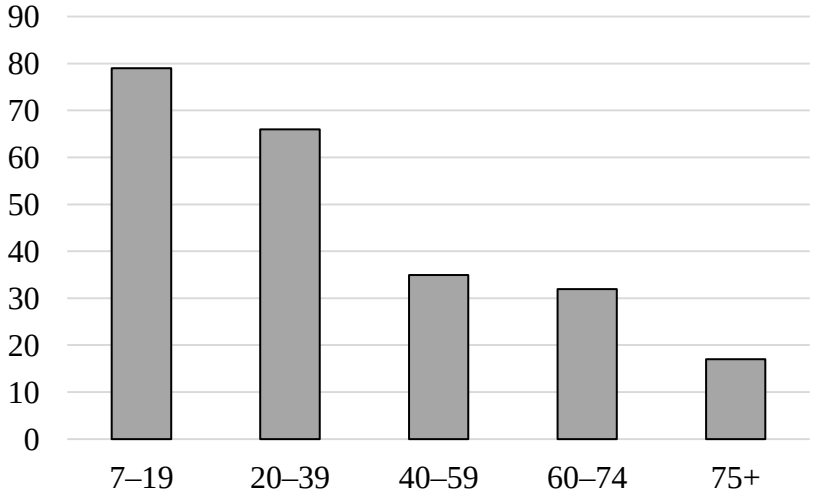

Figure 2. % cross-variable, inter-group comparison of dialect loss in Rhymney, by age (Jones, 1998: 101) [*n* figures not in original]

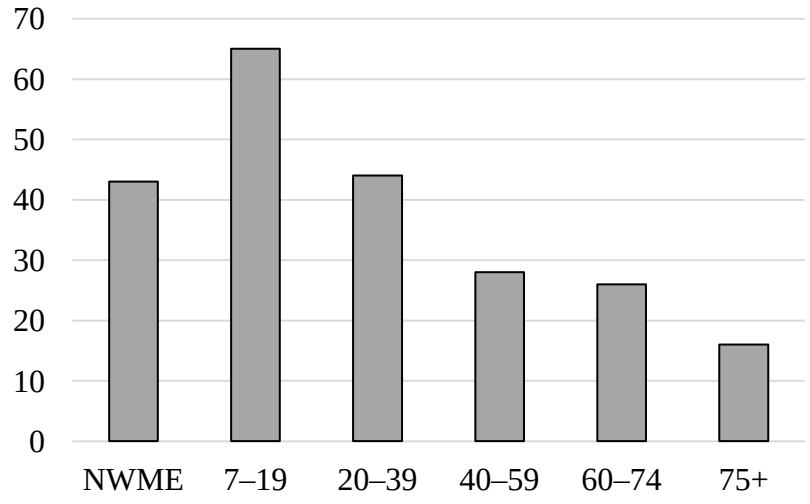

Figure 3. % cross-variable, inter-group comparison of dialect loss in Rhosllannerchrugog, by age (Jones 1998: 204) ['NWME' – no Welsh-medium education; *n* figures not in original]

Lastly, it may appear that attention is being unduly centred on the present situation, seeing issues such as the tension between standard and dialect as only an effect of recent policy. The situation is patently far more complex, with longstanding awareness among Welsh speakers of tensions between the standard language and natural, informal varieties (see e.g. Robert 2009). Our argument is not that these pressures on diversity are entirely new, but that they have only been magnified by contemporary language policy.

4.2 Cornish

We have seen dialects of Welsh weakening and converging amid a putatively successful revival effort, proceeding in alignment with the European Charter. Welsh is a well-resourced and long-standing revival, with a substantial body of speakers. But how does linguistic diversity fare at a different extreme, where the language died long ago, has lost its intergenerational link to native speakers, and has been painstakingly reconstructed for modern use? In this section, we discuss diversity in the Cornish language movement, illustrated by a mix of primary interview data with language activists, documentary data, and other published research.

The Cornish language died slowly from the 16th to the early 19th century (Treenoodle 1846: 1–4; Jago 1882: 13; Jenner 1904: 11–23). Successive mining booms in Cornwall over these centuries served to churn the Cornish population, and spur massive in-migration (Pounds 1943: 45), gradually diluting Cornish with English. There were efforts to catalogue the language in its twilight years, though fragmented and largely amateur; and many of these piecemeal records were subsequently lost (see e.g. Pryce 1790: iv). Nowadays, the surviving historical record of written Cornish is estimated at a meagre 176,000 words total (George & Broderick 2009: 754). From this punishingly scant corpus, revivalists with varying levels of linguistic training have attempted to reconstruct a full language. This has involved filling gaps in grammar and lexicon by extrapolation, as well as by borrowing and adapting from surviving related languages – principally Breton and Welsh (see Sayers & Renkó-Michelsén 2015).

Crucially this was never an orchestrated or centrally planned effort, but was conducted independently by different people, to different extents, at different times over the centuries. Given the scarcity of the corpus, and the variety of approaches to reconstruction, there are now different "versions" of reconstructed Cornish. So, although Cornish has no dialects

as such, nevertheless there is a form of diversity. Can these versions all be promoted within Charter-based revival activity? And in relation to variability, is the contemporary revival setting up the conditions for new forms of diversity to flourish, including spoken vernaculars as the reconstructed language re-enters daily use?

In our earlier research (e.g. Sayers 2012; Sayers & Renkó-Michelsén 2015), and in the work of others, we have found pressures on diversity in reconstructed Cornish linked to the funding and evaluation frameworks emanating from the Charter. Prior to 2001, Cornish was promoted by separate voluntary groups, each favouring a different one of the reconstructed versions noted above – with little mutual dialogue. There were occasional small grants from local government and other funders (Sayers 2012: 101), but otherwise no large-scale funding. Each version had its supporters, and the language revival proceeded along these parallel avenues, representing a nascent form of diversity.

The UK Government recognised Cornish under Part II of the Charter in 2001 (weaker than Part III, but still foregrounding education). Three years later, the *Strategy for the Cornish Language* was published, including a priority for "a single written form of Cornish for use in official documentation and formal education" (CCC 2004: 18). What would this mean for the existing versions of the language?

Between 2006 and 2009, a combined local, national and European funding package of £600,000 was awarded to decide how Cornish would be officially promoted. This figure dwarfed any funding previously afforded the language; and the central goal of this lengthy consultation was a singular standard form for official promotion.

But debates about standardisation between activists and officials were longer running. As one of our interviewees, a leading activist, put it in 2005: "Government use it as an excuse. Why hasn't Cornish been put in schools already? Well there are four different spelling systems, which one should we use? You'll have to choose one for official purposes. This is the answer that's been given."

As the 2004 *Strategy* foreshadowed, the 2006–2009 consultation was premised on the need for a singular standard. In large part, then, the consultation was an exercise in diplomacy between what had essentially become opposing factions. One eventual possibility was to select one of the existing versions for official use. Shortly before the consultation began, the above-quoted activist commented: "You'll never get an agreement on one type of Cornish, not within the next ten years anyway. That's pie in the sky,

because there are die hards." The same activist continued, "the last thing we want […] is an amalgamation of the current systems to provide yet another form of Cornish. […] It wouldn't have authenticity." This same indifference was echoed by other activists, a view perhaps coloured by their own investment in each version.

But as the consultation progressed, it eventually became clear that agreement on an existing version would be less expeditious than, after all, amalgamating existing versions into a new version. Two professional linguists from outside the UK were employed to synthesise what became known as the Standard Written form, SWF (commonly pronounced 'swoof'). Their report (Bock & Bruch 2008) outlined a new version which accommodated some variation in spellings (a diplomatic nod to existing versions), but ultimately crystallised Cornish into a standard vehicle for official adoption.

Adoption of SWF enabled longer-term funding from central government, to fulfil the requirements of the Charter – principally production and distribution of teaching resources. By 2016, twenty schools across Cornwall were involving Cornish to different extents in an extracurricular capacity. Official adoption and centralisation of resources enabled three learning packages, circulated to all primary schools in Cornwall, as well as taster sessions and other contributions. Cornish did not enter the national curriculum but its presence grew across Cornwall. The official revival placed no explicit constraints on the use of other versions of Cornish, but only leant specific support to SWF. So, from a base of limited diversity in this nascent revival, Charter recognition raised its profile and its budget, but drove new constraints on diversity.

The present moment is actually a strange time to be writing about Cornish. In April 2016, the UK Government ceased its annual funding. This was couched in terms of a wider devolution of central government responsibility to local authorities; but was later revealed by a former Cabinet member as simply the end of a rather cynical political deal to secure fiscal savings elsewhere (Sayers 2017). There remains a Cornish Language Office funded by Cornwall County Council, which administers small grants for one-off projects, but with much more limited scope. Road signs are still produced bilingually, though just when signs need replacing, so this incurs no extra cost. The prior central government funding for Cornish had come from the Department for Communities and Local Government; but in the government's spring budget of 2017, DCLG saw a cut of 24%. This department is otherwise predominantly responsible for

housing supply and public services. Given that Cornwall is also the poorest region of England (ONS 2011), and its economy continues to decline (Cornwall Council 2013), calls for reinstatement of this funding may fall on deaf ears.

Certain projects are funded through other means, for example a nine-day Cornish language festival in February/March 2017 funded by the National Lottery "Celebrate" fund. The Akademi Kernewek (www.akademikernewek.org.uk) – set up during the period of central government funding to develop dictionaries, terminology, etc. – is still active, though in a largely voluntary capacity. Various other pre-existing voluntary bodies continue to operate. Meanwhile the raised profile of Cornish, along with the relative accessibility of SWF, has enabled more self-sustaining activities, not least use of Cornish by some companies, for example certain bus announcements, signage in pubs, hotels and shops, and widespread use in email signatures.

But overall the light has dimmed, and the goal of substantially increasing everyday spoken use of Cornish is somewhat adrift. In a recent language action plan (Cornwall Council 2016), 10 out of 23 goals are listed as red (stalled), pending further lobbying of central government to reverse its funding withdrawal. Given the 24% cut to DCLG noted above, this is a period of existential uncertainty for the revival.

But to return to our overarching theme, the large-scale funding which did occur was only made possible by agreement on, and propagation of, SWF. The lack of a single agreed standard was the major logjam, and despite indifference towards an amalgamated standard, SWF nevertheless enabled a level of visibility and recognition previously confined to fantasy for Cornish activists.

So, what of linguistic diversity in reconstructed Cornish? Even if the language movement regains momentum in future, SWF remains the bedrock. Without that lies the unappealing prospect of reignited factional dispute, and dilution or diversion of institutional support. Eyes are focused on securing sustainable state funding for continued rollout of SWF, through education. This of course recalls all the factors outlined in the Welsh case, with normative pressures from above and social pressures from below against variation and variability. If the revival finds its feet again, linguistic diversity seems unlikely to follow.

5 Conclusion

Although the primary law of the EU urges respect for linguistic diversity, our legal analysis suggests this is something of a sandcastle at high tide. The language regime of the EU conceived sixty years ago, the respective jurisprudence of the European Court of Justice, and the practice of the institutions of the EU, result in a hierarchical language regime based on restricted multilingualism. The great divide is between official languages of the EU and non-official languages, resulting in a distinction between *de facto* working languages on the one hand and official languages of the EU on the other. The Commission sees language learning as the key to securing linguistic diversity, but this is problematic. Multilingualism conceived in this way actually undermines linguistic diversity overall, creating pressures towards greater linguistic homogeneity of Member States. This in turn increases the likelihood of language loss, and tacitly reinforces the dominance of English as a lingua franca in Europe. Meanwhile the significantly more numerous allochthonous languages spoken across Europe (crowding out autochthonous languages around four to one) are summarily excluded altogether.

Our sociolinguistic analysis curls up further question marks over linguistic diversity within the European regime of promoting minority languages. With its focus on formal language learning, it favours what Wright (2007a) refers to as a "language-as-system" approach, which is fundamentally at odds with the fluctuating and enigmatic reality of "language-as-practice".

None of this is intended to assert that the intention to enshrine multilingualism or promote specific minority languages is outright folly, or a waste of resources. Nor are we claiming that linguistic diversity itself is necessarily, incontrovertibly, a positive end in itself. We have made no case for these positions, nor do we advance them now. Our goal here has been more modest, to hold European language policy up to its own claim about linguistic diversity, to expose that claim to sustained scrutiny from different angles, and to demonstrate some wrinkles in its logic.

This is more than just pedantic heckling. Although we did not advance arguments about the value of this or that position towards linguistic diversity, nevertheless we have shown that a fundamental rhetorical basis of European language policy is ultimately rather threadbare – winnowed away and watered down by the rationalistic funnelling of resources in supra-national political institutions, the mundane practicalities of language

planning on the ground, and heightened social pressures to conform. This in turn is an important contribution to a wider debate on the shortfalls of such grand policy claims, and of grand narratives generally.

We end by repeating our central assertion, that linguistic diversity is complicated, much more so than the existence of a series of countable languages, or even dialects within them. Attempts to officially promote a circumscribed number of specific languages will do little to help this, and may introduce new pressures on diversity which, by every official measure, will be missed. Linguistic diversity transcends language boundaries, and may be harmed by institutional intervention. Thus, linguistic diversity may ultimately belong outside the discourse of contemporary European language policy.

Acknowledgements

The idea for our collaboration began in 2013, over coffee at the Language and Super-diversity conference at the University of Jyväskylä, Finland. We are grateful to the organisers for bringing us together. Our underlying ideas emerged previously from our respective doctoral research, for which we thank our supervisors and sponsors. In 2014, we presented the arguments in this article to Sociolinguistic Symposium 20 (coincidentally back in Jyväskylä); then at the Multidisciplinary Approaches in Language Policy & Planning Conference at the University of Calgary, Canada; and finally, in 2015 to the Changing English conference at the University of Helsinki, Finland. We are thankful to all these audiences for constructive feedback. We acknowledge the help and support of our employers, in allocating time for research and providing resources for conference attendance. Along the way our families have provided patient and loving support, and for that there can be no adequate thanks.

References

Aitchison, Jean & Carter, Harold. 2000. *Language, economy and society: The changing fortunes of the Welsh language in the twentieth century*. Cardiff: University of Wales Press.

Arzoz, Xavier. 2008. The protection of linguistic diversity through Article 22 of the Charter of Fundamental Rights. In Arzoz, Xabier (ed.), *Respecting linguistic diversity in the European Union*, 145–174. Amsterdam: John Benjamins.

Blommaert, Jan. 2001. The Asmara Declaration as a sociolinguistic problem: Notes in scholarship and linguistic rights. *Journal of Sociolinguistics* 5(1). 131–142.

—— 2005. From fieldnotes to grammar. Artefactual ideologies and the textual production of languages in Africa. (http://cscs.uchicago.edu/papers/blommaert.pdf) (Accessed 21 June 2017)

Blommaert, Jan & Verschueren, Jef. 1998. *Debating diversity: Analysing the discourse of tolerance*. London: Routledge.

Bock, Albert & Bruch, Benjamin. 2008. *An outline of the Standard Written Form of Cornish*. (www.kernowek.net/Specification_Final_Version.pdf) (Accessed 2017-03-06.)

Bradean-Ebinger, Nelu. 2011. Kisebbségi és regionális nyelvek az EU-ban [Minority and regional languages in the EU]. *International Relations Quarterly* 2(2). 1–9.

Broadbent, Jane & Laughlin, Richard. 2002. Public service professionals and the New Public Management: Control of the professions in the public services. In McLaughlin, Kathleen & Osborne, Stephen P. & Ferlie, Ewan (eds.), *New Public Management: Current trends and future prospects*, 95–108. Abingdon: Routledge.

Carrera, Sergio. 2014. Integration of immigrants in EU law and policy: Challenges to rule of law, exceptions to inclusion. In Azoulai, Loïc & de Vries, Karin (eds.), *Migration and EU law and policy*, 149–187. Oxford: Oxford University Press.

CCC (Cornwall County Council). 2004. *Strategy for the Cornish language*. Truro: Cornwall County Council. (www.linguae-celticae.org/dateien/Cornish_Language_Strategy.pdf) (Accessed 2015-06-17.)

CoE (Council of Europe). 1992a. *European Charter for Regional or Minority Languages: Explanatory Report*. European Treaty Series 148. Strasbourg: Council of Europe.

—— 1992b. *European Charter for Regional or Minority Languages*. European Treaty Series 148. Strasbourg: Council of Europe.

—— n.d. Monitoring the European Charter for Regional or Minority Languages. (https://www.coe.int/en/web/european-charter-regional-or-minority-languages/monitoring) (Accessed 2017-06-20.)

Cornwall Council. 2013. *Economy monitoring: Monthly update December 2013*. Truro: Cornwall Council.

—— 2016. Framework Convention for the Protection of National Minorities: Cornish national minority action plan. (www.cornwall.gov.uk/media/21282837/action-plan-august-2016-v10.pdf) (Accessed 2017-02-10.)

Coupland, Nikolas. 2011. Welsh linguistic landscapes 'from above' and 'from below'. In Jaworski, A. & Thurlow, C. (eds.), *Semiotic landscapes: Image, space, text*, 77–101. London: Continuum.

Coupland, Nikolas & Bishop, Hywel & Williams, Angie & Evans, Betsy & Garrett, Peter. 2005. Affiliation, engagement, language use and vitality: Secondary school students' subjective orientations to Welsh and Welshness. *International Journal of Bilingual Education & Bilingualism* 8(1). 1–24.

Coupland, Nikolas & Bishop, Hywel & Evans, Betsy & Garrett, Peter. 2006. Imagining Wales and the Welsh language: Ethnolinguistic subjectivities and demographic flow. *Journal of Language & Social Psychology* 25(4). 351–376.

Derlén, Mattias. 2011. In defence of (limited) multilingualism: Problems and possibilities of the multilingual interpretation of European Union law in courts. In Kjaer, Anne L. (ed.), *Linguistic diversity and European democracy*, 156–157. Farnham: Ashgate.

de Schutter, Helder. 2017. Global linguistic justice and English as a Lingua Franca. Presentation to *Brexit, English, and the language policy of the European Union*, Cardiff University, March 10, 2017.

de Witte, Bruno. 2004. Language law of the European Union: Protection or eroding linguistic diversity. In Smith, Rachael C. (ed.), *Culture and European Union law*, 205–241. Oxford: Oxford University Press.

—— 2008. The protection of linguistic diversity through provisions of the EU Charter other than article 22. In Xabier, Arzoz (ed.), *Respecting linguistic diversity in the European Union*, 175–190. Amsterdam: John Benjamins.

Dufva, Hannele & Aro, Mari & Suni, Minna & Salo, Olli-Pekka. 2011: Onko kieltä olemassa? Teoreettinen kielitiede, soveltava kielitiede ja kielen oppimisen tutkimus [Does language exist? Theoretical linguistics, applied linguistics and studies in language learning]. *Soveltavan kielitieteen tutkimuksia* 3. 22–34.

Dunbar, Robert. 2000. Implications of the European Charter for Regional or Minority Languages for British linguistic minorities. *European Law Review* 25 Supp (Human rights survey). 46–69.

EC (European Commission). 2002. Presidency conclusions: Barcelona European Council, Barcelona, 15 and 16 March 2002. (http://ec.europa.eu/invest-in-research/pdf/download_en/barcelona_european_council.pdf) (Accessed 2017-03-14.)

—— 2003. Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions – Promoting Language Learning and Linguistic Diversity: an Action Plan 2004–2006 (COM/2003/0449). (http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex:52003DC0449) (Accessed 2017-08-25.)

—— 2005a. Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions – A new framework strategy for multilingualism. (http://eur-lex.europa.eu/legal-content/GA/ALL/?uri=celex:52005DC0596) (Accessed 2016-10-21.)

—— 2005b. Communication from the Commission to the European Parliament and the Council – The European indicator of language competence. (http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52005DC0356) (Accessed 2016-01-.)

—— 2008. Reference document. European Commission staff working document on strengthening actions and tools to meet integration challenges – Report to the 2008 ministerial conference on integration. SEC(2008) 2626. (http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52008IR0212) (Accessed 2016-10-21.)

Eckert, Penelope. 2000. *Linguistic variation as social practice*. Oxford: Blackwell.

EEC Council. 1958. Regulation No 1 determining the languages to be used by the European Economic Community. (http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31958R0001) (Accessed 2017-03-10.)

Erker, Daniel. 2017. The limits of named language varieties and the role of social salience in dialectal contact: The case of Spanish in the United States. *Language & Linguistics Compass* 11: e12232.

European Parliament. 2003. Resolution with recommendations for European regional and lesser-used languages – the languages of minorities in the EU – in the context of enlargement and cultural diversity (2003/2057(INI). (http://www.europarl.europa.eu/sides/getDoc.do?reference=P5-TA-2003-0372&language=EN) (Accessed 2017-08-25.)

Eurostat. 2015. Migration and migrant population statistics in EU-28. (http://emn.ie/index.jsp?p=128&n=229) (Accessed 2016-07-.)

Extra, Guus & Verhoeven, Ludo (eds.). 1993. *Immigrant languages in Europe*. Clevedon: Multilingual Matters.

Farrell, Shaun & Bellin, Wynford & Higgs, Gary & White, Sean. 1997. The distribution of younger Welsh speakers in anglicised areas of south east Wales. *Journal of Multilingual and Multicultural Development* 18(6). 489–495.

FEL (Foundation for Endangered Languages) Canada. 2017. Winter Newsletter.

Felföldi, Rita. 2011. A kisebbségi nyelvek helyzetének változása az Európai Unióban [Changes in the status of minority languages in the European Union]. (https://btk.ppke.hu/uploads/articles/6414/file/felfoldirita.pdf) (Accessed 2013-05-04.)

Fidrmuc, Jan. 2011. The economics of multilingualism in the EU. *Brunel Economics and Finance Working Paper Series* 11-04. (https://www.brunel.ac.uk/__data/assets/pdf_file/0007/342763/CEDI_11-04.pdf) (Accessed 2017-06-21.)

Fidrmuc, Jan & Ginsburgh, Victor A. & Weber, Shlomo. 2006. Economic challenges of multilingual societies. Working Paper No.11/04. (www.researchgate.net/publication/228651837) (Accessed 2013-09-14.)

FUEN (Federal Union of European Nationalities). n.d. Language Diversity – a European campaign to raise awareness for multilingualism and linguistic diversity in Europe. (http://language-diversity.eu/en/wp-content/uploads/sites/3/2014/09/3_Our-language-diversity-campaign_Linguistic-diversity-and-multilingualism-in-Europe.pdf) (Accessed 2017-06-21.)

García, Ofelia & Lin, Angel M. Y. 2017. Translanguaging in bilingual education. In García, Ofelia & Lin, Angel M. Y. & May, Stephen (eds.), *Bilingual and multilingual education* (3rd ed.), 117–130. New York: Springer.

Garrett, Peter & Coupland, Nikolas & Williams, Angie. 2003. *Investigating language attitudes: Social meanings of dialect, ethnicity and performance*. Cardiff: University of Wales Press.

George, Ken & Broderick, George. 2009. The revived languages: Cornish and Manx. In Ball, Martin J. & Müller, Nicole (eds.), *The Celtic languages* (2nd ed.), 753–769. London: Routledge.

Grin, François. 2003. *Language policy evaluation and the European Charter for Regional or Minority Languages*. New York: Palgrave Macmillan.

Henrard, Kristin. 2004. An investigation into the desirable, and possible role of the Language Charter in expanding on Article 22 of the EU's Charter of Fundamental Rights. Presentation to *2nd Mercator international symposium: Europe 2004: A new framework for all languages?*, Tarragona, 27–28 February 2004.

Higgs, Gary & Williams, Colin & Dorling, Danny. 2004. Use of the census of population to discern trends in the Welsh language: an aggregate analysis. *Area* 36(2). 187–201.

Huws, Catrin Fflur. 2006. The Welsh Language Act 1993: A measure of success? *Language Policy* 5(2). 141–160.

Huws, Meri. 2016a. The position of the Welsh language – challenges and opportunities. (http://www.iwa.wales/click/2016/09/position-welsh-language-challenges-opportunities/) (Accessed 2017-02-02.)

—— 2016b. The position of the Welsh language 2012–2015: Welsh Language Commissioner's 5-year report. (www.comisiynyddygymraeg.cymru/English/Commissioner/Pages/5-year%20Report.aspx) (Accessed 2017-02-02.)

Jaffe, Alexandra. 2004. Language rights and wrongs: A commentary. In Freeland, Jane & Patrick, Donna (eds.), *Language rights and language survival: Sociolinguistic and sociocultural perspectives*, 273–282. Manchester: St. Jerome.

—— 2007. Discourses of endangerment: Contexts and consequences of essentializing discourses. In Duchêne, Alexandre & Heller, Monica (eds.), *Discourses of endangerment: Interest and ideology in the defence of languages*, 57–75. London: Continuum.

Jago, Fred W. P. 1882. *The ancient language, and the dialect of Cornwall*. Truro: Netherton & Worth. (www.archive.org/details/ancientlanguaged00jago) (Accessed 2015-06-17.)

Jenner, Henry. 1904. *A handbook of the Cornish language*. London: Ballantyne, Hanson & Co. (www.archive.org/details/handbookofcornis00jennuoft) (Accessed 2015-06-17.)

Jones, Mari C. 1994. Standardization in modern spoken Welsh. In Parry, M. M. & Davies, W.V. & Temple, R. A. M. (eds.), *The changing voices of Europe*, 243–264. Cardiff: University of Wales Press.

—— 1998. *Language obsolescence and revitalization: Linguistic change in two sociolinguistically contrasting Welsh communities*. Oxford: Clarendon Press.

Kasstan, Jonathan, 2017. New speakers: Challenges and opportunities for variationist sociolinguistics. *Language and Linguistics Compass* 11: e12249.

Király, Miklós. 2007. *Egység és Sokféleség* [Unity and Diversity]. Budapest: Új Ember Kiadó

Kraus, Peter A. 2008. *A union of diversity: Language, identity and polity-building in Europe*. Cambridge: Cambridge University Press.

Láncos, Petra L. 2009. Részvételi jogok és a nyelvi sokszínűség az Európai Unióban [Participation rights and linguistic diversity in the European Union]. *Miskolci Jogi*

Szemle [*Miskolc Legal Review*] 4(2). 112–128. (www.mjsz.uni-miskolc.hu/200902/8_lancospetra.pdf) (Accessed 2013-06-17.)

—— 2012. *Nyelvpolitika és nyelvi sokszínűség az Európai Unióban* [*Language Policy and Linguistic Diversity in the European Union*]. Budapest: Pázmány Press.

Lanstyák, István. 2004. Az Európai Unió nyelvpolitikája és a Szlovákiában beszélt nyelvek [The language policy of the European Union and the languages spoken in Slovakia]. *Társadalomtudományi Szemle* [*Social Science Review*] 6(4). 43–66.

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.

Mac Giolla Chríost, Diarmait. 2007. *Language and the city*. New York: Palgrave Macmillan.

Marcellesi, Jean-Baptiste. 2003. *Sociolinguistique: Épistémologie, langues régionales, polynomie* [*Sociolinguistics: Epistemology, regional languages, polynomy*]. Paris: l'Harmattan.

Marí, Isidor & Strubell, Miguel. 2002. The linguistic regime of the European Union: Prospects in the face of enlargement. Presentation given at *Linguistic proposals for the future of Europe*, Barcelona, May 31 2002.

May, Stephen. 2000. Accommodating and resisting minority language policy: The case of Wales. *International Journal of Bilingual Education and Bilingualism* 3(2). 101–128.

—— 2003. Language, nationalism and democracy in Europe. In Hogan-Brun, Gabrielle & Wolff, Stefan (eds.), *Minority languages in Europe: Frameworks – status – prospects*, 210–232. New York: Palgrave Macmillan.

Mayer, Franz C. 2006. Europäisches Sprachverfassungsrecht [European Constitutional Law]. *Walter Hallenstein Institut Papers*. 2006/1. 22.

McLeod, Wilson. 2008. A new multilingual United Kingdom? The impact of the European Charter for Regional or Minority Languages. In Pertot, Susanna & Priestly, Tom & Williams, Colin H. (eds.), *Rights, promotion and integration issues for minority languages in Europe*, 44–59. Basingstoke: Palgrave Macmillan.

Milian-Massana, Antoni. 2008. Recognition of the Basque language in EU law: A pending issue? In Totoricaguena, Gloria & Urrutia, Iñigo (eds.), *The legal status of the Basque language today: One language, three administrations, seven different geographies and a diaspora*, 93–112. Boise, ID: Cenarrusa.

Muehlmann, Shaylih. 2007. Defending diversity: Staking out a global common interest? In Duchêne, Alexandre. & Heller, Monica. (eds.), *Discourses of endangerment: Ideology and interest in the defence of languages*, 14–34. London: Continuum.

Murray, Robert W. 2006. The place of historical linguistics in the age of structuralism. In Auroux, Sylvain & Koerner, Ernst F. K. & Niederehe, Hans-Joseph & Versteegh, Kees (eds.), *History of the language sciences* vol. 3, 2430–2445. Berlin: De Grutyer.

Nettle, Daniel. 1999. *Linguistic diversity*. Oxford: Oxford University Press.

Nettle, Daniel & Romaine, Suzanne. 2000. *Vanishing voices: The extinction of the world's languages*. Oxford: Oxford University Press.

Nic Shuibhne, Niamh. 2004. Case comment on Kik v Office for Harmonisation in the Internal Market. *Common Market Law Review* 41: 1093–1111.

—— 2008. EC Law and Minority Language Policy: Some Recent Developments. In Arzoz, Xabier (ed.), *Respecting Linguistic Diversity in the European Union*, 123–143. Amsterdam: John Benjamins.

Office for National Statistics. 2004. *Welsh language – Welsh speakers increase to 21%*. (https://web.archive.org/web/20040118055724/www.statistics.gov.uk/CCI/nugget.asp?ID=447) (Accessed 2017-11-15.)

—— 2011. *Regional, sub-regional and local Gross Value Added*, 2010. (www.ons.gov.uk/ons/dcp171778_250356.pdf) (Accessed 2017-04-06.)

—— 2012. 2011 Census: Key Statistics for Wales, March 2011. (www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuskeystatisticsforwales/2012-12-11) (Accessed 2017-04-06.)

ONS (see Office for National Statistics).

Ó Riagáin, Dónall. 2002. The lesser used languages of Europe and their participation in the programmes of the European Union. Presentation to *Linguistic proposals for the future of Europe*, Barcelona, 31 May - 1 July.

Perley, Bernard C. 2012. Zombie linguistics: Experts, endangered languages and the curse of undead voices. *Anthropological Forum* 22(2). 133–149.

Petrovic, John E. 2005. The conservative restoration and neoliberal defences of bilingual education. *Language Policy* 4(4). 395–416.

Pounds, Norman J. G. 1943. Population movement in Cornwall, and the rise of mining in the 18th century. *Geography* 28(2). 37–46.

Pryce, William. 1790. *Archaeologia Cornu-Britannica*. Sherborne: Crutwell. (www.archive.org/details/archaeologiacor01prycgoog) (Accessed 2015-06-17.)

Robert, Elen. 2009. Accommodating "new" speakers? An attitudinal investigation of L2 speakers of Welsh in south-east Wales. *International Journal of the Sociology of Language* 195. 93–116.

Sallabank, Julia. 2005. Prestige from the bottom up: a review of language planning in Guernsey. *Current Issues in Language Planning* 6(1). 44–63.

Sayers, Dave. 2012. Standardising Cornish: The politics of a new minority language. *Language Problems & Language Planning* 36(2). 99–119.

—— 2015. Getting past the 'indigenous' vs. 'immigrant' language debate. *Language on the Move*. (www.languageonthemove.com/getting-past-the-indigenous-vs-immigrant-language-debate) (Accessed 2017-04-06.)

—— 2017. Cornish language funding 2010–2016 as a political 'bribe'. (https://academia.edu/31944206) (Accessed 2017-04-18.)

Sayers, Dave & Renkó-Michelsén, Zsuzsanna. 2015. Phoenix from the ashes: Reconstructed Cornish in relation to Einar Haugen's four-step model of language standardisation. *Sociolinguistica* 29. 17–38.

Schedler, Kuno & Proeller, Isabella. 2002. The new public management: A perspective from mainland Europe. In McLaughlin, Kathleen & Osborne, Stephen P. & Ferlie, Ewan (eds.), *New Public Management: Current trends and future prospects*, 163–180. Abingdon: Routledge.

Skutnabb-Kangas, Tove. 2002. Why should linguistic diversity be maintained and supported in Europe? Some arguments. Strasbourg: Council of Europe.

(www.coe.int/t/dg4/linguistic/Source/Skutnabb-KangasEN.pdf)
(Accessed 2017-11-15.)

Smith, Jennifer & Durham, Mercedes & Fortune, Liane. 2007. "Mam, my trousers is fa'in doon!": Community, caregiver, and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change* 19(1). 63–99.

Stanford, James N. & Preston, Dennis R. 2003. Introduction: The lure of a distant horizon – Variation in indigenous minority languages. In Stanford, James N. & Preston, Dennis R. (eds.), *Variation in indigenous minority languages*, 1–20. Amsterdam: John Benjamins.

Statistics for Wales. 2012. 2011 Census: First Results on the Welsh Language. (http://wales.gov.uk/docs/statistics/2012/121211sb1182012en.pdf) (Accessed 2017-08-25.)

Strubell, Miquel. 2007. The political discourse on multilingualism in the European Union. In Castiglione, Dario & Longman, Christopher (eds.), *The language question in Europe and diverse societies: Political, legal and social perspectives*, 149–184. Oxford: Hart.

Thomas, A. R. 1987. A spoken standard for Welsh: Description and pedagogy. *International Journal of the Sociology of Language*, 66(4). 99–113.

Toggenburg, Gabriel N. 2005. Who is managing ethnic and cultural diversity in the European condominium? The moments of entry, integration and preservation. *Journal of Common Market Studies* 43(4). 717–738.

Torgersen, Eivind & Kerswill, Paul. 2004. Internal and external motivation in phonetic change: Dialect levelling outcomes for an English vowel shift. *Journal of Sociolinguistics* 8(1). 23–53.

Treenoodle, William 'Uncle Jan'. 1846. *Specimens of Cornish provincial dialect*. London: John Russell Smith.
(www.archive.org/details/specimenscornis00sandgoog) (Accessed 2015-06-17.)

Trudgill, Peter. 1999. *The dialects of England* (2nd ed.). Oxford: Blackwell.

van Els, Theo. 2003. Language policy of and for the European Union: Consequences for foreign language teaching in the Member States. In Ahrens, Rüdiger (ed.), *European language policy*, 45–55. Heidelberg: Universitätsverlag.

van Parijs, Philippe. 2008. Linguistic diversity as curse and as by-product. In Arzoz, Xabier (ed.), *Respecting linguistic diversity in the European Union*, 17–46. Amsterdam: John Benjamins.

Welsh Government. 2017. Cymraeg 2050: Welsh language strategy. (http://gov.wales/topics/welshlanguage/welsh-language-strategy-and-policies/cymraeg-2050-welsh-language-strategy/?lang=en) (Accessed 2017-09-25.)

Williams, Colin H. 2008. *Linguistic minorities in democratic context*. New York: Palgrave Macmillan.

Woolard, Kathryn & Schieffelin, Bambi. 1994. Language ideology. *Annual Review of Anthropology* 23. 55–82.

Wright, Sue. 2007a. The right to speak one's own language: Reflections on theory and practice. *Language Policy* 6(2). 203–224.

—— 2007b. What is a language? The difficulties inherent in language rights. In Castiglione, Dario & Longman, Christopher (eds.), *The language question in*

Europe and diverse societies: Political, legal and social perspectives, 81–100. Oxford: Hart.

Contact Information:

Dave Sayers
Department of Humanities
Sheffield Hallam University
Sheffield S1 1WB
United Kingdom
e-mail: dave(dot)sayers(at)antab(dot)net

Petra Lea Láncos
German Research Institute for Public Administration
Freiherr-vom-Stein-Straße 2
67346 Speyer
Germany
e-mail: lancos(at)foev-speyer(dot)de

The dynamics of linguistic contact: Ancient Greek *-ízein* and Latin *-issāre*/*-izāre*/*-idiāre*

Liana Tronci
University for Foreigners of Siena

Abstract

This paper proposes a re-examination of contact phenomena in Ancient Greek and Latin through a description of the Greek verbs in -ίζειν [*-ízein*] and the Latin loans in *-issāre*/*-izāre*/*-idiāre*. This subject has been much debated, especially from the point of view of the recipient language, whereas the donor language has not yet been adequately taken into consideration. This paper intends to fill the gap, by describing the occurrences of Latin loan verbs and comparing them with their Greek sources. In order to understand the mechanisms of interference between the two languages, it is necessary to analyse the textual and cultural significance of both Greek and Latin verbs, and to investigate the pathways followed by Greek verbs in -ίζειν [*-ízein*] to penetrate into Latin. The cultural and textual domains involved in the borrowing process were, on the one hand, the so-called technical languages, which range from that of Christian religion to that of the treatises on medicine, architecture, agriculture, and grammar, and, on the other hand, the language spoken by the Greeks who inhabited Magna Graecia and, after the Roman occupation, transmitted, as slaves and preceptors, their language and culture to the Roman society. The paper discusses how and to what extent this borrowing process influenced the Latin lexicon and, through it, the lexicon of Romance languages. Some new insights are also given concerning the relationship between lexical borrowing and language change. On the one hand, Greek loanwords increased the Latin lexicon; on the other hand, Latin morphology was also involved, because a new derivational process arose through reanalysis. The spreading of the new derivational pattern in Latin appears to be constrained by sociolinguistic factors. Data from Romance languages provide evidence of the relevance of the new pattern for the Latin language and support the idea that spoken Latin was influenced by the Greek language much more than Classical Latin texts show.

Keywords: Ancient Greek, Latin, lexical borrowing, language change

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

1 Introduction

This paper aims to re-examine the general subject of language contact between Ancient Greek and Latin, with the study of a contact-induced language change, namely the arising of the Latin verbs in *-issāre*/*-izāre*/*-idiāre* from lexical borrowing of Greek verbs in -ίζειν [*-ízein*] (Greek words or morphemes are given in Greek alphabet, followed by their transliteration in Latin alphabet in square brackets). Such verbs include, e.g. Lat. *atticissāre* 'to speak Attic dialect', *citharizāre* 'to play the cithara', and *gargaridiāre* 'to gargle' from Greek ἀττικίζειν [*attikízein*], κιθαρίζειν [*kitharízein*], and γαργαρίζειν [*gargarízein*]. This topic has been much debated, especially from the point of view of the recipient language; however, the donor language and its relationship with the recipient language have not yet been adequately taken into consideration. Moreover, scholars have almost exclusively adopted the perspective of external linguistics, by taking into account particularly the social circumstances of the borrowing, and any considerations on language change have been neglected. Evidence of how Greek loanwords entered the Latin lexicon and changed its structure is given not only by Latin, but also by modern languages, such as Romance languages, English, and German, whose lexicon was influenced by that of Latin. The borrowing process considered here not only changed the lexical inventory of Latin, but also gave birth to a new way of creating verbs, which became highly productive in Romance languages.

The aims of this paper are both to account for the lexical and structural influence of Greek on Latin and to contribute to the debate on language contact and its relation with language change, from the point of view of the interplay between external and internal factors (see Chamoreau & Goury 2012; Chamoreau & Léglise 2012; 2013; De Smet et al. 2013). The structure of this paper is as follows: in §2 I present the main topics investigated by scholars and put forward some suggestions based on methodological grounds; in §3 I illustrate the syntactico-semantic values of the verb forms examined here in Greek, Latin, and Romance languages, with the aim of accounting for the paths of borrowing; §4 is dedicated to a discussion of the effects of language contact on language change, and §5 to concluding remarks.

2 An overview of previous studies and some methodological remarks

The subject discussed here has attracted the interest of many scholars, particularly specialists of Latin taking a sociolinguistic perspective. The main topics hitherto investigated are: (a) the morpho-phonemic adaptation of loan verbs in Latin, and their integration within the Latin morpho-phonemic system; (b) the morpho-lexical types of Latin verbs, e.g. loans and calques, in order to determine the degree of their independence towards the donor language; (c) the syntactic and semantic functions of Latin verbs in *-issāre*/*-izāre*/*-idiāre*; and (d) the cultural paths of borrowing. These topics are briefly discussed in §2.1.

2.1 A brief discussion of the literature

As far as morpho-phonemic shapes are concerned, Latin verbs are characterised by three derivational suffixes, *-iss*(*āre*), *-iz*(*āre*), and *-idi*(*āre*), which have been explained as follows (cf. particularly Arena 1965; Mignot 1969: 330–339; Biville 1990: 99–136). The first one (*-issāre*) is a diatopic variant of verb forms borrowed from the Doric Greek of Great Greece: forms such as (Doric) Greek σαλπίσσειν [*salpíssein*] ‘to sound the trumpet’ and λακτίσσειν [*laktíssein*] ‘to kick with the foot’ attested in Heraclides of Taranto and corresponding to the (Attic) Greek σαλπίζειν [*salpízein*] and λακτίζειν [*laktízein*] give evidence of the pronunciation [ts] of the Greek consonant <ζ> and are assumed to be the sources for Latin verbs in *-iss*(*āre*). The second shape of the suffix (*-izāre*) is the normalised form, which occurs in Latin since the grapheme <z> [z] was introduced into the Latin alphabet in 81 BCE. The third one (*-idiāre*) is a diastratic variant of *-izāre* that presumably reflected the popular pronunciation [dz] of Latin <z>, foreshadowing the phonemic changes in Romance languages (for more details, see Tronci 2015). The suffixes *-issāre* and *-idiāre* did not spread as much as *-izāre* in the Latin lexicon because of diachronic and diastratic constraints: *-issāre* was only used in Early Latin and then disappeared, while *-idiāre* could not occur in literary texts because of its popular and spoken-language nuance. In Latin texts, there are very few verbs in *-idiāre*: according to Cockburn (2012), only three types (*catomidiāre* ‘to strike on the shoulders’, *lactidiāre* ‘to strike with foot’, and *gargaridiāre* ‘to gargle’) are attested, but some verbs in *-izāre* also have forms in *-idiāre* as their diastratic variants, e.g. *baptidiāre* alongside *baptizāre*, and *exorcidiāre* alongside *exorcizāre*.

As for the morphological classification of words, scholars recognise the existence of four types: loans, calques, pseudo-calques (or hybrids), and autonomous Latin formations, along a scale of both greater independence from the model and progressive integration within the Latin system (cf. Funck 1886; Dardano 2008). This classification refers to the traditional sociolinguistic studies on modern languages (e.g. Haugen 1950; Weinreich 1953; Deroy 1956). I give here examples illustrating the four types (meanings of the Greek verb forms that are not present in the Latin counterparts are given in parentheses): Lat. *atticissāre* 'to speak the Attic dialect' is a loan from Greek ἀττικίζειν [*attikízein*] 'to speak Attic (/to side with the Athenians)', Lat. *graecissāre* 'to speak Greek' is a calque on Greek ἑλληνίζειν [*hellēnízein*] 'to speak Greek (/to make Greek)', Lat. *moechissāre* 'to commit adultery with' is a hybrid formation, created on Lat. *moechus* 'adulterer' (loanword from Greek μοιχός [*moikhós*] 'adulterer'), and Lat. *trullissāre* 'to plaster' is an autonomous formation from the Latin word *trulla* 'drawing tool'. According to Dardano (2008: 54), Latin loanwords in *-issāre*/*-izāre*/*-idiāre* can be classified as both cultural and core borrowings, which are defined by Myers-Scotton (2006: 212, 215) as "words that fill gaps in the recipient language's store of words because they stand for objects or concepts new to the language's culture" and "words that duplicate elements that the recipient language already has in its word store", respectively. The former are loanwords pertaining to the technical domains of Christian religion, medicine, and architecture, whilst the latter have been borrowed because of their prestige or foreign allure. The morphological integration of these verbs within the Latin lexicon was probably favoured by the co-existence of another class of Greek loanwords, that of the nouns in *-ismus*/*-ista*, such as *atticismus* 'Atticism' (*atticissāre*), *gargarismus* 'a gargle' (*gargaridiāre*), *citharista* 'a player on the cithara' (*citharizāre* 'to play the cithara'), *euangelista* 'an evangelist' (*euangelizāre* 'to evangelise'), and so on (see André 1971: 64–65 and Dardano 2008: 56–57). They were borrowed from Greek nouns in -ισμός/-ιστής [*-ismós*/*-istḗs*], which were morpho-lexically related to the verbs in -ίζειν [*-ízein*] within the Greek system (for examples, see Necker & Tronci 2012; 2017).

From the point of view of syntax and semantics, both Greek and Latin verbs have unpredictable values; the same lexical item can occur in very different syntactic structures with very different semantic values, e.g. Greek ξενίζειν [*ksenízein*] '(a) to receive someone as a guest, (b) to be a stranger, to speak with a foreign accent' (see §3.1). One semantic classification of Latin verbs (cf. Leumann 1948; Dardano 2008; Cockburn

2012) resembles that of Ancient Greek verbs (cf. Schmoll 1955). Three classes are traditionally recognised: (1) *Faktitiva*, i.e. verbs of doing/making, such as *moechissāre* 'to commit adultery with' and *martyrizāre* 'to make somebody a martyr'; (2) *Instrumentativa*, i.e. verbs denoting the conventional action performed using the instrument designated in the stem, such as *citharizāre* 'to play the cithara' and *trullissāre* 'to plaster'; (3) *Zustandsverba*, i.e. stative verbs, such as *martyrizāre* 'to be a martyr' and *graecissāre* 'to speak Greek'. A great part of this latter class is constituted by the so-called *Imitativa* (i.e. imitative verbs), which have both proper and common nouns as lexical bases, and whose basic meaning may be 'to behave like x' (and, by extension, 'to speak like x', 'to dress like x', and so on): illustrated by verbs like *patrissāre* 'to behave like a father, to play the father', *bētizāre* lit. 'to behave like a Swiss chard', and *lentulizāre* 'to imitate Lentulus, to play the Lentulus', it is one of the most productive types. This classification is, however, too rigid and interpretation-oriented to provide a satisfactory account of the semantic and syntactic variability of verbs (see §3.2).

In Latin literature, verb forms in *-issāre*/*-izāre*/*-idiāre* occur principally in Plautus' comedies, in Christian literature (translations and commentaries of the Bible, works of the Church Fathers), and in Late Latin technical treatises, but they are not found in texts written during the Classical period, or modelled on Classical Latin (on the notion of Classical Latin, see Clackson 2011a). Scholars have therefore suggested that these verbs were perceived by Latin speakers as foreign-sounding words, and that they were only used by authors who wished to make an explicit reference to the Greek language, literature, and culture (cf. Biville 1990; Cockburn 2012). Plautus made reference to Greek and used Greek loans to claim that he was Greek and that the Attic comedy was the model for his works. In Christian literature, translations of sacred books and religious traditions had to be as close as possible to the original text, and new concepts and practices compelled translators to introduce loans from Greek into Latin (e.g. *baptizāre* 'to baptise', *anathematizāre* 'to anathematise, to curse', *euangelizāre* 'to preach/to evangelise', *iudaizāre* 'to live in the Jewish manner', *scandalizāre* 'to cause to stumble'). Late Latin technical treatises were also mostly translated from Greek (cf. Fruyt 2011: 151), especially those dealing with medicine, and they are characterised by many technical loanwords (e.g. *elleborizāre* 'to poultice with hellebore', *sinapizāre* 'to poultice with mustard') and hybrids (e.g. *clysterizāre* 'to apply a clyster', *cauterizāre* 'to burn with a hot iron, to brand').

In summary, many important results have been obtained by scholars in understanding how Latin verbs arose, as loans from Greek or as Latin autonomous formations. However, questions have been left unanswered concerning, on the one hand, the interplay between external and internal factors in the dynamics of Latin language change and, on the other hand, the interface between sociolinguistics and diachronic linguistics, i.e. the relationship between the diastratic, diamesic, and diaphasic dimensions of variation and linguistic change. By diastratic I refer to variation across social classes or groups (e.g. educated vs. uneducated), by diamesic to variation across the medium of communication (e.g. written vs. spoken), and by diaphasic to variation in degrees of formality (depending on, e.g. communicative situation, interlocutor, and topic).

2.2 Questions, aims, and method of this study

Within the traditional views illustrated above, Latin verbs appear to be some sort of butterfly collection: there is a list of ca. 140 types that are mostly *hapax legómena* (tokens with a frequency of 1) or, in a small number of cases, verbs with many tokens. The latter, however, occur in translations, commentaries, and quotations of biblical texts, i.e. in Latin texts that closely reproduce the original Greek versions. Because of the strong dependence of the Latin occurrences on their Greek sources, it is not feasible to explain the linguistic and sociolinguistic values of Latin occurrences without taking into account their Greek sources and models. In order to capture the linguistic values of Latin occurrences and, in this way, the social meaning of the language contact that yielded them, I adopt a comparative approach and investigate both Ancient Greek and Latin, following the idea of "conspiracy" between contact-induced phenomena and internal linguistic change (Chamoreau & Léglise 2012: 9).

In order to distinguish the roles of internal and external factors in linguistic change, Johanson (2002: 286) claimed that "[i]nternal factors should probably not be regarded as "reasons" or "forces", but rather as inherent proclivities or tendencies". According to Johanson (2002: 286), "[c]ases in which the data seem to admit both external and internal motivations […] are often instances of externally motivated internal tendencies". This perspective recalls that suggested by Roman Jakobson (1990 [1938]: 208) and quoted by Weinreich (1953: 25), that a language "accepts foreign structural elements only when they correspond to its tendencies of development". Within this perspective, the emergence of the

Latin verbs investigated here can be seen as a contact-induced language change: in the Latin system, the structural conditions for creating these verbs existed, but their birth was also made possible by the long-lasting contact with the Greek language (cf. Kaimio 1979; Biville 1990; 1992; 2002; Dubuisson 1992a; 1992b; Adams 2003; for an overview, see Tronci 2015). By structural conditions, I mean the capacity of the Latin language to create new verbs by deriving them from nouns, adjectives or verbs through suffixation (e.g. causative verbs in *-fic*(*āre*) formed from both nouns and adjectives, and frequentative verbs in *-it*(*āre*) formed from verbs). Even though Latin did not have recourse to derivational strategies as much as Ancient Greek or Sanskrit in forming new verbs, the existence of these Latin derivational patterns and the ability of speakers, who presumably were mostly bilingual, to analyse the verbs borrowed from Greek worked together in triggering the new Latin derivational process.

This study accounts for the occurrences of Latin verbs by describing them from both external and internal points of view and by comparing them with their lexical and textual Greek sources. Within this comparative perspective, Latin verbs in *-issāre*/*-izāre*/*-idiāre* are not regarded as "merely lexical" items of the recipient language, but rather as the outcomes of the convergence between Greek and Latin, which was favoured by the long-lasting contact between the two languages within the Roman society – in accordance with the idea that "[g]rammatical replication is most likely to occur if there is a large degree of intensive and extensive bilingualism among the speakers of the replica language and if contact extends over a longer period of time" (Heine & Kuteva 2005: 13). In spite of the convergence between Greek and Latin, the verbs in *-issāre*/*-izāre*/*-idiāre* did not have an even distribution in Latin texts: as often noted, they were prevented from occurring in Classical Latin texts. This uneven distribution is the result of multiple factors, which concern the relationship between the two languages within Roman society and over time, involving diastratic, diaphasic, and diamesic variations.

3 Ancient Greek, Latin, and the paths of borrowing (with an appendix on Romance languages)

In this section, I provide an account for the paths of lexical borrowing, through an in-depth examination of the Greek source verbs and the Latin loans, from both internal and external points of view. My investigation on Greek verbs (§3.1) is restricted primarily to the internal structure of words

(i.e. the relationship between form and function), their occurrences in the texts and their spreading into the lexicon. As far as Latin is concerned (§3.2), external factors are particularly taken into account. I discuss language contact and bilingualism as triggers of lexical borrowing, the role of the Greek language within Roman society, the sources of loanwords, and the literary models for the new Latin formations. The issue of the outcomes of Latin verbs in *-issāre*/*-izāre*/*-idiāre* in the Romance languages is also touched upon (§3.3) because of their relevance for understanding the sociolinguistic status of these verbs in the Latin language and society.

3.1 Ancient Greek verbs in -ίζειν [*-ízein*]: lexicon, syntax, and semantics

The derivational suffix -ίζ(ειν) [*-íz*(*ein*)] arose in Ancient Greek from a morphological reanalysis of verb forms such as ἐλπίζειν [*elpízein*] 'to hope' and συρίζειν [*surízein*] 'to play the pipe', where -ίζ(ειν) [*-íz*(*ein*)] may be diachronically explained as due to the phonetic coalescence of the nominal stem ending in a stop (either dental, ἐλπιδ- [*elpid-*], or velar, συριγγ- [*surigg-*]) and the inherited verbal suffix *-je*/*o-*: Ancient Greek -δ-/-γ- [*-d-*/*-g-*] + *-j-* > -ζ- [*-z-*] [z]. Once this phonetic coalescence made the two morphemes indistinguishable, the verbs were synchronically reinterpreted as ἐλπ-ίζειν [*elp-íz*(*ein*)] and συρ-ίζειν [*sur-íz*(*ein*)], and thus arose the verbal suffix -ίζ(ειν) [*-íz*(*ein*)], which was very productive during the history of Greek, starting from Homeric poems until the Hellenistic period and beyond (e.g. Schmoll 1955). Evidence of this productivity is provided by both the morphological and the syntactico-semantic levels of analysis. As for morphology, nominal, adjectival, verbal, adverbial stems, and also proper nouns, numerals, and idioms could combine with -ίζειν [*-ízein*]. As for syntax and semantics, the syntactic values of these verbs are so variable that they are unpredictable out of context and their meanings are therefore strongly dependent on the context. The same lexeme can show very different values in different contexts and the verb ξενίζειν [*ksenízein*] provides a good example of this. The two meanings of the verb 'to receive someone as a guest' and 'to be a stranger, to speak with a foreign accent' (cf. Liddell et al. 1996 [1843], s.v.) are due to two different lexical-syntactic processes, as the transitive vs. intransitive syntax of the verb clearly shows. These two meanings reflect the two different but related meanings 'guest' and 'foreign' of the lexical basis ξένος [*ksénos*], but a

verb ξενίζειν [*ksenízein*] with the meaning 'to be a guest, to behave like a guest' would not a priori be excluded.

Besides the lexicalised verbs, e.g. πολεμίζειν [*polemízein*] 'to wage war, to fight', ὀργίζειν [*orgízein*] 'to make angry, to irritate', ὀνειδίζειν [*oneidízein*] 'to make a reproach', νομίζειν [*nomízein*] 'to use customarily, to practise', and κομίζειν [*komízein*] 'to take care of, to provide for' (see Tronci 2010; 2012 for a lexico-syntactic analysis), evidence of the extraordinary productivity of -ίζειν [*-ízein*] is provided by occasional new formations, as the following examples show. The examples include the original text in Greek and Latin, the transliteration for the Greek, and the translation into English. Translations are taken from the *Cambridge Edition of Greek and Latin Classics and the World English Bible*, with some adjustments. Original texts and abbreviations of Greek and Latin works are available on the website of the Perseus Project.[1]

(1) οὐκ ἔστιν **ἀλωπεκίζειν**,
οὐδ' ἀμφοτέροισι γίγνεσθαι φίλον. (Aristoph. *Wasps* 1241–1242)
ouk éstin ***alōpekízein***,
oud'amphotéroisi gígnesthai phílon.
'I know not how to play the fox, nor call myself the friend of both parties.'

(2) εἰ γὰρ μὴ νύμφαι γε θεαὶ Βάκιν ἐξαπάτασκον,
μηδὲ Βάκις θνητούς, μηδ' αὖ νύμφαι Βάκιν αὐτὸν–
ἐξώλης ἀπόλοι', εἰ μὴ παύσαιο **βακίζων**. (Aristoph. *Peace* 1070–1072)
ei gàr mḕ númphai ge theaì Bákin eksapátaskon,
mēdè Bákis thnētoús, mēd'aû númphai Bákin autòn–
eksṓlēs apóloi', ei mḕ paúsaio ***bakízōn***.
'Nay, nay! if only the Nymphs had not fooled Bacis, and Bacis mortal men; and if the Nymphs had not tricked Bacis a second time…
May the plague seize you, if you don't stop Bacizing!'

(3) πάσας δ' ὑμῖν φωνὰς ἰεὶς καὶ ψάλλων καὶ πτερυγίζων
καὶ **λυδίζων** καὶ ψηνίζων καὶ βαπτόμενος βατραχειοῖς
οὐκ ἐξήρκεσεν, […] (Aristoph. *Kn*. 522–524)
pásas d'humîn phōnàs hieìs kaì psállōn kaì pterugízōn
kaì ***ludízōn*** *kaì psēnízōn kaì baptómenos batrakheioîs*
ouk eksḗrkesen, […]
'he had sung in all keys, played the lyre and fluttered wings; he turned into a Lydian and even into a gnat, daubed himself with green to become a frog. All in vain!'

[1] See www.perseus.tufts.edu.

These verbs are all formed on a nominal stem (both common and proper nouns) and occur in intransitive structures. Although their semantic values appear to be different from each other ('to play the fox', 'to prophesy like Bacis', and 'to turn into a Lydian'), they can all be reduced to an essential value 'to play the x' (where 'x' is the lexical basis), and thus to 'to play the fox', 'to play the Bacis', and 'to play the Lydian'. The processes of antonomasia and its opposite, archetypal name, involve the nouns 'fox', 'Bacis' and 'Lydian' and, then, give birth to the verbs. In antonomasia, the noun replacing 'x' is functionally a proper noun, although it is categorially a common noun (e.g. ἀλωπεκίζειν [*alōpekízein*] 'to play the fox'). In archetypal name, the noun replacing 'x' is both categorially and functionally a proper noun (e.g. βακίζειν [*bakízein*] 'to play the Bacis'). In both, the nouns are functionally proper nouns, but their creation processes are different. In the case of primary common nouns (e.g. ἀλώπηξ [*alṓpēks*] 'fox' in ἀλωπεκίζειν [*alōpekízein*] 'to play the fox'), their denotative value is lost and their connotative value becomes relevant: in the case of the noun ἀλώπηξ [*alṓpēks*] 'fox', its connotative value 'to be sly' becomes the commonplace associated with the new proper noun that occurs in the derived verb (e.g. ἀλωπεκίζειν [*alōpekízein*] 'to play the fox', that is, 'to be as sly as a fox'). Regarding primary proper nouns (e.g. Βάκις [*Bákis*] 'Bacis' in βακίζειν [*bakízein*] 'to play the Bacis'), one should assume two functional processes: firstly, the proper noun functionally becomes a common noun, and, secondly, the common noun functionally becomes a new proper noun. Given that common nouns are characterised by a denotative value, the common noun arising from the proper noun Βάκις [*Bákis*] 'Bacis' denotes a prophet, Βάκις [*Bákis*] being a prophet. That is, the common noun Βάκις [*Bákis*] (e.g. 'to be a Βάκις [*Bákis*]', that is, 'to be a prophet') denotes whatever 'prophet' and does not necessarily refer to the prophet called Βάκις [*Bákis*]. Once the proper noun functionally becomes a common noun, antonomasia can occur and, thus, a new proper noun arises (see La Fauci 2007; 2008 for the "proper to common to proper noun" cycle).

The meanings of the verbs in -ίζειν [*-ízein*] are sometimes difficult to understand, because the connotations to which they are related depend on encyclopedic knowledge, which is common among people sharing the same culture but can vary from one culture to another. In other words, it is essential to know that Bacis is a prophet to understand the meaning of the verb βακίζειν [*bakízein*] 'to play the Bacis', and hence 'to prophesy like Bacis'. Likewise, some ethnonymic verbs, such as ἑλληνίζειν [*hellēnízein*]

'to speak Greek' in (4), refer to the language ('to play the x by speaking', hence 'to speak x'), as it is one of the most important signs of ethnic identity, but other ethnonymic verbs have different connotations, e.g. κρητίζειν [*krētízein*] 'to play the Cretan', that is, 'to lie', in (5):

(4) Ἕλλην μέν ἐστι καὶ **ἑλληνίζει**; (Plat. *Meno* 82b)
Héllēn mén esti kaì ***hellēnízei****?*
'He is a Greek, I suppose, and speaks Greek?'

(5) πρὸς Κρῆτα δὲ ἄρα, τὸ τοῦ λόγου, **κρητίζων** ἠγνόει τὸν Φαρνάβαζον. (Plut. *Lys.* 20.2)
pròs Krêta dè ára, tò toû lógou, ***krētízōn*** *ēgnóei tòn Pharnábazon.*
'but in thus 'playing the Cretan against a Cretan', as the saying is, he misjudged Pharnabazus.'

Besides the intransitive ethnonymic type, there is also the transitive ethnonymic one:

(6) ἀποδρὰς γὰρ ἐς τὴν γωνίαν τυρὸν πολὺν
κατεσικέλιζε κἀνέπλητ' ἐν τῷ σκότῳ. (Aristoph. *Wasps* 910–911)
apodràs gàr es tèn gōnían turòn polùn
katesikélize *kanéplēt'en tôi skótōi.*
'He sought refuge in a dark corner to glutton on a big Sicilian cheese, with which he sated his hunger.'

Verbs such as κατασικελίζειν [*katasikelízein*] 'to play the Sicilian, by doing/dealing with (something)' are a sort of double predication, implying antonomasia ('to play the x') on a lexical-syntactic level and a two-argument structure on a syntactic level. This type of verb can be seen as a transitivization of the type in (3).

In addition to the antonomasia type, -ίζειν [*-ízein*] is productive in creating verbs from whatever lexical basis and with no matter what syntactico-semantic value. The verbs can occur in either transitive or intransitive structures. As for the transitive ones, besides the factitive/causative meaning (e.g. βεμβικίζειν [*bembikízein*] 'let someone be a top' (from βέμβιξ [*bémbiks*] 'top') in (7), many other kinds of relationship between lexical basis and derived verb are possible, e.g. γαστρίζειν [*gastrízein*] 'to burst the bell' (from γαστήρ [*gastḗr*] 'bell') in (8), and σιφωνίζειν [*siphōnízein*] 'to draw off with a siphon' (from σίφων [*síphōn*] 'siphon') in (9). As for the intransitive ones, there is a broad variety of meanings: evidence for this is given in (10–12), in which

παιωνίζειν [*paiōnízein*] ‘to chant the Paean’ (from παιών/παιάν [*paiṓn*/*paián*] ‘Paean’), παππίζειν [*pappízein*] ‘to say papa’ (from πάππας [*páppas*] ‘papa’), and γρυλίζειν [*grulízein*] ‘to grunt’ (from γρῦλος [*grûlos*] ‘swine’), respectively, are attested. All examples are taken from Aristophanes.

(7) φέρε νυν, ἡμεῖς αὐτοῖς ὀλίγον ξυγχωρήσωμεν ἅπαντες,
ἵν’ ἐφ’ ἡσυχίας ἡμῶν πρόσθεν **βεμβικίζωσιν** ἑαυτούς. (Aristoph. *Wasps* 1516–1517)
phére nun, hēmeîs autoîs olígon ksugkhōrḗsōmen hápantes,
hín’eph’hēsukhías hēmôn prósthen ***bembikízōsin*** *heautoú.*
‘Let us stand out of the way a little, so that they may twirl at their ease.’

(8) ὦ πόλις καὶ δῆμ’, ὑφ’ οἵων θηρίων **γαστρίζομαι**. (Aristoph. *Kn.* 273)
ô pólis kaì dêm’, huph’hoíōn thēríōn ***gastrízomai****.*
‘Oh citizens! oh people! see how these brutes are bursting my belly.’

(9) ἐπεὶ τάδ’ οὐκ εἴρηχ’, ὁρᾷς, ὡς στλεγγίδας λαβοῦσαι
ἔπειτα **σιφωνίζομεν** τὸν οἶνον. (Aristoph. *Thes*. 556–557)
epeì tád’ouk eírēkh’, horâis, hōs stleggídas laboûsai
épeita ***siphōnízomen*** *tòn onion.*
‘Have I said how we use the hollow handles of our brooms to draw up wine?’

(10) εὐφημεῖν χρὴ καὶ στόμα κλείειν καὶ μαρτυριῶν ἀπέχεσθαι,
καὶ τὰ δικαστήρια συγκλείειν, οἷς ἡ πόλις ἥδε γέγηθεν,
ἐπὶ καιναῖσιν δ’ εὐτυχίαισιν **παιωνίζειν** τὸ θέατρον. (Aristoph. *Kn*. 1316–1318)
euphēmeîn khrὲ kaì stóma kleíein kaì marturiôn apékhesthai
kaì tà dikastḗria sugkleíein, hoîs hē pólis hḗde gégēthen,
epì kainaîsin d’eutukhíaisin ***paiōnízein*** *tò théatron.*
‘Maintain a holy silence! Keep your mouths from utterance! call no more witnesses; close these tribunals, which are the delight of this city, and gather at the theater to chant the Paean of thanksgiving to the gods for a fresh favour.’

(11) […] καὶ πρῶτα μὲν ἡ θυγάτηρ με
ἀπονίζῃ καὶ τὼ πόδ’ ἀλείφῃ καὶ προσκύψασα φιλήσῃ
καὶ **παππίζουσ’** ἅμα τῇ γλώττῃ τὸ τριώβολον ἐκκαλαμᾶται
(Aristoph. *Wasps* 607–609)
[…] *kaì prôta mèn hē thugátēr me*
aponízēi kaì tṑ pód’aleíphēi kaì proskúpsasa philḗsēi
kaì ***pappízous’*** *háma têi glṓttēi tò triṓbolon ekkalamâtai*
‘first my daughter bathes me, anoints my feet, stoops to kiss me and, while she is calling me “her dearest father”, fishes out my triobolus with her tongue’

(12) ὑμεῖς δὲ **γρυλίζοντες** ὑπὸ φιληδίας
ἕπεσθε μητρί, χοῖροι. (Aristoph. *Pl.* 307–308)
*humeîs dè **grulízontes** hupò philēdías*
hépesthe mētrí, khoîroi.
'And do you too grunt with joy and follow your mother, my little pigs.'

A great amount of productivity is also evident when new concepts and tools need to be named, e.g. in Christian religion, philosophy, and medicine. In the context of religion, new meanings are attributed to already existing verbs, cf. βαπτίζειν [*baptízein*] 'to baptise' instead of 'to dip' in (13) and δαιμονίζεσθαι [*daimonízesthai*] 'to be possessed by a demon' instead of 'to be deified' in (14), and new verbs are created as well, e.g. σκανδαλίζειν [*skandalízein*] 'to give offence or scandal to anyone' in (15) and γαμίζειν [*gamízein*] 'to give a daughter in marriage' in (16):

(13) ἐγὼ μὲν ὑμᾶς **βαπτίζω** ἐν ὕδατι εἰς μετάνοιαν· (*Matthew* 3.11)
*egṑ mèn humâs **baptízō** en húdati eis metánoian;*
'I indeed baptise you in water for repentance.'

(14) ὀψίας δὲ γενομένης, ὅτε ἔδυ ὁ ἥλιος, ἔφερον πρὸς αὐτὸν πάντας τοὺς κακῶς ἔχοντας καὶ τοὺς **δαιμονιζομένους**· (*Mark* 1.32)
*opsías dè genoménēs, hóte édu ho hḗlios, épheron pròs autòn pántas toùs kakôs ékhontas kaì toùs **daimonizoménous**;*
'At evening, when the sun had set, they brought to him all who were sick, and those who were possessed by demons.'

(15) εἰ δὲ ὁ ὀφθαλμός σου ὁ δεξιὸς **σκανδαλίζει** σε, ἔξελε αὐτὸν (*Matthew* 5.29)
*ei dè ho ophthalmós sou ho deksiòs **skandalízei** se, éksele autòn*
'if your right eye causes you to stumble, pluck it out and throw it away from you'

(16) ὅταν γὰρ ἐκ νεκρῶν ἀναστῶσιν, οὔτε γαμοῦσιν οὔτε **γαμίζονται**, ἀλλ' εἰσὶν ὡς ἄγγελοι ἐν τοῖς οὐρανοῖς. (*Mark* 12.25)
*hótan gàr ek nekrôn anastôsin, oúte gamoûsin oúte **gamízontai**, all'eisìn hōs ággeloi en toîs ouranoîs.*
'For when they will rise from the dead, they neither marry, nor are given in marriage, but are like angels in heaven.'

All these types of verbs occur in Latin, as both loanwords and new Latin formations, and constitute a consistent type within the Latin verbs in *-issāre*/*-izāre*/*-idiāre*.

3.2 Latin verbs in *-issāre*/*-izāre*/*-idiāre*: texts, morphological patterns, and syntactico-semantic values

The first occurrences of Latin verbs in *-issāre*/*-izāre*/*-idiāre* date back to the 3rd century BCE, and are found in Plautus' comedies and in other Early Latin texts (fragments of Accius', Pacuvius', and Lucilius' works). There are loanwords and calques, both of them reflecting a strong relationship with their Greek model, but there are also new Latin formations. By creating these words, Plautus was allegedly referring to Aristophanes' pieces, and his puns.

(17) *idne tú mirare, si* ***patrissat*** *filius?* (Pl. *Ps.* 442)
'Are you surprised at it, if the son does take after the father?'

(18) *atque adeo hoc argumentum* ***graecissat****, tamen*
non ***atticissat****, verum* ***sicilicissitat*** (Pl. *Men*. 11–12)
'and, in fact, this subject is a Greek one; still, it is not an Attic, but a Sicilian one'

(19) *mi vir, unde hoc ornatu advenis?*
quid fecisti scipione aut quod habuisti pallium?
in adulterio, dum ***moechissat*** *Casinam, credo perdidit.* (Pl. *Cas.* 974–976)
'My good man, whence come you in this guise? What have you done with your walking-stick, or how disposed of the cloak you had?
While he was playing his loving pranks with Casina, he lost it, I fancy.'

The syntactico-semantic features of these forms are clearly similar to those of the Greek verbs above. The shape of Lat. *patrissāre* in (17) recalls that of Greek παππίζειν [*pappízein*] in (11) and πατερίζειν [*paterízein*], but their values are different: Lat. *patrissāre* 'to play the father' belongs to the antonomasia type, while Greek παππίζειν [*pappízein*] 'to say papa' and πατερίζειν [*paterízein*] 'to call someone father' do not. Although they are traditionally interpreted as 'to speak Greek/the dialect of Attica/the dialect of Sicily', respectively, Lat. *graecissāre*, *atticissāre*, and *sicilicissitāre* in (18) also belong to the antonomasia type, like the ethnonymic verb form of ἑλληνίζειν [*hellēnízein*] in (4). Finally, the verb form of *moechissāre* in (19) is transitive, like that of κατασικελίζειν [*katasikelízein*] in (6), so both of them imply a transitivization of the antonomasia type.

Antonomasia-type verbs have had a longstanding durability in diachrony and across languages: they entered Latin through lexical borrowing, and were subsequently inherited by Romance languages,

through regular language change. In both Greek and Italian, the antonomasia type is very productive (cf. Tronci 2015). In Latin, however, the antonomasia type does not appear to be so productive, with the exception of Plautus' creations, and some other later occurrences like the following (see Clackson 2011b: 507):

(20) *ponit assidue et pro stulto 'baceolum apud pullum pulleiaceum' et pro Cerrito 'uacerrosum' et 'uapide' se habere pro male et '**betizare**' pro languere, quod uulgo 'lachanizare' dicitur*. (Suet. *Aug.* 87.2)
'He [Augustus] constantly puts *baceolus* for *stultus*, *pullejaceus* for *pullus*, *vacerrosus* for *cerritus*, *vapide se habere* for *male*, and *betizare* for *languere*, which is commonly called *lachanizare*.'

This passage from Suetonius is sociolinguistically interesting for several reasons. First of all, it speaks to the fact that verbs in *-issāre*/*-izāre*/*-idiāre* (and particularly the antonomasia type) were not only used by slaves in Plautus' comedies but also by Roman people belonging to the ruling class (here, the emperor Augustus). The differences in using these forms depend on sociolinguistic and diachronic factors. In Plautus' performances, the characters belonged to people of the lower classes, being in most cases Greek slaves, so their speech reproduced that of the lower-class and Graecising people who lived in Rome in the 3rd century BCE. Two centuries later, the Roman ruling class was also Graecised, as evidenced by the passage in (20). According to Suetonius, the emperor Augustus used the verb *bētizāre* instead of the Latin verb *languēre*, or the vernacular loanword *lachanizāre*. Thanks to the metalinguistic remarks of Suetonius, the quasi-synonym Latin verbs *bētizāre*, *languēre*, and *lachanizāre* can find their places within the diasystem of the Latin language. The verb *bētizāre* is the Latin form corresponding to the Greek loanword *lachanizāre*, by means of morpheme induction: they have the same Graecising suffix *-izāre*, but the former has a Latin lexical basis (*bēta* 'beet'), whilst the latter is a Greek loanword. Both forms were considered as belonging to the lower-level language and therefore were avoided in written language, in which only *languēre* was accepted. As regards the verb *bētizāre*, its creation presupposes the ability of the speaker to both analyse the Greek loanword *lachanizāre* (*lachan-izāre*, from Greek λάχανον [*lákhanon*] 'garden herbs, vegetables') and create the new lexeme *bētizāre* by replacing the Greek lexical basis λάχανον [*lákhanon*] with the Latin one *bēta*. According to Suetonius, the emperor preferred to use the Latin form *bētizāre* rather than the Greek loanword *lachanizāre*. The reasons for his lexical choice are not

given by Suetonius; however, it may be suggested that either the Latin form sounded more expressive than the Greek loanword, or the Greek loanword was considered vernacular Latin, and therefore unsuitable for the emperor (cf. Tronci 2017 for more details).

In sum, Latin played a very important role in ensuring the continuation of the lexical process occurring from Ancient Greek to modern languages. For this reason, it may be assumed that many loanwords and Latin new formations in *-issāre*/*-izāre*/*-idiāre* existed in spoken and non-literary Latin, even though they did not find a place in literary texts because of their foreign sounding and low-class nuance. The development of these forms in Romance languages is, however, consistent with the hypothesis of their alleged high frequency in spoken Latin.

It is traditionally recognised by scholars (cf. Cockburn 2010; 2012) that most of the verb forms in *-issāre*/*-izāre*/*-idiāre* were created when the Bible was translated from Greek into Latin, and when the clergymen and theologians started to write commentaries on it (see Burton 2011: 489). These verbs are mostly loanwords from Ancient Greek, have many occurrences in Latin, and should be considered technical words, as they are words that Latin borrowed from Greek to refer to Christian religious practices (see Mohrmann 1961). Some Latin examples and their Greek correspondences are given below, in (a) and (b), respectively; they are all extracted from the Bible.

(21) a. *si tu cum Iudaeus sis gentiliter et non iudaice vivis quomodo gentes cogis* ***iudaizare****?* (*Galatians* 2.14)
b. εἰ σὺ ᾽Ιουδαῖος ὑπάρχων ἐθνικῶς καὶ οὐχὶ ᾽Ιουδαϊκῶς ζῇς, πῶς τὰ ἔθνη ἀναγκάζεις **᾽Ιουδαΐζειν**;
ei sù Ioudaîos hupárkhōn ethnikôs kaì oukhì Ioudaikôs zêis, pôs tà éthnē anagkázeis ***Ioudaïzein****?*
‘If you, being a Jew, live as the Gentiles do, and not as the Jews do, why do you compel the Gentiles to live as the Jews do?’

(22) a. ***thesaurizat*** *et ignorat cui congregabit ea*. (*Psalm* 38.7)
b. **θησαυρίζει** καὶ οὐ γινώσκει τίνι συνάξει αὐτά.
thēsaurízei *kaì ou ginṓskei tíni sunáksei autá.*
‘He heaps up, and doesn’t know who shall gather.’

(23) a. *praemium enim tibi bonum* ***thesaurizas*** *in die necessitatis*; (*Tobit* 4.11)
b. θέμα γὰρ ἀγαθὸν **θησαυρίζεις** σεαυτῷ εἰς ἡμέραν ἀνάγκης·
théma gàr agathòn ***thēsaurízeis*** *seautôi eis hēméran anágkēs;*
‘So you will be laying up a good treasure for yourself against the day of necessity.’

(24) a. *ille autem coepit **anathematizare** et iurare quia nescio hominem istum quem dicitis*. (*Mark* 14.70)
b. ὁ δὲ ἤρξατο **ἀναθεματίζειν** καὶ ὀμνύναι ὅτι Οὐκ οἶδα τὸν ἄνθρωπον τοῦτον ὃν λέγετε.
*ho dè érksato **anathematízein** kaì omnúnai hóti Ouk oîda tòn ánthrōpon toûton hòn légete.*
'But he began to curse, and to swear, "I don't know this man of whom you speak!"'

(25) a. *et adplicuit ad eos et **anathematizavit** eos* (*1 Maccabees* 5.5)
b. καὶ παρενέβαλεν ἐπ' αὐτοὺς καὶ **ἀνεθεμάτισεν** αὐτοὺς
*kaì parenébalen ep' autoùs kaì **anethemátisen** autoùs*
'and he marshaled his troops against them and anathematised them'

(26) a. *et dixit illis angelus nolite timere ecce enim **evangelizo** vobis gaudium magnum*. (*Luke* 2.10)
b. καὶ εἶπεν αὐτοῖς ὁ ἄγγελος, Μὴ φοβεῖσθε, ἰδοὺ γὰρ **εὐαγγελίζομαι** ὑμῖν χαρὰν μεγάλην.
*kaì eîpen autoîs ho ággelos, Mḕ phobeîsthe, idoù gàr **euaggelízomai** humîn kharàn megálēn.*
'The angel said to them, "Don't be afraid, for behold, I bring you good news of great joy".'

(27) a. *multa quidem et alia exhortans **evangelizabat** populum*.
b. πολλὰ μὲν οὖν καὶ ἕτερα παρακαλῶν **εὐηγγελίζετο** τὸν λαόν· (*Luke* 3.18)
*pollà mèn oûn kaì hétera parakalôn **euēggelízeto** tòn laón;*
'Then with many other exhortations he preached good news to the people.'

These words spread rapidly in both the commentaries on the Bible and the Christian liturgies, which were addressed to clergymen and theologians, and, for the latter, also to the public. The fact that the Latin language was preserved during centuries in the Christian liturgy helped these words enter Romance languages as loans, as *-izzare*, *-iser*, and *-izar* types in Italian, French, and Spanish, respectively.

The syntactico-semantic values of these verbs are variable, as (21–27) show. Close to the antonomasia type, here exemplified in (21) by *iudaizāre* (see also *christianizāre* 'to profess Christianity', *barbarizāre* 'to play the barbarian, to speak a barbarian language', *epicurizāre* 'to play the Epicurus, to behave like Epicurus', *admartyrizāre* and *martyrizāre* 'to play the martyr, to be a martyr'), there are verbs like *thesaurizāre* 'to treasure up, to store' and *anathematizāre* 'to curse, to devote to evil', which are intransitive in (22) and (24), and transitive in (23) and (25), as well as

euangelizāre ‘to proclaim glad tidings, to proclaim as glad tidings’, which has two different transitive structures illustrated in (26) and (27) (see also *baptizāre* ‘to baptise’). As far as their Greek correspondences are concerned (θησαυρίζειν [*thēsaurízein*] ‘to treasure up’, ἀναθεματίζειν [*anathematízein*] ‘to curse, to devote to evil’, εὐαγγελίζεσθαι [*euaggelízesthai*] ‘to bring good news, to preach’, and also βαπτίζειν [*baptízein*] ‘to baptise’), I suggest an analysis taking the internal point of view. If we assume that the intransitive type arose first, and that a transitivization process happened afterwards, alongside lexicalization, it is reasonable to think that the intransitive type is related to either light verb constructions or cognate object constructions. For instance, ἀναθεματίζειν [*anathematízein*] (τινί [*tiní*]: intransitive) can be related to ἀνάθεμα ἀνατιθέναι τινί [*anáthema anatithénai tiní*] ‘to put a curse on someone’, whilst ἀναθεματίζειν [*anathematízein*] (τινά [*tiná*]: transitive) probably arose from transitivization. This internal analysis cannot be applied to Latin occurrences, since they are loanwords and, for this reason, lack any relationship with Latin lexical items and syntactic structures. However, because of the widespread bilingualism of Roman society, which concerned both upper and lower classes, Latin speakers were able to analyse loanwords and reproduce their morpho-semantic models in creating calques or genuine Latin formations, e.g. *hymnizāre* ‘to sing hymns’ (a hybrid formation derived from the loan *hymnus*, Gr. ὕμνος [*húmnos*] ‘hymn’).

Let us now turn to the Latin verbs in *-issāre*/*-izāre*/*-idiāre* which are not borrowed or calqued from Greek, that is, verbs that are formed on Latin lexical bases without any Greek counterpart. According to Mignot (1969: 330), less than twenty *types* formed on Latin lexical bases are attested during the history of Latin, which means that this derivational process was not productive in Latin. Cockburn (2012) pointed out that most of these verbs are attested in Late Latin. This is an interesting fact because it confirms the idea that Classical Latin authors acted as a sort of filter with respect to the Graecising *-issāre*/*-izāre*/*-idiāre* verbs, by avoiding them in their texts.

In Early and Classical Latin, only six verbs formed on Latin lexical bases are found, i.e. *exuibrissāre* ‘to shake the voice (in singing)’ from the Latin verb *uibrāre* ‘to shake’; *patrissāre* ‘to take after one’s father’ from the noun *pater, patris* ‘father’; *matrissāre* ‘to become like one’s mother’ from the noun *mater, matris* ‘mother’; *certissāre* ‘to inform’ from the

adjective *certus* ‘fixed’; *purpurissāre*[2] ‘to paint with purple’ from the noun *purpura* ‘purple’; and *trullissāre* ‘to plaster’ from the noun *trulla* ‘dipper’. The new Latin formations derive from both nouns and adjectives, similarly to the loanwords: for instance, Lat. *cyathissāre* ‘to fill a *cyathus*’, borrowed from the denominal Greek verb κυαθίζειν [*kuathízein*] (lexical basis: the noun κύαθος [*kúathos*] ‘small ladle’) or Lat. *malacissāre* ‘to render soft’, borrowed from Greek μαλακίζειν [*malakízein*] (lexical basis: the adjective μαλακός [*malakós*] ‘soft, sweet’). Even though deverbal formations are very rare in Latin, as are Greek deverbal verbs in -ίζειν [-*ízein*], some examples exist, e.g. *uibrissāre* and *exuibrissāre* ‘to shake the voice (in singing)’. These six Latin formations in -*issāre* do not seem to have been productive in language use: *patrissāre*, for instance, is attested three times in Plautus, and *purpurissāre* is attested once in Plautus and then disappeared.

With respect to Classical Latin, a turnaround occurred during the first two centuries CE: fifteen new *types* of verbs in -*issāre*/-*izāre*/-*idiāre* are attested in that period (Cockburn 2012: 162). Most of them are loanwords which show not only the lexical relationship with the donor language but also its inflectional morphology, e.g. the Greek-like participles *aerizousa* which designates a kind of precious stone (from Gr. ἀερίζειν [*aerízein*] ‘to resemble air’), *amethystizontas* ‘resembling the amethyst in color’ (from an unattested Gr. verb *ἀμεθυστίζειν [*amethustízein*] formed on ἀμέθυστος [*améthustos*] ‘amethyst’), and *astragalizontes* ‘the dice-players’ (from Gr. ἀστραγαλίζειν [*astragalízein*] ‘to play with dice’). All these forms occur in the *Naturalis Historia* by Pliny the Elder, who is well-known for his Graecising language (see Cockburn 2012: 167–179). However, other genuine Latin forms occurred in that period, e.g. the verbs attested by Suetonius, *bombizāre* ‘to buzz (said of bees)’ from the noun *bombus* ‘deep sound’ (which is a loanword from Gr. βόμβος [*bómbos*]), and *tetrissitāre* ‘to cackle’, which presumably refers to the model of the Gr. verbs τρίζειν [*trízein*], τρύζειν [*trúzein*], and τερετίζειν [*teretízein*], all of them designating some human or bird sounds, whilst bearing the Latin frequentative suffix -*it*(*āre*). In the Latin language of that period, there are also some interesting forms attested in the *Satyricon* by Petronius. Besides the loanword *catomidiāre* ‘to strike on the shoulders’ (from Gr. κατωμίζειν

[2] Some scholars have suggested that the verb derives from the noun *purpurissum* ‘a kind of dark purple color’ (e.g. Funck 1886: 406, 413; Leumann 1948: 373; Cockburn 2012: 119–120), but I follow Biville (1990: 111), according to whom the verb is a loanword or a calque from the reconstructed Greek verb *πορφυρίσσειν [*porphurísseιn*].

[*katōmízein*]) and the hybrid formation *excatarissāre* 'to clean', which is formed by the Latin prefix *ex-* and the Greek loan καθαρίζειν [*katharízein*] 'to purify', the genuine deverbal Latin form *exopinissāre* 'to think' (from *opināri* 'to think') shows that the derivational process is morpho-lexically meaningless and serves the purpose of providing the new form with a Greek-like sound.

In summary, both loanwords/calques and genuine Latin formations appear to be comparable to their Greek counterparts, as far as both their morphological patterns and their syntactico-semantic values are concerned. The derivational pattern concerns mainly nominal and adjectival lexical bases. The derived verbs can be both transitive and intransitive, like their Greek models. The meanings of the verbs also range from the imitative type ('to behave/speak/act like x') to the causative one ('to make something x'). There is a difference, however, between the Early Latin forms and those belonging to Christian literature: the former were mainly of the antonomasia type, while the latter had a greater variety of meanings. Plautus' loanwords and new formations were considered as amusing and foreign-sounding by Latin speakers, so they were allegedly used in vernacular and spoken language. As far as Christian literature is concerned, the use of Greek loanwords was a requirement imposed by translation, more precisely by the fact that the Latin version of the Bible had to be as close as possible to the Greek source text. Latin speakers who converted to the Christian religion presumably knew the Greek language and viewed it as a feature characterising the lexicon of their religion, because of many Greek-sounding neologisms.

3.3 The evidence of Romance languages

The Latin derivational suffixes *-iss*(*āre*)/*-iz*(*āre*)/*-idi*(*āre*) gave rise to two different suffixes in most Romance languages, e.g. It. *-eggiare* and *-izzare*, Fr. *-oyer* and *-iser*, and Sp. *-ear* and *-izar*. This fact is very interesting for my research perspective, because it can be considered as a consequence of the different sociolinguistic spaces of Latin verbs in *-issāre*/*-izāre*/*-idiāre*. Here, I limit myself to giving some general insights into this topic, my main issue being to determine the dynamics of language contact vs. language change in Latin.

The two series of suffixes in the three Romance languages arose from two different diachronic paths: regular morpho-phonetic change (It. *-eggiare*, Fr. *-oyer*, Sp. *-ear*) and reanalysis through lexical borrowing

from Latin (It. *-izzare*, Fr. *-iser*, Sp. *-izar*). The latter suffixes are still productive in all three languages, with both nouns and adjectives as lexical bases, e.g. It. *memorizzare* 'to memorise' from the noun *memoria* 'memory' and *civilizzare* 'to civilise' from the adjective *civile* 'civil'; Fr. *étatiser* 'to nationalise' from the noun *état* 'state, nation' and *européaniser* 'to Europeanise' from the adjective *européen* 'European'; Sp. *carbonizar* 'to carbonise' from the noun *carbón* 'carbon' and *legalizar* 'to legalise' from the adjective *legal* 'legal'. Most verbs occur in transitive structures and are semantically oriented towards factitive and causative values. However, there are also some intransitive forms, e.g. It. *ironizzare*, Fr. *ironiser*, and Sp. *ironizar* 'to be ironic' which are presumably learned words. In French, some new formations in *-iser* belong to the imitative type, e.g. *gidiser* 'to resemble (the style of) André Gide'. In Spanish, the suffix *-izar* became more productive in the 20th century (Bergua Cavero 2004: 183). However, even in past centuries forms in *-izar* existed which were borrowed from Latin or created by reanalysis. Alvar & Pottier (1983: §311) argue that in the 17th century "there are as many verbs in *-izo* as one desires to form" (my translation). Rainer (1993: 592–596) distinguishes two types of derived verbs in *-izar* in Modern Spanish: deadjectival verbs with a factitive meaning, e.g. *culpabilizar* 'to make somebody feel guilty' (from the adj. *culpable* 'guilty'), and *castellanizar* 'to make something/somebody Castilian' (from the adj. *castellano* 'Castilian'); and denominal verbs, whose meanings range from 'to make something/somebody x', e.g. *pulverizar* 'to pulverise', to 'to treat somebody as x', e.g. *tiranizar* 'to tyrannise' (cf. also Pharies 2002: 373–374). Verbs derived from proper nouns also belong to this group, e.g. *galvanizar* 'to galvanise' and *pasteurizar* 'to pasteurise', which are common to other European languages, e.g. Fr. *galvaniser* and *pasteuriser*, It. *galvanizzare* and *pastorizzare*, and German *galvanisieren* and *pasteurisieren*, and can be considered to be pan-European words. As far as Italian verbs in *-izzare* are concerned, their high productivity depends on their occurrence in both common language (e.g. *polemizzare* 'to argue about', from the noun *polemica* 'argument', *fraternizzare* 'to fraternise', from the adjective *fraterno* 'fraternal') and specialised languages (e.g. *scannerizzare* 'to scan', from the Engl. loanword *scanner*, *digitalizzare* 'to digitise' from the adjective *digitale* 'digital'), according to Dardano (2009: 47–48, 54–55; cf. also Tekavčić 1980: 87–88).

Unlike the verbs formed with the learned suffixes It. *-izzare*, Fr. *-iser*, and Sp. *-izar*, which are productive in all three languages, the verbs

suffixed by Fr. *-oyer* (e.g. *foudroyer* 'to strike by lightning' derived from the noun *foudre* 'lightning', *rougeoyer* 'to glow red' derived from the adjective *rouge* 'red') were productive in past centuries but are not anymore.[3] According to Pharies (2002: 184), occurrences such as It. *guerreggiare*, Fr. *guerroyer*, and Sp. *guerrear* (and Cat. *guerrejar*) 'to war' or It. *verdeggiare*, Fr. *verdoyer*, and Sp. *verdear* (and Cat. *verdejar*) 'to become green' evidence the high productivity of the suffix *-idiāre* in Late Latin (see Tronci 2015 for more details on *-eggiare* in Ancient Italian). In Spanish, the morpho-phonetic change from Lat. *-izāre* has given the suffix *-ear* which is productive as both denominal (e.g. *pasear* 'to go for a walk', derived from the noun *paso* 'walk') and deadjectival suffix (*blanquear* 'to glow white', from the adjective *blanco* 'white'). Spanish also preserves a couple of words derived from the same Latin source, such as the popular inherited verb *batear* (in Catalan *batejar*) and the learned loan *bautizar* 'to baptise' (cf. Rainer 1993: 458–465; Pharies 2002: 184–186; Bergua Cavero 2004: 185). The phonetic convergence of both Latin suffixes *-idiāre* and *-igāre* into *-ear* increased even more the class of derived verbs in *-ear* (cf. Pharies 2002: 185–186; Cockburn 2013) which counts ca. 829 types in the Spanish language spoken in Chile (cf. Morales Pettorino et al. 1969).

Let us now come back to Italian verbs in *-eggiare*. They are either deadjectival or denominal, occur in transitive and intransitive structures, and carry various semantic values (cf. Tekavčić 1980: 88; Dardano 2009: 47, 53). In some cases, they have the generic factitive nuance ('to do/to make x') and can be replaced by a light verb construction containing the noun which is the lexical basis of the verb: for instance, It. *guerreggiare* 'to war' can be paraphrased by *fare la guerra*, lit. 'to make war'. In other cases, the verbs in *-eggiare* belong to the imitative type, e.g. *toscaneggiare* 'to imitate the Tuscan people' (from the ethnonym *toscano* 'Tuscan'), *fellineggiare* 'to imitate (the style of) Fellini' (from the proper noun *Fellini*). According to Dardano (2009: 47), the latter type has become very frequent in the language of newspapers in recent decades. The distribution of the verbs formed by *-izzare* and *-eggiare* in Italian is particularly interesting because the two suffixes are both productive and specialise in two different functions. Combined with ethnonyms and proper nouns as lexical bases, *-eggiare*, i.e. the suffix deriving from the vernacular Latin *-idiāre* through regular morpho-phonetic change, specialises in the

[3] See www.cnrtl.fr/definition/-oyer.

antonomasia-type function (e.g. *americaneggiare* 'to behave like an American'); on the other hand, *-izzare*, that is the suffix deriving from upper-class Latin *-izāre* through borrowing, specialises in the transitive/causative function (e.g. *americanizzare* 'to Americanise'). Both the form and function of the two Italian suffixes mirror the two different sociolinguistic spaces of Latin verbs (see §4). The labels "vernacular" and "upper-class" Latin are not just related to the social classes of speakers. It is well known that the language of the Bible could not be too popular because it was used to deal with religion and to speak of sacred subjects. As pointed out by Burton (2011: 487), one should assume that "[m]any features of biblical Latin […] are probably best identified as belonging to a sort of post-Classical *koiné* rather than to any definitely stigmatised register". Thus, "upper-class" and "vernacular" Latin are not absolute labels, but relative to one another. That means that the verbs in *-issāre*/*-izāre*/*-idiāre* occurring in Christian literature reflect a "higher" level of language than those occurring in Plautus' comedies, and this is not surprising.

4 Lexical borrowing and language change: explaining their relationship

The picture drawn above does not exhaust the subject but is sufficient to capture some regularities of the linguistic change that took place in the Latin language as a consequence of lexical borrowing. A new derivational class of verbs arose in Latin through reanalysis of borrowed items, extraction of the suffixes and their application to genuine Latin lexical bases. This class of verbs spread through Latin into Romance languages, and then, through French, into English and German. The result of these long-standing processes is that many European languages share today the derivational patterns whose common shapes are the suffixes borrowed from Greek -ίζειν [*-ízein*] into Latin and then inherited or borrowed from Lat. *-izāre* into Romance languages.

4.1 Borrowing and language change: from Greek to Latin (and to Romance languages)

First of all, it must be underlined that lexical borrowing did not involve the lexicon only: syntax and semantics were also concerned because the borrowed items were associated with syntactic and semantic values that

were formerly either unknown or expressed in a different way in Latin. As seen in §3.2, Latin verb forms in *-issāre*/*-izāre*/*-idiāre* have various semantic and syntactic values. They cannot be reduced to one type but imply different processes. Both the use of these verbs and their distribution within the texts depend on sociolinguistic variables that concern the diastratic, diaphasic, and diamesic dimensions. An important parameter to evaluate is the relationship with the Greek model, regarding both the values of the source verb and its use within the texts.

One of the most widespread values is the one found in the antonomasia type, which is so persistent across centuries that verbs in *-eggiare*, such as *catoneggiare* 'to play the Cato' (cf. Latin *lentulizāre* 'to play the Lentulus'), still exist in Italian. It is not surprising that the antonomasia type spread into Romance languages by means of a regular morpho-phonemic change: the Latin verb forms of this type belonged to spoken and popular language, namely the so-called Vulgar Latin, as appears from both their presence in Plautus' comedies, and their absence in Classical texts (on the label Vulgar Latin, see Herman 2000: 7; Adams 2013: 10–11). Among Romance languages, Italian inherited from Latin this kind of form-function relation, which became very productive in Old Italian, more than it appears to have been in Latin. From the comparison between Latin and Italian, it can be assumed that the lower productiveness of the antonomasia type in Latin is not caused by internal (systemic) constraints, it is in fact an optical illusion due to external factors, like the predominance of Classical literature, on the one hand, and the lack of popular texts, on the other hand, in our knowledge of Latin. This assumption is in line with both the (poor) evidence provided by Latin texts and the outcomes of Romance languages. Moreover, it can explain why the antonomasia type verbs are patterned on the *-eggiare* form in Italian, and why they never occurred with the *-izzare* form: their diastratic connotation in Latin correlates with their diachronic developments, in other words with the fact that they underwent the regular morpho-phonemic change and were not borrowed by Romance languages.

From Ancient Greek to Latin and from Latin to Romance languages, there exists a long-lasting persistence of some verbs (Gr. -ίζειν [*-ízein*], Lat. *-izāre*, It. *-izzare*, Fr. *-iser*, etc.), precisely those that belong to Christian literature. These verbs appear to be unchanged across languages in both form and function: the reason for this is that the religious practices and the ways they were labelled have been long-lastingly maintained across centuries and cultures. As opposed to the antonomasia-type verbs, verbs in

Christian literature did not undergo the regular morpho-phonemic change because they entered Latin and then Romance languages through the translation of the Bible and other sacred books, that is, through written texts. The written transmission of texts preserved these verb forms from morpho-phonemic and semantic change. It is interesting to note that the morphological opposition between suffixes developed by Romance languages (e.g. It. *-eggiare* vs. *-izzare*) existed as a sociolinguistic variation within the Latin system: see, for instance, the two Latin verbs *baptizāre* and *baptidiāre*.

Secondly, the study of the relationship between lexical borrowing and language change sheds new light on the social dynamics of the language and its diachrony. As we have seen, the paths through which these verbs were borrowed and spread into Latin are diverse. This fact correlates with the various sociolinguistic values of verb forms and is reflected in the form of the suffix (*-izāre* vs. *-idiāre*), in the different syntactico-semantic functions of verbs, in their distribution in literary texts, and finally in their Romance outcomes. Moreover, this sociolinguistic variation is evidence of the deep integration of the new word class within the language system as a whole, that is, within its system and diasystem. Besides the lexical entries, the inventory of Latin morphemes also increased. The new derivational suffix maintained the manifold semantic and syntactic values of the original Greek one. The difference with the Greek counterpart concerns the sociolinguistic markedness of Latin verbs in *-issāre*/*-izāre*/*-idiāre*, which is relevant not only for explaining the phonetic variability of the suffix and the uneven distribution of verbs within the Latin texts, but also for accounting for the Romance outcomes. In agreement with Matras (2007: 31), it can be claimed that "[t]here is a link between the sociolinguistic norms of a speech community, the intensity of cultural contacts, and the outcomes of structural processes of change".

4.2 Borrowing and language change: Latin phenomena and theoretical implications

In order to provide a classification of the borrowing process from Greek to Latin, I follow the five-step scale proposed by Thomason & Kaufmann (1988: 74). The phenomenon discussed here reaches the third step because it involves structural borrowing, which is defined by the assumption that "derivational suffixes may be abstracted from borrowed words and added to native vocabulary". From a synchronic point of view, this borrowing

results in a change of the Latin lexical system: a new set of derived verbs arose and, with them, a new form-function relation. Latin verbs borrowed from Ancient Greek are in fact lexical items, but they also triggered a structural change in derivational mechanisms of the Latin verb system. Lexical borrowing thus also entailed structural borrowing. Nevertheless, the categories of lexical and structural borrowing are sometimes too clear-cut: especially if the language contact involves ancient languages, the speakers are assumed to be bilingual, but their bilingualism cannot be accurately evaluated (see Moravcsik 1978: 120).

The discussion on the "borrowability" of grammatical features dates back, at least, to Whitney (1881), who claimed that "[w]hatever is more formal or structural in character remains in that degree free from the intrusion of foreign material" (quoted in Haugen 1950: 224). The idea that lexical borrowing is one of the factors triggering linguistic change, besides analogy and grammaticalization, dates back to Meillet (1958 [1905–1906]), on the topic of lexical and structural borrowing, and Meillet (1958 [1912]), on the internal factors that entail linguistic change. However, the suggestion that borrowed items or structures induce some changes in the system of the recipient language was unacceptable as it stood to scholars supporting the Structuralist paradigm, e.g. Jakobson (1990 [1938]), Weinreich (1953), and, more recently, Johanson (2002). In their opinion, borrowing is allowed to entail some changes in the recipient language only if these changes existed as internal tendencies in the recipient language itself. According to Weinreich (1953: 25), "[s]ince such latent internal tendencies, however, by definition exist even without the intervention of foreign influence, the language contact and the resulting interference could be considered to have, at best, a trigger effect, releasing or accelerating developments which mature independently". Scholars have devoted much attention to this topic during the last century (see Gardani et al. 2015 for a detailed overview). Some important aspects of the debate were pointed out by Campbell (1993), who particularly addressed the issue of the borrowability of elements between languages which are not structurally similar. Against the traditional (structuralist) opinion that borrowing requires some structural similarity between donor and recipient language, Campbell demonstrated that the universals and principles which have been proposed to account for constraints on borrowing have been denied by some studies, which display several cases of borrowing between languages that are structurally different (e.g. Finnish and American English in Campbell 1980; Pipil and Spanish in Campbell 1987). Some studies have

also shown that borrowing can be used to fill gaps in the recipient language, particularly when the languages in contact are structurally different (cf., among others, Heath 1978; Muysken 1981; Stolz & Stolz 1996). In Campbell's view, "given enough time and intensive contact, virtually anything can (ultimately) be borrowed" (1993: 103–104; cf. also Thomason & Kaufmann 1988: 14).

In the case study at stake here, the languages concerned are structurally similar, in that both of them are characterised by derivational processes in the domain of verbal morphology and are able to derive verbs from adjectives, nouns, and verbs. That said, it can be argued that the borrowing and the subsequent process of reanalysis were triggered by the long-standing and intensive contacts between Greek and Latin and the sociolinguistic status of the Greek language within Roman society.

5 Concluding remarks

In this article, I have attempted to investigate the general subject of lexical borrowing and its relationship with language change from both the synchronic and the diachronic points of view. By assuming that lexical borrowing from Ancient Greek in Latin was due to the presence of many bilingual Latin speakers, I have illustrated how Greek verb items in -ίζειν [*-ízein*] entered Latin and how Latin speakers considered them. Lexical borrowing can be the source for changes that involve the structures of language, in the lexicon as well as on other levels of linguistic analysis. The borrowing of lexical items does not just concern the lexicon, it also has an impact on morphosyntax and semantics because it implies the emergence of new form-function relations. Once the borrowed lexical items and their form-function relations are established in the language system, new formations can be patterned on them. Structural borrowing is at this point completed, and its consequence is a change in the synchronic system of the recipient language.

I also argued for an analysis of the borrowing process and borrowed words that takes into account both internal and external factors. Within this perspective, it was possible to distinguish two classes of loanwords, whose differences concern both synchronic features and diachronic outcomes. The first group of loanwords arose in Early Latin and is composed of impromptu formations, occurring particularly in the language of Plautus, who used Greek-sounding words so as to imitate the Greek language spoken by his characters. The verbs in *-issāre*/*-izāre* which date back to this

period are mostly loanwords and calques; genuine Latin new formations are very rare. For the most part, they are *hapax legómena* and belong to the imitative type. The second group of loanwords penetrated later into Latin, in the first centuries CE, through Christian literature, which was translated from Greek into Latin at that time. Even in this case, the loanwords and calques are more frequent than the new Latin formations. The reason for this is that several Greek verbs in -ίζειν [-*ízein*] attested in the Bible and other Christian texts designated notions and practices which were new for Greek thought and a fortiori for the Latin one. Because of this, they did not have correspondences in the Latin lexicon and could be translated only by means of loans. The verbs belonging to the second group occur frequently in the texts: this is an important difference with respect to the verbs belonging to the first group. They also became a sort of stamp of Latin Christian language. Through borrowing from Latin, most of these verbs spread into European modern languages, e.g. Engl. *to evangelise*, *to demonise*, *to anathematise*. The different outcomes of the two waves of Greek loanwords in Latin depend on external factors, especially the role of the Greek language within Roman society in the last two centuries BCE and the first two centuries CE, and the different Greek textual sources for Latin loans and calques. In Plautus and Early Latin texts, Greek was perceived as the language of slaves and preceptors. Plautus' characters came from the Greek *milieu* of Southern Italy, so their speeches are filled with Greek or Greek-sounding words. The new verbs in *-issāre*/*-izāre* are an instance of this tendency: by creating these verbs, Plautus made a clear reference to Aristophanes, who created many new verbs in -ίζειν [-*ízein*]. Like the latter, the verbs in *-issāre*/*-izāre* created by Plautus were short-lived: they did not resist the purism required by Classical Latin authors, who did not allow Greek-sounding words to occur in their works. In Christian literature, by contrast, the need to translate the new religious concepts and practices which were still unknown to Roman culture led translators to render the Greek verbs in -ίζειν [-*ízein*] through loans and calques which started the new lexicon of Christian religion. The high-level sociolinguistic status of this latter type is evidenced by the fact that Latin loanwords from Greek penetrated into Romance languages as learned words (e.g. It. *-izzare* verbs) and did not undergo morpho-phonetic changes, as was the case for the majority of verbs attested in Early Latin and belonging to the imitative type (e.g. It. *-eggiare* verbs).

Finally, my study corroborates the idea that the investigation of language contact should contemplate an approach that integrates internal

and external evidence, on the one hand, and synchronic variability and diachronic change, on the other hand. As I have shown, internal and external evidence converge towards parallel results. From the internal viewpoint, the high productivity of the verbs concerned here in both Greek and Romance languages allows us to suggest that Latin verbs in *-issāre*/*-izāre*/*-idiāre* were also productive, much more than Latin texts give evidence for. From the external viewpoint, the diachronic changes from Latin into Romance languages correlate with the sociolinguistic status of Latin verbs. In this case study, the sociolinguistic variation between the learned Latin suffixes *-issāre*/*-izāre*, on the one hand, and the vernacular suffix *-idiāre*, on the other hand, corresponds to the two different diachronic outcomes of Latin verbs into Romance languages, i.e. the verbs which were borrowed into It. *-izzare*, Fr. *-iser*, and Sp. *-izar*, and the verbs which morpho-phonetically developed into It. *-eggiare*, Fr. *-oyer*, and Sp. *-ear*. Latin has been shown to have been essential for the continuity of the long-standing processes of language interference and change, despite the lack of verbs in *-issāre*/*-izāre*/*-idiāre* in Classical Latin and their low productivity in the first centuries of Latin history, until Christian literature and Late Latin.

Acknowledgements

This research was carried out within the project *Multilingualism and Minority Languages in Ancient Europe* [HERA.29.015| CASSIO], funded by *Hera Joint Research Programme "Uses of the Past"*, Horizon 2020 – 649307. I would like to thank Elizabeth Rowley-Jolivet for her help with the English, the two anonymous reviewers and the editors for their helpful comments.

References

Adams, James N. 2003. *Bilingualism and the Latin language*. Cambridge: Cambridge University Press.

—— 2013. *Social variation and the Latin language*. Cambridge: Cambridge University Press.

Alvar, Manuel & Pottier, Bernard. 1983. *Morfologia histórica del español* [Spanish historical morphology]. Madrid: Gredos.

André, Jacques. 1971. *Emprunts et suffixes nominaux en latin* [Loanwords and nominal suffixes in Latin]. Genève: Librairie Droz & Librairie Minard.

Arena, Roberto. 1965. Contributi alla storia di lat. *-isso* [Contribution to the history of Lat. *-isso*]. *Helikon* 5. 97–122.

Bergua Cavero, Jorge. 2004. *Los Helenismos del Español* [Graecisms in Spanish]. Madrid: Gredos.

Biville, Frédérique. 1990. *Les emprunts du latin au grec. Approche phonétique. Tome I: Introduction et consonantisme* [Loanwords from Greek to Latin. Phonetic approach. Volume 1: Introduction and consonantism]. Louvain: Peeters.

—— 1992. Le grec parlé en latin vulgaire. Domaines lexicaux, structures linguistiques d'accueil [Uses of Greek in Vulgar Latin. Lexical domains, linguistic structures of the target language]. In Iliescu, Maria & Marxgut, Werner (eds.), *Latin vulgaire – latin tardif III. Actes du IIIème Colloque international sur le latin vulgaire et tardif, Innsbruck, 2–5 septembre 1991*) [Vulgar Latin – Late Latin III. Proceedings of the third international conference on Vulgar and Late Latin, Innsbruck, 2–5 September 1991], 25–40. Tübingen: Max Niemeyer.

—— 2002. The Graeco-Romans and Graeco-Latin: A terminological framework for cases of bilingualism. In Adams, James N. & Janse, Mark & Swain, Simon (eds.), *Bilingualism in ancient society*, 77–102. Oxford: Oxford University Press.

Burton, Philip. 2011. Christian Latin. In Clackson, James (ed.), *A companion to the Latin language*, 485–501. Malden, MA: Blackwell.

Campbell, Lyle. 1980. Towards new perspectives on American Finnish. In Paunonen, Heikki & Suojanen, Matti (eds.), *Central problems in bilingualism*, 43–54 (vol. 3, part 4 of Osmo Ikola (ed.), *Congressus Quintus Internationalis Fenno-Ugristarum*). Turku: Suomen Kielen Seura.

—— 1987. Syntactic change in Pipil. *International Journal of American Linguistics* 53. 253–280.

—— 1993. On proposed universals of grammatical borrowing. In Aertsen, Henk & Jeffers, Robert J. (eds.), *Historical linguistics 1989: Papers from the 9th International Conference on Historical Linguistics, New Brunswick, 14–18 August 1989*, 91–109. Amsterdam: John Benjamins.

Chamoreau, Claudine & Goury, Laurence (eds.). 2012. *Changement linguistique et langues en contact* [Language change and languages in contact]. Paris: CNRS Éditions.

Chamoreau, Claudine & Léglise, Isabelle (eds.). 2012. *Dynamics of contact-induced language change*. Berlin: Mouton de Gruyter.

—— (eds.). 2013. *The interplay of variation and change in contact settings*. Amsterdam: John Benjamins.

Clackson, James. 2011a. Classical Latin. In Clackson, James (ed.), *A companion to the Latin language*, 236–256. Malden, MA: Blackwell.

—— 2011b. The social dialects of Latin. In Clackson, James (ed.), *A companion to the Latin language*, 505–526. Malden, MA: Blackwell.

Cockburn, Olivia C. 2010. The use of the Latin *-izare* (*-issare, -idiare*) suffix in early Christian literature. *Revista de Estudios Latinos* (*RELat*) 10. 105–116.

—— 2012. *Los verbos latinos en -IZARE* (*-ISSARE, -IDIARE*) [Latin verbs in *-izare* (*-issare, -idiare*)]. Madrid: Universidad Autónoma de Madrid.

—— 2013. Los sufijos verbales *-ficare* e *-izare* (*-issare*, *-idiare*) y su propagación en el español [The verbal suffixes *-ficare* and *-izare* (*-issare*, *-idiare*) and their spread in Spanish]. In Casanova Herrero, Emili & Calvo Rigual, Cesáreo (eds.), *Actas del XXVI Congreso Internacional de Lingüística y de Filología Románicas* [Proceedings of the 26th International Congress of Romance Linguistics and Philology], vol. 4, 505–512. Berlin: Mouton de Gruyter.

Dardano, Paola. 2008. Contatti tra lingue nel mondo mediterraneo antico: i verbi in *-issare/-izare* del latino [Language contact in the Ancient Mediterranean world: Latin verbs in *-issare*/*-izare*]. In Orioles, Vincenzo & Toso, Fiorenzo (eds.), *Circolazioni linguistiche e culturali nello spazio mediterraneo* [Linguistic and cultural circulation in the Mediterranean area], 49–61. Genova: Le Mani Editore.

Dardano, Maurizio. 2009. *Costruire parole. La morfologia derivativa dell'italiano* [How to create words. Italian derivational morphology]. Bologna: il Mulino.

Deroy, Louis. 1956. *L'emprunt linguistique* [Linguistic borrowing]. Paris: Les Belles Lettres.

De Smet, Hendrik & Ghesquière, Lobke & Van de Velde, Freek (eds.). 2013. *On multiple source constructions in language change* [special issue]. *Studies in Language* 37(3).

Dubuisson, Michel. 1992a. Le grec à Rome à l'époque de Cicéron. Extension et qualité du bilinguisme [Greek in Rome in the time of Cicero. Extension and quality of bilingualism]. *Annales ESC* 47(1). 187–206.

—— 1992b. Le contact linguistique gréco-latin: problèmes d'interférences et d'emprunts [Linguistic contact between Greek and Latin: interferences and loanwords]. *Lalies* 10. 91–109.

Fruyt, Michèle. 2011. Latin vocabulary. In Clackson, James (ed.), *A companion to the Latin language*, 144–156. Malden, MA & Oxford, Chichester: Blackwell.

Funck, Anton. 1886. Die Verba auf *-issare* und *-izare* [Verbs in *-issare* and *-izare*]. *Archiv für lateinische Lexikographie und Grammatik* 3. 398–442.

Gardani, Francesco & Arkadiev, Peter & Amiridze, Nino. 2015. Borrowed morphology: an overview. In Gardani, Francesco & Arkadiev, Peter & Amiridze, Nino (eds.), *Borrowed morphology*, 1–26. Berlin: Mouton de Gruyter.

Haugen, Einar. 1950. The analysis of linguistic borrowing. *Language* 26(2). 210–231.

Heath, Jeffrey. 1978. *Linguistic diffusion in Arnhem Land.* Canberra: Australian Institute of Aboriginal Studies.

Heine, Bernd & Kuteva, Tania. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.

Herman, Jozsef. 2000. *Vulgar Latin*. The Pennsylvania State University Press. (Translation of Herman 1967.)

Jakobson, Roman. 1990. On the theory of phonological affinities. In Waugh, Linda R. & Monville-Burston, Monique (eds.), *On language. Roman Jakobson*, 202–213. Cambridge: Cambridge University Press. (Translation of Jakobson, Roman 1962 [1938].)

Johanson, Lars. 2002. Contact-induced change in a code-copying framework. In Jones, Mari C. & Esch, Edith (eds.), *Language change: the interplay of internal, external, and extra-linguistic factors*, 285–313. Berlin: Mouton de Gruyter.

Kaimio, Jorma. 1979. *The Romans and the Greek language*. Helsinki: Societas Scientiarum Fennica.

La Fauci, Nunzio. 2007. Antonomasie [Antonomasia]. In Cresti, Emanuela (ed.), *Prospettive nello studio del lessico italiano. Atti del IX Convegno Internazionale della SILFI, Firenze, 14-17 giugno 2006* [Perspectives on the study of the Italian lexicon. Proceedings of the 9th international congress of SILFI, Florence, June, 14–17, 2006], Vol. I, 279–283. Firenze: Firenze University Press.

—— 2008. Nomi propri, luoghi comuni [Proper names, common places]. In Arcamone, Maria Giovanna & De Camilli, Davide & Porcelli, Bruno (eds.), *Atti del XXII Congresso Internazionale di Scienze Onomastiche, Pisa, 28 agosto-4 settembre 2005* [Proceedings of the 22nd International Congress of Onomastics, Pisa, August, 28 - September, 4, 2005], vol. I, 605–613. Pisa: Edizioni ETS.

Leumann, Manu. 1948. Griechische Verben auf *-ίζειν* im Latein [Greek verbs in *-ίζειν* in Latin]. In *Mélanges de philologie, de littérature et d'histoire anciennes offerts à J. Marouzeau* [Essays in ancient philology, literature and history in honour of J. Marouzeau], 371–389. Paris: Les Belles Lettres.

Liddell, Henry G. & Scott, Robert & Jones, Henry S. 1996 [1843]. *Greek-English lexicon*. Oxford: Clarendon Press.

Matras, Yaron. 2007. The borrowability of grammatical categories. In Matras, Yaron & Sakel, Jeanette (eds.), *Grammatical borrowing in cross-linguistic perspective*, 31–74. Berlin: Mouton de Gruyter.

Meillet, Antoine. 1958 [1905–1906]. Comment les mots changent de sens [How words change their meaning]. In Meillet, Antoine (ed.), *Linguistique historique et linguistique générale* [Historical and general linguistics], 230–271. Paris: Champion.

—— 1958 [1912]. L'évolution des formes grammaticales [The evolution of grammatical forms]. In Meillet, Antoine (ed.), *Linguistique historique et linguistique générale* [Historical and general linguistics], 130–148. Paris: Champion.

Mignot, Xavier. 1969. *Les verbes dénominatifs latins* [Latin denominal verbs]. Paris: Klincksieck.

Mohrmann, Christine. 1961. *Études sur le latin des Chrétiens* [Studies on Christian Latin]. Vol. I. Roma: Edizioni di Storia e Letteratura.

Morales Pettorino, Félix & Quiroz Mejías, Oscar & Mayorga Aravena, Dora. 1969. *Los verbos en* -ear *en el español del Chile* [The verbs in *-ear* in Chilean Spanish]. Santiago: Editorial del Pacífico.

Moravcsik, Edith. 1978. Universals of language contact. In Greenberg, Joseph H. & Ferguson, Charles A. & Moravcsik, Edith A. (eds.), *Universals of human language. Volume 1: Method and theory*, 93–122. Stanford, California: Stanford University Press.

Muysken, Pieter. 1981. Halfway between Quechua and Spanish: The case for relexification. In Highfield, Arnold & Valdman, Albert (eds.), *Historicity and variation in Creole studies*, 52–78. Ann Arbor: Karoma.

Myers-Scotton, Carol. 2006. *Multiple voices: An introduction to bilingualism.* Malden, MA: Blackwell.

Necker, Heike & Tronci, Liana. 2012. Italian *-ismo* and *-ista* and Ancient Greek *-ismós* and *-istḗs*. Morphological processes and diachronic relationships. In Ralli, Angela & Booij, Geert & Scalise, Sergio & Karasimos, Athanasios (eds.), *Morphology and the architecture of grammar. Online proceedings of the Eighth Mediterranean Morphology Meeting* (*Cagliari, Italy, 14–17 September 2011*), 203–221. (http://www3.lingue.unibo.it/mmm2/wp-content/uploads/2013/09/MMM8_Cagliari_Proceedings.pdf) (Accessed 2017-09-15.)

—— 2017. Proper names and derivational morphology. Italian *-ismo*/*-ista* and Ancient Greek *-ismós*/*-isté:s*. *Studi Italiani di Linguistica Teorica e Applicata* 46(1). 41–60.

Pharies, David. 2002. *Diccionario etimológico de los sufijos españoles y de otros elementos finales* [Etymological dictionary of Spanish suffixes and other word endings]. Madrid: Gredos.

Rainer, Franz. 1993. *Spanische Wortbildungslehre* [Spanish word formation]. Tübingen: Niemeyer.

Schmoll, Hans. 1955. *Die griechischen Verba auf -ίζω* [Greek verbs in -ίζω]. Tübingen: Universität Tübingen Thesis.

Stolz, Thomas & Stolz, Christel. 1996. Funktionswortentlehnung in Mesoamerika Spanisch-amerindischer Sprachkontakt (Hispanoindiana II) [The borrowing of function words in the contact between Spanish and indigenous Mesoamerican languages (Hispanoindiana II)]. *Sprachtypologie und Universalienforschung* 49(1). 86–123.

Tekavčić, Pavao. 1980. *Grammatica storica dell'italiano. III. Lessico* [Historical grammar of Italian. III. The lexicon]. Bologna: il Mulino.

Thomason, Sarah G. & Kaufmann, Thomas. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Tronci, Liana. 2010. Funzioni, forme, categorie. Una nota su costrutti con verbi in *-ίζω* [Functions, forms, and categories. Some remarks on constructions with verbs in *-ίζω*]. In Putzu, Ignazio & Paulis, Giulio & Nieddu, Gian Franco & Cuzzolin, Pierluigi (eds.), *La morfologia del greco tra tipologia e diacronia* [Greek morphology between typology and diachrony], 495–511. Milano: FrancoAngeli.

—— 2012. Valori differenziali di costrutti con forme verbali in *-ίζω* [Differential values in constructions with verbs in *-ίζω*]. In Lorenzetti, Luca & Mancini, Marco (eds.), *Discontinuità e creolizzazione nella formazione dell'Europa linguistica* [Discontinuity and creolization in the formation of European languages], 273–289. Roma: Il Calamo.

—— 2015. Greco *-ίζω* e latino *-isso*/*-izo*/*-idio*. Note preliminari per lo studio di un caso di contatto interlinguistico [Preliminary remarks on Greek *-ίζω* and Latin *-isso*/*-izo*/*-idio* as a case-study of language contact]. In Consani, Carlo (ed.), *Contatto interlinguistico tra presente e passato* [Language contact between present and past], 173–195. Milano: LED. (http://www.ledonline.it/Il-Segno-le-Lettere/allegati/728-Contatto-Interlinguistico-Tronci.pdf) (Accessed 2017-09-15.)

—— 2017. Quelques remarques pour une reconsidération des verbes latins en *-isso*/*-izo*/*-idio* [Some remarks for a new look at Latin verbs in *-isso*/*-izo*/*-idio*]. In

Spevak, Olga (ed.), *Études de linguistique latine II* [Studies in Latin linguistics II], *Pallas* 103, 293–300.
Weinreich, Uriel. 1953. *Languages in contact. Findings and problems*. New York: Linguistic Circle of New York.
Whitney, William D. 1881. On mixture in language. *Transactions of the American Philological Association* 12. 5–26.

Contact information:

Liana Tronci
Università per Stranieri di Siena
Piazza Rosselli 27/28
I-53100 Siena
Italy
e-mail: tronci(at)unistrasi(dot)it

Detection of non-native speech in a familiar language and in an unfamiliar language

Melike Uzal
University of Helsinki

Erkki Komulainen
University of Helsinki

Mehmet Akif Kılıç
Istanbul Medeniyet University

Olli Aaltonen
University of Helsinki

Abstract

This study explores the question of whether native and non-native listeners, i.e. natives familiar with the language they are judging and non-natives who are not, manage to distinguish a foreign accent from a native accent in the speech of native speakers (NSs) and nonnative speakers (NNSs). Participants included 21 speakers (11 NSs and 10 NNSs who were native Turkish speakers) as well as two listener groups that consisted of 61 Finnish listeners (FLs), and 10 Turkish listeners (TLs) without Finnish experience. This study compares accent ratings by these two listener groups that evaluated the 21 spontaneous speech samples for foreign accent using a 9-point scale. The results showed a very significant difference between the listener groups for the NSs but no significant difference for the NNSs. The difference between the FL and the TL groups was because the FLs managed to distinguish the NSs from the NNSs, but otherwise these two listener groups exercised statistically similar ratings. Therefore, these results demonstrate that the listeners' familiarity with Finnish, the target language, hence listeners' native speaker status strongly affect ratings of foreign accents, since native listeners could distinguish the NSs, whereas non-native listeners could not. The results suggest that listeners' familiarity with the target language plays a much more

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

SKY Journal of Linguistcs 20 (2017) 109–136

profound role in accent detection than their familiarity with the accent language. Moreover, the results show that contrary to previous research, in the absence of listeners' familiarity with the target language, it is much more challenging to detect a foreign accent. The results also showed that speech rate correlated with the judgments provided by the TLs but not with the judgments provided by the FLs. This result raises the possibility that there are salient universal features of non-native speech such as speech rate that even non-native listeners unfamiliar with the language they are judging utilize while judging a foreign accent.

Keywords: L2 speech, accent detection, L2 listener, L1 listener, familiarity with the accent language, familiarity with the target language, listeners' native speaker status

1 Introduction and background

The overall rating for degree of nativeness is often termed *global foreign accent* (Major 2007). Extensive research has demonstrated that global foreign accent correlates with a number of linguistic phenomena. Some of these include segmentals (Major 1987; 2001; Flege et al. 1995; González-Bueno 1997; Riney & Takagi 1999; Munro et al. 1999; Riney et al. 2000; Bunta & Major 2004), syllable structure (Magen 1998), and prosody (e.g. Anderson-Hsieh et al. 1992; Jilka 2000). Foreign accent is found to correlate also with subsegmental information including deviances observed in voice onset time difference in stop consonants, formant frequencies and vowel durations in vowels, and suprasegmental information including deviations in prosodic phenomena such as stress, phrasing, rhythm and intonation, as well as temporal aspects of speech such as segmental length, tempo, loudness, juncture and pitch differences (Moyer 2013; Schmid & Hopp 2014). For instance, Toivola (2011: 3) found that both temporal aspects of speech, such as speech rate, the number of pauses, the duration of pauses, and the number of single deviant phonetic segments contributed to the perceived degree of non-native accent in the speech of 10 Russian L2 learners of Finnish. Likewise, Trofimovich & Baker (2006: 2) found that two suprasegmentals (the duration of pauses and speech rate) were the variables that contributed the most to the perception of foreign accent by native listeners. In addition, non-native speakers often speak more slowly than native speakers, and previous studies (e.g. Munro & Derwing 1998; 2001) have shown temporal aspect of speaking rate to be the variable that contributes to the perceived degree of accentedness; listeners rate more slowly produced speech to indicate more accent than faster speech. Moreover, as Schmid & Hopp (2014: 4) state, native listeners detect a

foreign accent from features such as disfluency markers of filled pauses and repetitions (Lennon 1990) and hesitation (Dewaele 1996). Therefore, as Flege (1981: 445) observes, it seems that native listeners base foreign accent judgment on a combination of segmental, subsegmental, and suprasegmental differences that distinguish the speech of native speakers from that of non-native speakers. Furthermore, Munro & Derwing's studies (Munro & Derwing 1995; Derwing & Munro 1997) on intelligibility have shown that poor accent ratings are associated with phonetic, phonemic and grammatical errors, as well as problems with intonation.

Scovel (1995: 170) described some very broad general paralinguistic and phonetic features (e.g. the speed or the fundamental frequency of the voice listeners hear, some voice quality settings such as low voice, phonetic data such as the speaker's degree of retroflection, tone and pitch of voice) that affect the native listeners' decisions about whether or not the voice they heard was accented. His study raised the possibility that such paralinguistic features are *universal* and that these salient paralinguistic features are very telling of accentedness. All of the previous studies mentioned above are related to the current study because it also provides information on whether foreign accent judgment made by both native listeners and non-native listeners is based on speech rate, as has been found in previous research.

In second language acquisition research, it is commonly accepted that the most reliable distinguisher of accentedness is a native speaker of the language in question (Major 2007: 540), meaning that a listener's native speaker status has a profound effect on second language (L2) perception. This is because researchers have assumed that non-native speakers cannot be reliable listeners, since most fall short of native production and competence in the L2 themselves (Major 2007: 540). Munro et al. (2006: 114), however, have challenged the view that the ratings of native listeners are more valid than those of non-native listeners. When it comes to empirical evidence, only a small number of studies (e.g. Flege 1988; Riney et al. 2005; MacKay et al. 2006; Munro et al. 2006; Kang 2008) have compared the ratings of native listeners and non-native listeners to determine the effect of listeners' native speaker status on ratings of accentedness (Munro et al. 2006: 126). Surprisingly, with the exception of Kang's study (2008), all found that the ratings of native listeners and non-native listeners familiar with the target language were quite similar. In contrast, Kang (2008: 184) found that non-native speaker (NNS) listeners were harsher than the native speaker (NS) listeners in accentedness ratings.

There are a few studies (Major 2007; Bond et al. 2008; Weber & Pöllmann 2010) that have suggested that even non-native listeners lacking familiarity with the language spoken (L2), i.e. the target language, were able to distinguish a foreign accent accurately from a native one, albeit with a lower success rate than that of native listeners. According to the above-mentioned studies, non-native listeners may have performed this successful accent detection by utilizing general markers of non-nativeness such as speech rate and sentence/utterance duration as an estimate of fluency – a clue from the speech (Bond et al. 2008: 7). For instance, in Major's (2007: 539) study, American English listeners unfamiliar with Portuguese could detect an English accent in Portuguese. Major (2007: 552) interpreted this finding to mean that native and non-native listeners have similar abilities in rating foreign accents and that their L1s and L2s do not dramatically affect the ratings. Voice quality, for example, has been suggested as a potential marker of non-nativeness, though its role in L2 production has not been thoroughly investigated yet (see e.g. Esling 2000). Articulatory effort and carefulness are other potential markers (Weber & Pöllmann 2010: 541). All of the studies mentioned above are related to the current study because it also shows whether non-native listeners lacking familiarity with the target language can distinguish a foreign accent accurately from a native one.

1.1 Aims of the present study

This study has been set with two aims. The first aim was to discover how some listener background factors such as the listeners' familiarity with the accent language and native speaker status, i.e. familiarity with the language spoken (L2)/ the target language spoken, affected listeners' perception of foreign accent. The native language of the non-native speaker is termed *accent language*. One of the aims of the study was to find out whether the listeners' familiarity with the accent language gave them any advantage of detecting accentedness. Due to this reason, none of the listeners were informed that the non-native speakers of Finnish were native speakers of Turkish. With this kind of experimental design, it is important that non-native listeners do not know beforehand that they will hear their own mother tongue as an accent language.

The second aim was to obtain a preliminary assessment on whether markers of non-nativeness are language-specific or language-independent. That is, this study aimed to explore whether markers of non-nativeness are language-independent, innate and universal, in which case some general

markers would reveal a speaker's native status regardless of the listener's language abilities, or whether they are language-specific markers, in which case only listeners who have formed a native perception of the language could identify a speaker's native speaker status. The varied findings of the studies addressing the effects of the listener's familiarity with the target language spoken on foreign accent ratings (e.g. Major 2007; Bond et al. 2008; Weber & Pöllmann 2010) were one of the motivations of the present study. All of these studies' findings on listener familiarity with the target language spoken have suggested the existence of universal general markers for non-nativeness, which in turn implies that the Turkish listeners in the present study unfamiliar with Finnish would be expected to make use of these universal general markers of non-nativeness, enabling them to detect accentedness. This study explores whether this prediction, supported also by Roy C. Major from Arizona State University (personal communication, 2016), correlates with the findings. If the findings of the current study show that non-native listeners with no familiarity in the language they judge use speech rate as a clue to judge accentedness, it will also provide further empirical proof for the existence of universal general markers of non-nativeness.

1.2 Research question and hypotheses

The study sought to discover the degree to which some listener background factors (listeners' NS status, hence familiarity with the target language spoken, and familiarity with the accent language) affect the degree of perceived accent ratings in L2 Finnish. In this study, listeners possessing varying familiarity with the target language spoken were asked to assess the degree of perceived foreign accent in Finnish spoken by native speakers and non-native speakers. One item of interest was how accurately native Turkish listeners (non-native listeners) lacking familiarity with Finnish could detect Finnish L2 speakers in Finnish speech samples. Non-native listeners unfamiliar with the target language spoken are not familiar with the native accent and therefore might not be expected to form reliable, accurate and valid perceptions of that language. Thus, it was expected (the null hypothesis) that in the absence of any familiarity with the target language spoken, the non-native listeners would be unable to identify a foreign accent reliably and accurately, even though they had excellent familiarity with the accent language of Turkish. Indeed, as Major (2007) observes, in theory the idea of asking listeners to rate foreign accents in an

unfamiliar language is strange, even ludicrous. Unlike Major's study (2007), studies by Bond et al. (2008) and Weber & Pöllmann (2010) did not have listeners with excellent familiarity in the accent language and no familiarity with the target language to allow any direct comparisons with the TLs in the present study. In line with previous research on accent detection, it was expected (the alternate hypothesis) that the TLs' excellent familiarity with the accent language (linguistic experience) might help them differentiate between native and non-native speakers of Finnish. Therefore, if the TLs in this study are successful in detecting L2 speakers of Finnish, it will be possible to argue that the TLs' excellent familiarity with the accent language enabled them to distinguish between native and non-native speakers of Finnish in a reliable manner.

Compared with previous studies, this study is perhaps the first to include listeners who had excellent familiarity with the target language (they shared the same L1) and who had no familiarity with the accent language, namely the Finnish listeners (FLs). Also, this study is perhaps the first to employ non-native listeners to judge the degree of perceived accent from spontaneous speech samples in a language unfamiliar to them. To date, only Major (2007), Bond et al. (2008) and Weber & Pöllmann (2010) have used non-native listeners unfamiliar with the language they rated; however, as for their speech sample choices, Major (2007) used read passages of varying durations (M = 22 s) and both Weber & Pöllmann (2010) and Bond et al. (2008) used read speech of sentences. In contrast, this study used spontaneous speech samples of the same duration for both NSs and NNSs.

1.3 Significance of the study

First, the study will provide knowledge of how foreign accent (in this case Finnish with a Turkish accent) is perceived by both Finnish native speakers and non-native speakers unfamiliar with Finnish, which will contribute to the accent detection literature regarding perceptual studies of accented speech. Second, since non-native speakers unfamiliar with Finnish are native speakers of Turkish, it will be further explored whether the listeners' excellent familiarity with the accent language has an effect on their accentedness ratings. In addition, a debated issue in accent-rating studies is whether the listeners' foreign accent ratings indicate something about the speech itself because they are influenced by its acoustic and phonological properties, or whether they indicate something about the listener and

therefore vary with the listeners' language experience (Weber & Pöllmann 2010). Researchers should have an understanding of the factors that figure in listeners' accentedness judgment and, in particular, how properties of speech and the characteristics of listeners influence that accentedness judgment.

2 Methodology

2.1 Speakers

The 21 adult speakers were 11 native speakers (NSs) and 10 non-native speakers (NNSs) of Finnish. All the NSs were native speakers of Finnish from the Helsinki Metropolitan Area with no knowledge of Turkish, eight female speakers and three male speakers aged 22–39 (M = 27.6). All the NNSs were bilingual NSs of Turkish from a wide variety of Turkish cities; they consisted of five female speakers and five male speakers aged 27–66 (M = 40.2). All 10 adult NNSs were either first or second-generation Turkish immigrants to Finland. To sum up, all 21 speakers resided in the Helsinki Metropolitan Area and spoke standard Finnish (Karlsson 2008).

2.2 Speech samples

The speech samples consisted of pieces of spontaneous speech on a topic chosen by the individual speaker from three options (see Appendix). The speech samples from the 21 speakers were 40-second segments extracted from 1-minute recordings. The reason for using spontaneous speech samples was that since spontaneous speech is the most authentic form of natural speech, the use of spontaneous speech as a stimulus would be the most meaningful way to learn about accentedness. A 40-second piece of spontaneous speech is a stretch long enough for accent characteristics to emerge.

2.3 Listeners

Two main listener groups participated in the study: native listeners, referred to as the Finnish listeners, (n = 61) and non-native listeners, referred to as the Turkish listeners, (n = 10) unfamiliar with Finnish. Thus, this study had two listener groups with respect to both listener familiarity with the accent language and the language being spoken (L2) dimension: the Turkish

listeners who had excellent familiarity with the accent language and no familiarity with the language being spoken (L2) and the Finnish listeners who had no familiarity with the accent language and excellent familiarity with the language being spoken (L2), Finnish being their L1. However, since the number of non-native listeners was much smaller than the number of native listeners, this resulted in more foreign accent ratings given by native listeners than non-native listeners. That is, a total of 1321 foreign accent ratings were rated (61 native listeners rating 21 spontaneous speech productions and 10 non-native listeners rating 21 spontaneous speech productions).[1]

The non-native listeners were monolingual NSs of Turkish with no familiarity in Finnish. All the non-native listeners were from Giresun, in the Black Sea region of Turkey. None of the non-native listeners had studied Finnish or had been to Finland, and none reported familiarity with Finnish (in terms of hearing or recognizing it). Thus, these Turkish non-native listeners were good examples of laymen.

2.4 Procedure

The 61 native listeners completed the rating task individually in a soundproof recording studio in the Helsinki Metropolitan Area, Finland, with a total of 21 speech samples (21 speakers × 1 spontaneous speech task recording) presented via headsets. For the 10 non-native listeners, the foreign-accent rating task was completed in a quiet room in Giresun, Turkey, again with each non-native listener individually performing the rating, with a total of 21 speech samples presented via headsets. For the native listeners, a preliminary information form and a short training session were provided before the rating began. To avoid unrelated linguistic factors affecting accent ratings, in the preliminary information form, the native listeners were instructed to ignore all non-phonological speech content and only assess foreign accent. The native listeners were also instructed to use the entire scale while rating the samples and were told to guess if they were uncertain. A short training session consisted of 30 practice speech samples of single sentences (13 from NSs, 17 from NNSs) to help familiarize the

[1] There were eight ratings with negative reaction times and other 162 premature ratings given before the end of the sample. These 170 ratings (25% of all ratings) were excluded from the analyses. All in all, due to the excluded 170 ratings, there were a total of 1321 foreign accent ratings rated (61*21*10=1491) instead of 1491.

native listeners with the rating process and the range of accents; these were not analyzed.

A 9-point scale was used for the rating of accent. The listeners were told that they would hear productions spoken by either non-native or native speakers of Finnish, but they were not informed about the proportion of native and non-native speakers. They were asked to rate each production for the degree of accent by pushing one of the nine buttons representing a scale from 1 (no foreign accent) to 9 (very strong foreign accent). The same procedure was followed with the non-native listeners as it was with the native listeners, the only difference being that the non-native listeners did not participate in the training session. This was because the express aim of the study was to discover if non-native listeners who had never heard the target language could identify a foreign accent. Due to this design, the non-native Turkish listeners did not listen to any speech samples of standard Finnish, even though many of them so requested to have a benchmark for their accent judgements. Both the native and non-native listener rating sessions consisted of one block lasting 16–25 minutes. The runs including speech samples and the speakers were randomized.

2.5 Data analysis

71 listeners (61 FLs and 10 TLs) rated 21 speech samples (11 NSs and 10 NNSs). Each listener contributed to many data-points. Consequently, the data were not independent because they came from the same listener who rated. Statistical analysis of such correlated (nested, clustered) data requires methods that can properly account for the intra-subject correlation of response measurements. If such correlation is ignored, then inferences such as statistical tests or confidence intervals can be grossly invalid. On the other hand, the use of averaged ratings (aggregated data) leads to association indices that are too high (Iversen 1991).

Generalized linear mixed models (GLMM) were formulated by Nelder & Wedderburn (1972) as a way of unifying various other statistical models. These methods are now available in statistical packages such as statistical package for the social sciences (SPSS) (IBM 2016). They handle correlated data structures (O'Dwyer & Parker 2014). They also apply to cases where non-normality of distributions exists and/or the scale is ordinal. The linear mixed model (LMM) is more restricted. It can handle nested data, but it assumes the normal distribution of residual values (Madsen & Thyregod 2010).

The nature of the data obtained in the present study required the use of LMM (because the data was correlated and nested) and GLMM (because in addition to being nested data, non-normality of distributions existed as well). Consequently, analyses of foreign accent ratings were performed using GLMM (response variable ordinal, link logit) and LMM in SPSS 24. Results were practically the same. Thus, the results here were reported from the LMM analyses, since it is better suited to present results showing means and standard deviations. This choice also affected the way the results were depicted in graphical form, namely histograms and scatterplots in this study. Scatterplots use averaged values. Regression lines (X on Y) in them show how strong the depicted associations are.

There were four groups (combinations of FL/TL and NS/NNS) to compare. This was done by analyzing six possible pairwise comparisons using Bonferroni corrected post-hoc tests. The magnitude of the mean difference (effect size) was expressed by Cohen's d (difference divided by the pooled standard deviation) (Cohen 1988; Sawilowsky 2003).

The present situation in which there were more listeners than the spontaneous speech samples presented to be rated is not usual. Due to this, reliability had to be assessed by using several reliability indices. Therefore, there were three different approaches to the reliability of the ratings. First, the mean correlation with other listeners was calculated for each listener. The mean correlation grew if ratings bore similarity. Second, all foreign accent ratings (both NS and NNS) were split into two random groups. Reliability existed if the splits correlated. Third, intra-class correlation (ICC) was calculated (Shrout & Fleiss 1979); this is often used in situations where raters are used. The obvious problem is that it gives values that are too high when the number of raters exceeds the number of the stimuli. The ratio here was quite high, 71 to 21 (3.38). Consequently, in such situations very low actual reliability can produce quite a high reliability coefficient.

Speech rates of the speakers were measured with Praat software. The syllables and silences were segmented and labeled from the recorded speech samples. Syllable and silence durations and numbers were measured automatically with a script, which measures interval durations. Later, the speech rate was calculated manually by dividing the number of syllables by the total duration of a certain speech sample, in syllables/second. The duration of silence was not used here, but the silences had to be labeled as well to find the number of syllables clearly.

3 Results

3.1 Comparison of combined speaker and listener groups

Figure 1 illustrates the relative frequency of accent ratings of the four listener-speaker combinations. Most of the FL group successfully rated the NSs as natives (84%). The mean was 1.25. In contrast, the TL group, who were unfamiliar with Finnish, rated the NSs as having a moderate degree of foreign accent according to the foreign accent rating scale (20% gave 1, M = 4.79). The number of 1s for the NNSs was far less in both listener groups: FLs= 10% (M = 5.64), TLs: = 14% (M = 5.55).

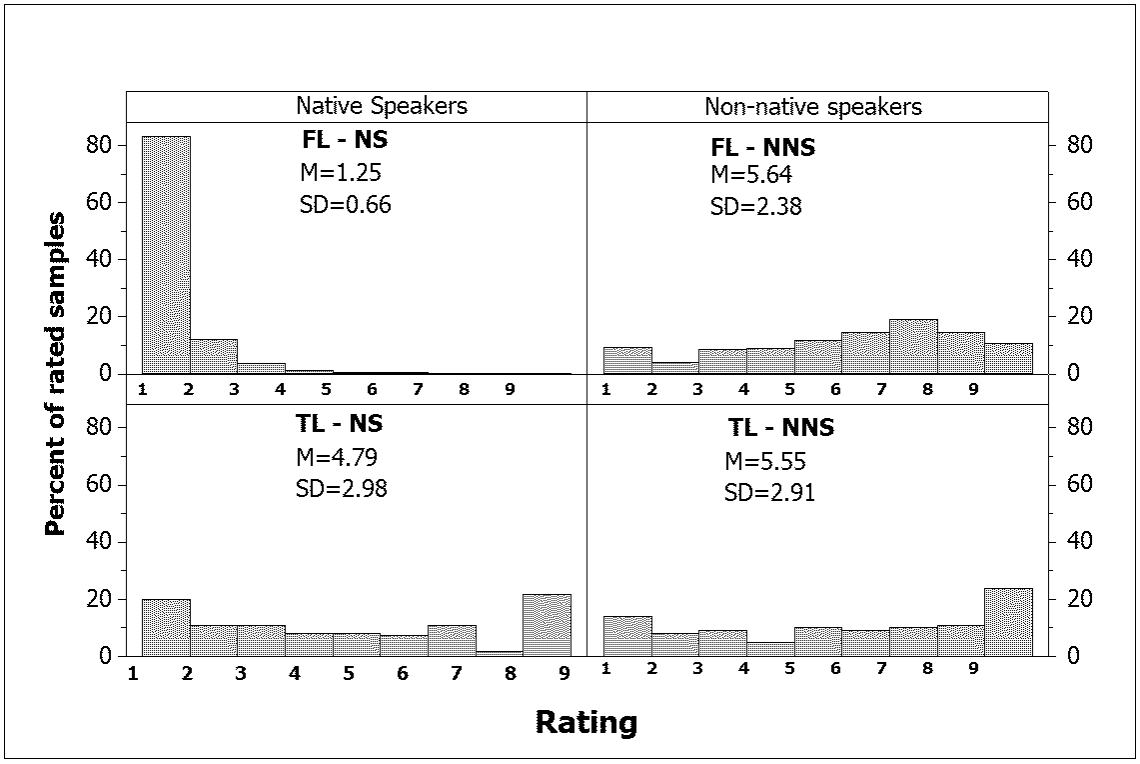

Figure 1. Rating distribution, means and standard deviations according to listener groups

The LMM analysis was carried out to evaluate the statistical significance and the magnitude of the mean differences described above. Three groups (FL-NNS, TL-NS and TL-NNS) were quite similar to each other in their accent rating means, but differed very clearly from FL-NS group. The pairwise post-hoc comparisons (LMM) of these four groups are shown in Table 1. Three pairs were not significantly different from each other and all their pairwise calculated effect sizes were small. These three groups, however, all differed very significantly statistically (p <.001) from the FL-

NS mean. These three effect sizes (*d*) ranged from 1.24 to 1.53 and could be characterized as very large (Cohen 1988; Sawilowsky 2003). The comparisons confirmed that the listeners' familiarity with Finnish, the target language, was a very strong factor in the ratings since native listeners could distinguish NSs, whereas non-native listeners could not. Figure 1 and Table 1 show that the TLs did not identify and distinguish the NSs (M = 4.79) from the NNSs (M = 5.55). There was a tendency in the right direction, though. This difference in ratings, however, was statistically not significant (p = .440), and the effect size was very small (d = .27) (Sawilowsky 2003). The LMM analysis clearly supports the fact that the non-native listeners were not able to distinguish the accent difference between the NSs and NNSs.

Table 1. Post hoc pairwise LMM analyses of four groups and their effect sizes

Post hoc pairs		**Difference**	***p****	**Cohen's *d***
FL – NS	FL - NNS	-4.39	<.001	1.53
FL – NS	TL – NS	-3.54	<.001	1.24
FL – NS	TL – NSS	-4.30	<.001	1.50
FL – NNS	NS – TL	-0.85	=.980	0.30
FL – NNS	TL – NNS	0.09	=.999	0.03
TL – NS	TL - NNS	-0.76	=.440	0.27

* = Bonferroni corrected

3.2 Reliability analysis

Table 2 shows reliability results of listener ratings carried out in three different ways. The harshest index was the mean correlation between the listeners. As Table 2 shows, TLs' reliability was low for NNSs (.31), for NSs (.14) and combined (.23). FLs' values were high except FL-NS value (.19). These NS samples were very homogenous, which explains this low value. The split procedure was perhaps the most realistic index of reliability. There was a low degree of reliability for TLs (.58 for NNS, .51 for NSs and .67 for both NSs and NNSs combined). The split total in the FL-group was .93, which is very high. All ICC values expressed high degrees of inter-rater reliability, even for TLs (ICC = .68) for NNSs, which is the lowest ICC of TLs, which is still a good inter-rater agreement. All the ICC values, however, were unrealistically high because of the nature of the data at hand. This bias typically presents itself when the number of raters is greater than the number of items to be rated. The values in mean correlations and ICC were not comparable either, since there were 61 FLs

as raters and only 10 TLs. The split method seems to be fair and dependable, i.e. FL ratings had a high reliability, whereas TL ratings had only some reliability. TL value did not reach the level (.7) which is a typical lower limit in split-type situations (Nunnally & Bernstein 1994).

Table 2. Three reliability indices

Mean correlation

	FL	TL	Total
NNS	.84	.31	.62
NS	.19	.14	.11
Total	.92	.23	.72

Split

	FL	TL	Total
NNS	.55	.58	.48
NS	.74	.51	.83
Total	.93	.67	.89

ICC

	FL	TL	Total
NNS	1.0	.68	.99
NS	1.0	.89	.99
Total	1.0	.74	.99

3.3 Speech rate and accent ratings

Due to the stimuli being of the same duration for all speakers, 40-second spontaneous speech samples, it was not possible to measure varying utterance durations and their correlation with accent ratings as in Major (2007), Bond et al. (2008) and Weber & Pöllmann (2010). Instead, speech rate was measured in Praat to see whether mean accent ratings and speech rate correlated, which would imply the use of speech rate as a universal clue to accentedness, hence non-nativeness. Multiple factors can influence foreign accent ratings, one of which might be a slower speaking rate: it might be perceived as less fluent and, therefore, more accented. To find out whether speech rate and accent ratings had any relationship with each other according to listener groups, both Pearson r analysis and LMM analyses were conducted. These two analyses were done to provide comparability among the results of previous studies because previous accent detection studies used Pearson r values to examine the effect of speech rate on accent ratings. The values of these two analyses are shown in Figure 2. In Figure 2, x-axis shows speech rates of speakers calculated in syllables/second ranging from 1.84 to 4.87, whereas y-axis shows foreign accent ratings

ranging from 1 to 9. Figure 2 shows that for the NSs one of the Pearson r correlations was significant (r = -.63, a moderately strong correlation), meaning that the slower the NSs' speech rate, the higher the accent score (the worse and harsher accent rating) was given by TLs. This same relationship in LMM analysis shows a lower value (-.22)**.** Likewise, for the NNSs, the Pearson r correlation was very significant (r = -.89), meaning that the slower the NNSs' speech rate, the higher the accent score was given again by TLs. The same relationship in LMM was likewise of lower value (-.47). Regression lines (X on Y) indicated the same thing. Lines from FL ratings were horizontal while corresponding lines from TL ratings showed a clear negative slope. Thus, LMM results, Pearson r correlations and regression lines show that TLs made use of speech rate, i.e. fluency, as a cue to native speaker status for both the NSs and the NNSs.

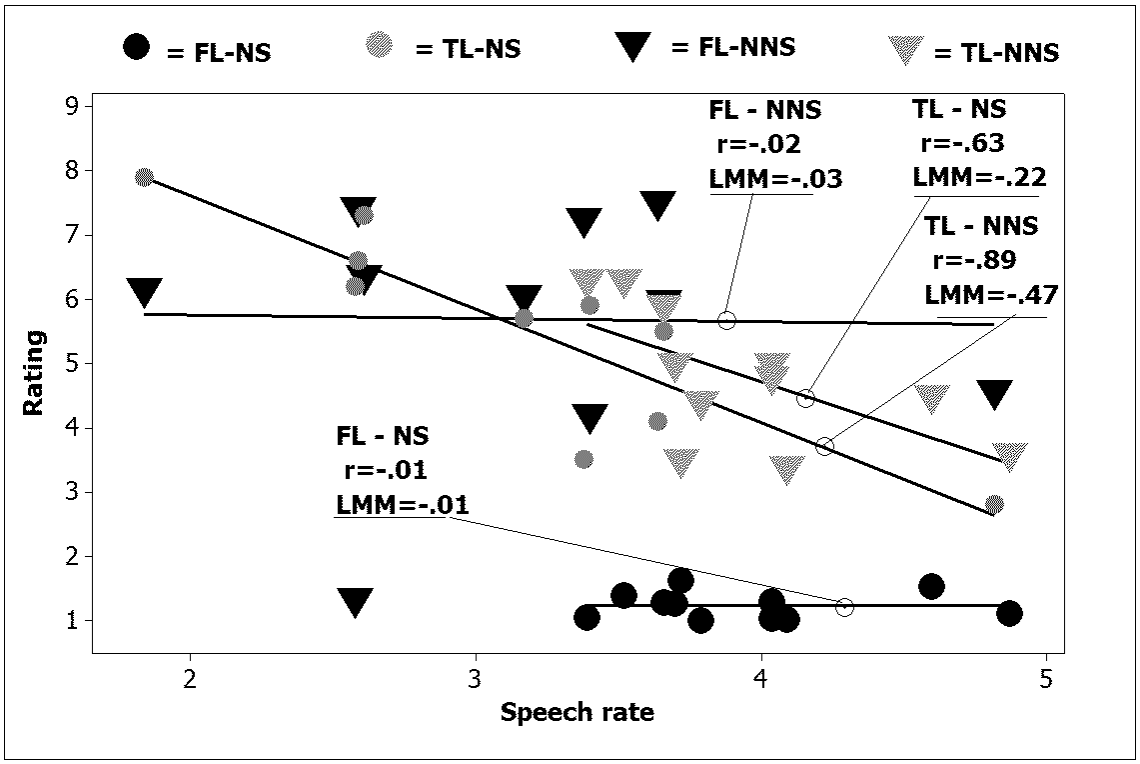

Figure 2. A scatterplot of speech rate versus foreign accent rating

As shown in Figure 2, TLs relied slightly more on speech rate when they rated NNSs when compared to NSs (LMM analysis values of -.22 for NSs vs. -.47 for NNSs). Indeed, high correlations between speech rates of both NSs and NNSs and mean accent ratings for TLs showed that they used speech rate as a basis for foreign accent judgement. Only TLs were found to make use of speech rate. They might have relied on these universal

perceptual features thinking that they do not have any linguistic information to rely on, such as their L1 or L2, as high correlations between speech rate and accent ratings made clear. It is important to note that LMM-analysis coefficients were considered more correct than Pearson r correlations because when counting correlations with mean values as was done here, a part of variance is lost when using aggregated values.

Figure 3 depicts the agreement in ratings between FLs and TLs using regression lines and correlations coefficients.

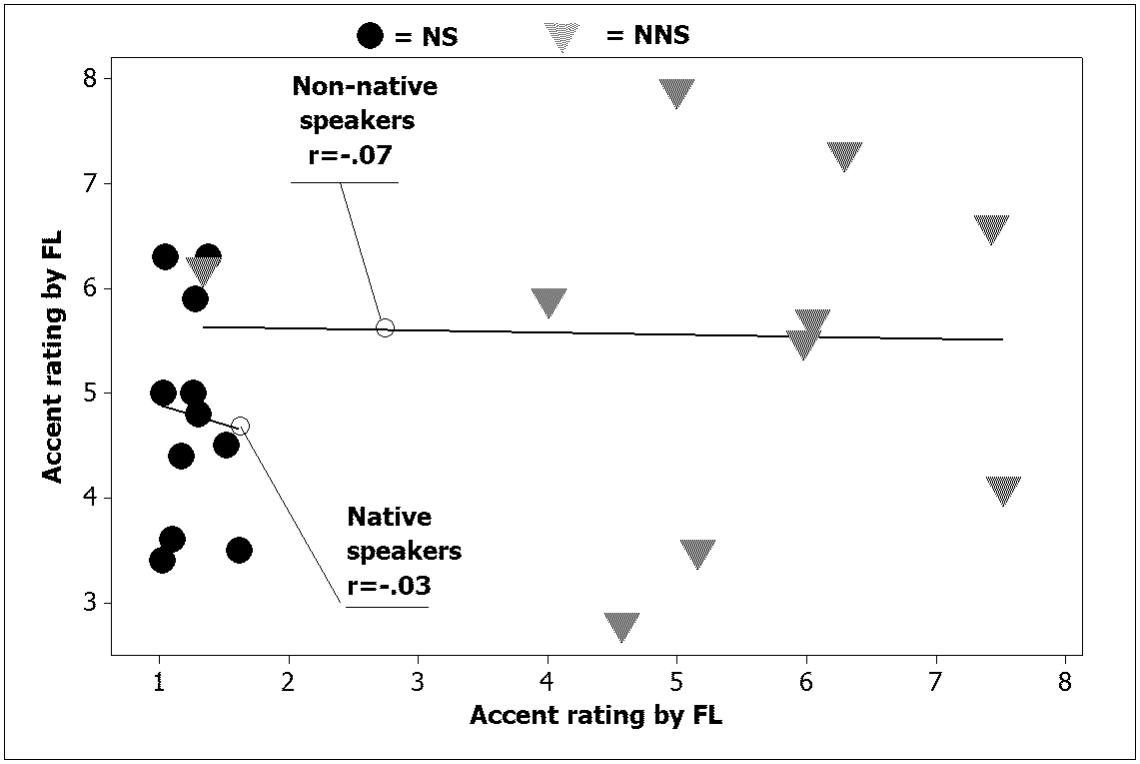

Figure 3. A scatterplot of foreign accent ratings FLs versus TLs by NSs and NNSs

In Figure 3, it can be seen immediately that in both speaker groups the agreement was close to zero. Regression lines (FL on TL) go almost horizontally or have a slight negative slope, and the corresponding correlation coefficients are very small and insignificant, i.e. when it comes to agreement of ratings between TLs and FLs the finding was that there was no agreement.

4 Discussion

4.1 The effect of listeners' familiarity with the language spoken

It was hypothesized that, in the absence of any familiarity with the target language spoken, listeners would be unable to detect accentedness accurately. The present findings confirm this prediction, showing that listeners' familiarity with the target language spoken is a prerequisite for detecting accentedness. The only listener background factor that affected the perception of foreign accent was listeners' NS status, hence listeners' familiarity with Finnish, the target language. The TLs, who had no familiarity with Finnish, were unable to distinguish between native speakers of Finnish and Turkish non-native speakers of Finnish, whereas the FLs managed to identify them extremely accurately. This finding agrees with Kang (2008: 196) that even non-native listeners familiar with the target language, in Kang's case L2 English, had different perceptions (were more stringent) than native listeners in accentedness ratings. Therefore, this study concludes that in the absence of any familiarity with the target language (L2), accent detection might be inaccurate even if the listeners' L1 is the same as that of the non-native L2 speakers; i.e. even the listeners' excellent familiarity with the accent language, Turkish, did not help them to detect accentedness correctly.

The findings of the present study on listener familiarity with the language spoken (L2) – that TLs unfamiliar with Finnish were unable to distinguish between the native and non-native speakers of Finnish – is still in line with the findings of Bond et al. (2008) and supplement the accent detection literature as well by showing that non-native listeners unfamiliar with the target language spoken indeed use speech rate as a clue while making accentedness judgments. However, in contrast to Bond et al.'s (2008) non-native listeners unfamiliar with the target language spoken whose accent ratings were above the level of chance, this study's TLs' accent ratings were not accurate. In what follows, two reasons are discussed that could account for some of the differences in rating patterns between Bond et al. (2008) and the present study. The first reason is that, in the present study, a much more representative sample of non-native listeners was used than in Bond et al. (2008). The 10 TLs in this study represented laymen, and the 61 FLs (some of them Finnish-as-a-second-language teachers and some students at the University of Helsinki) represented educated people with a general familiarity in other languages. The FLs were very reliable, accurate and successful at rating the degree of

Turkish foreign accent in the Finnish speech samples, and the TLs, who had no idea how Turkish-accented Finnish would sound, or for that matter any other accented language in Finnish, were not so reliable and accurate. The TLs had studied German and/or English (world languages) as second or foreign languages at school, but those two languages were of no use to them when judging accentedness in an unknown language, such as Finnish. Similar to the present study's FLs, in Bond et al. (2008), the American listeners were college students who represented educated people with familiarity and experience in dealing with other languages. That is, there was a difference in rating patterns between the present study and Bond et al. (2008) study because TLs represented laymen who were not experienced in dealing with other languages and thus were a more representative sample when compared to American listeners in Bond et al. (2008) who were all college students and relatively experienced in dealing with other languages.

Unmarked accents and *marked accents* may help in the identification of a foreign accent, regardless of the language. If a listener perceives a marked accent, it is language-specific, whereas if a listener perceives an unmarked accent, it is *language-independent.* Accent refers here to a universal aspect of accent in any given language. Marked and unmarked accents apply regardless of a listener's L1 background, familiarity with the language spoken (the target language, L2) and familiarity with the native accent of that language (L1), i.e. the accent language. First, it is suggested that when an accent is unmarked, it is the type of accent that prevents a listener unfamiliar with both the language being spoken (L2) and the accent language (L1) from identifying the speaker's L1 in the L2 speech. Therefore, it is not so telling and salient to cause listeners to detect the accent in question. Examples could be Finnish with a Turkish accent, as in this study, Turkish with a Finnish accent or even Polish with a Turkish accent. In fact, the findings of this study have clearly shown that even Turkish listeners were unable to detect a Turkish accent in Finnish. The Turkish listeners in this study were ordinary laymen, hence non-native listeners with no linguistic experience and linguistic sophistication, and not everyone knows how a Turkish accent (the Turkish accent supposedly being an unmarked accent universally proposed in this study) sounds in different languages or specifically in Finnish L2, as was the case in this study.

Second, it is suggested that Russian, Chinese, Japanese, French, German, Taiwanese, English, Dutch, Italian, Spanish and some others

might be examples of marked accents. In general, these languages have such strong, distinct, revealing and solid accents that any educated listener familiar with these world languages would recognize the accents of their speakers in any given language. Some accents are so telling that it is easier to identify them. According to this model, it would be easier to detect the L1 (accent language) of German-accented Turkish speakers than that of Turkish-accented Finnish speakers because of the distinct quality of the former accent. Likewise, Turkish, Finnish, Hungarian, Polish, Kyrgyz, and Swedish could be examples of unmarked accents. There is little knowledge at a global level about how these languages sound in terms of accent. Moreover, since both Turkish and Finnish are far from being world languages, listeners with no linguistic experience and linguistic sophistication might fail to detect them as accent languages simply because they have no idea what Turkish or Finnish sound like. At present, this is a matter of speculation, as our study does not attempt to uncover these possible accents but simply addresses this possibility. It was beyond the scope of this study to identify these accents, and this speculation comes as a byproduct of the findings in this study. Adjunct Professor Zinny Bond (personal communication, 2016) from Ohio State University, however, concurred with these suggestions and observed that this analysis makes good sense of both the results of the present study and those of Bond et al. (2008).

The second reason that could account for some of the differences in rating patterns between Bond et al. (2008) and the present study is that in taking the proposition of marked and unmarked accent into consideration, it is suggested that in the present study Turkish might have constituted an unmarked accent in L2 Finnish for the Turkish listeners. Likewise, the possibility of marked and unmarked accent might explain how monolingual American listeners managed to identify a native vs. non-native background with a success rate of 63%, significantly above the level of chance in Bond et al. (2008) study. These American listeners with no familiarity in either the accent language, Russian, and the target language, Latvian, might thus have had some idea of how a Russian accent (supposedly being a universally marked accent) would sound in any given language. Since Russian and Latvian belong to the same language family, intuition suggests that for American listeners it might have been relatively more difficult to distinguish between native speakers of Russian and native speakers of Latvian. Nevertheless, the American listeners in Bond et al. (2008) managed to give the highest evaluations to native Latvians and

distinguished between low and high-proficiency Russians, just as the Latvian listeners did. If monolingual Americans, however, had heard Finnish with a Turkish accent, Turkish with a Finnish accent or Swedish with a Polish accent, their success in distinguishing native speakers from non-native speakers might have been lower, perhaps only at the level of chance. A direction for further studies in accent detection research could be to find out the possible existence of marked and unmarked accents.

Scovel (1995: 175) described that accent detection success depends on native listener sophistication in languages, i.e. their overall linguistic experience and linguistic sophistication. Likewise, Scovel's argumentation (1995) that accent detection success depends on native listener sophistication in languages can be broadened to include non-native listeners as well. Therefore, it can be argued that accent detection success depends on non-native listener sophistication in languages as well exemplified with the findings of this study on non-native listeners judging accentedness in a language unfamiliar to them. In other words, in accent detection studies involving the rating of accents in unfamiliar languages, the addition of one new term is proposed. That term is *listener familiarity with language accents.* This term is the same as Scovel's (1995) final third stage of identification in accent recognition process in which, depending on the overall linguistic experience and linguistic sophistication of the native listeners, they may identify the native language of the accented voices. The only difference is that this term applies to all types of listeners regardless of their native speaker status whereas the term used by Scovel (1995) only applies to native listeners.

4.2 The effect of listener familiarity with the accent language

The findings in this study showed that even excellent familiarity with the accent language did not afford listeners any advantage in reliably detecting accentedness in the absence of familiarity with the target language spoken. The TLs, who had no familiarity with the target language, Finnish, but who had excellent familiarity with the accent language, Turkish (they shared the same L1, i.e. the advantage of sharing an L1 language background with the L2 speaker) did not identify foreignness successfully. Given that this study produced a negative finding with respect to listener familiarity with the accent language, Turkish, its argumentation of stating that listener familiarity with the accent language does not necessarily cause listeners to rate accentedness reliably comes from a position of weakness because it

produced a negative finding. An absence of evidence is not evidence of absence. The failure to produce similar positive findings using different language combinations does not refute previous research and Major's (2007) findings on listener familiarity with the accent language. The first author of this study discussed the findings of the present study with Roy C. Major (personal communication, 2016), and he found them unexpected but observed that this discussion makes good sense of both the findings in the present study and those of his and previous research. Although the findings in this study are unexpected, they are nevertheless compatible with findings from previous studies (Major 2007; Bond et al. 2008; Weber & Pöllmann 2010) on listener familiarity with the accent language and provide fresh insights into the issue. This is because the study findings fill the gap in the accent detection literature by showing that even non-native listeners with the potential advantage of excellent familiarity with the accent language can fail to identify non-native speakers, in the absence of familiarity with the target language. By showing this, the present study also supplements the findings of previous research and adds new pieces to the puzzle of how accent detection occurs.

While it might seem counterintuitive that listener familiarity with the accent language failed to enhance their ability to detect foreignness, it should be noted that previous studies also varied on whether they found a correlation between familiarity with the accent language and accent ratings (e.g. Major 2007; Bond et al. 2008 for studies finding such an effect, Munro et al. 2010 for no effect, and Weber & Pöllmann 2010 for a study that failed to find such a strong effect). It is suggested that there are several reasons for the findings of the present study that the listeners' excellent familiarity with the accent language conferred no advantage in the absence of any familiarity with the target language. First, it is proposed that the phonological properties of particular language pairs (in this case Turkish and Finnish) might make accent detection more difficult. This first suggestion creates the space for the second suggestion: if a language has a marked accent, as in the study case of Russian by Bond et al. (2008), it might be easier to detect it. Likewise, the possibility of Turkish being an unmarked accent, at least on a global level, might have made its detection more difficult for the TLs, who did not realize that it was in fact their own mother tongue which was foreign in the L2 speech of Finnish. Since the TLs had never heard Finnish before it sounded completely foreign, incomprehensible and unmarked.

4.3 Possible cues for accentedness: listeners' rating behavior

The findings of this study showed that there was a statistically significant difference between the ratings of the TLs and those of the FLs. Not only did the TLs fail to distinguish between the NSs and NNSs of Finnish, but their agreement with FL group ratings was also bad or nonexistent, as TLs gave foreign accent ratings relying on speech rate. The findings also showed that foreign accent ratings of FL group were among themselves reliable and the FLs were able to identify a foreign accent, whereas ratings of the TL group were only to some extent reliable. That is, when making their accent judgment on the stream of spontaneous speech there was a big difference in rating behavior. Naturally, the TLs had no possibility of relying on segmental information because they had never heard the target language before. Before the rating began, all the TLs had asked to hear a model voice representing standard Finnish pronunciation, as they wished to have a yardstick on which to base their accent judgment when rating the spontaneous speech samples. Furthermore, they reported difficulty in deciding on the accentedness scores, feeling that they were rating at random with no clue of what Finnish sounded like, which made them feel uneasy and uncomfortable. They had no expectations of how Finnish would sound, so they had no chance of knowing whether the speech signals corresponded to the pronunciation norms of Finnish. It was clear that the TLs were puzzled by the rating task.

Taken together, the findings of this study show how the TLs and FLs perceived the degree of foreign accent in Finnish in fundamentally dissimilar ways, each based on different phonetic parameters. There was a clear disparity and no agreement in the rating strategy between the FLs and TLs, who had never heard Finnish before and thus had no knowledge of its phonological structure. Although they had one source of language information (Turkish) available to them, they did not seem to utilize this familiarity with the L1 of the NNSs. Based on previous research on non-native listeners judging accentedness from languages unfamiliar to them, in the present study apart from the same L1 advantage, one could expect non-native listeners to make use of obvious perceptual cues such as slower tempo, i.e. speech rate (universal, non-linguistic speech characteristics as Bond et al. 2008 term them). The results of the study showed that the non-native listeners had indeed utilized general traces of non-native speech such as speech rate, but their use of speech rate did not improve their ability to identify natives from non-natives. That is, even though general traces of

non-nativeness might exist, and even though TLs used speech rate as a clue to non-nativeness, it did not lead to accurate accent ratings, an important distinction to make. Therefore, one cannot argue that the use of speech rate improved a listener's ability to detect accentedness.

Bond et al. (2008) found non-native listeners with no familiarity in the language they judge use utterance duration to make foreign accent judgment. Thus, Bond et al. (2008) concluded that their non-native listeners lacking familiarity with the accent and the target languages used fluency as a general marker of non-native speech, which they roughly estimated by utterance duration. The present study has found non-native listeners with no familiarity in the language they judged (the target language) use speech rate to judge a foreign accent. All in all, it seems that what is probably observable from speech without knowing the language is speech rate, tempo, utterance duration and fluency as these two studies have shown. Additionally, the reason for these features to signal non-native speech is because previous studies have found non-native listeners with no familiarity in the target language they judged to utilize these features while making their foreign accent judgments. In fact, what the speech rate findings of this study showed was plainly that the non-native listeners made use of speech rate to judge a foreign accent, so slower speech rate signaled more non-native speech to them, whereas a faster speech rate signaled more native speech to them. This finding shows that general traces of non-nativeness (speech rate) might exist. While it might seem counterintuitive that TL's use of speech rate as a cue to non-nativeness failed to enhance their ability to detect foreignness, it should be noted that previous studies also varied on whether they found a correlation between utterance duration and accent ratings. For instance, previous studies that measured utterance duration as an example of general markers for non-native speech noticed that the correlation between utterance duration and accent ratings were either insignificant (Weber & Pöllmann 2010: 540) or weak (Major 2007: 549). Only Bond et al. (2008: 6) found a high correlation between utterance duration and accuracy in identifying speakers as native or non-native for their American listeners. Thus, it seems that only Bond et al. (2008) were justified in concluding that their non-native listeners lacking familiarity with the accent and the target languages used fluency as a general marker of non-native speech.

5 Conclusion

In light of the findings in the present study, the following conclusions can be made about foreign accent detection: If listeners have excellent familiarity with the accent language (it is their L1) and no familiarity with the target language – as the TLs had in this study – they have both their L1 language source (if they can detect this accent language and utilize their familiarity with it) and some possible universal non-segmental information available to them in their use. That is, when non-native listeners consider they have no linguistic information on which to base their accent judgement, it is natural for them to rely on universal non-segmental information, such as speech rate or fluency, as indeed TLs did in the present study. There is a critical point here, however. Because this study demonstrated that the TLs' use of these general markers of non-nativeness, i.e. speech rate, did not improve their ability to detect foreignness, such argumentation is weak. However, the failure of the present study to reproduce the findings of Bond et al. (2008) adds to the accent detection literature by showing that the possible use of these universal non-segmental perceptual cues does not necessarily allow listeners to distinguish between native and non-native speakers accurately. The findings of the present study are in line with those of Bond et al. (2008) in the sense that the TLs based their judgment on some universal perceptual factors, i.e. speech rate. The difference in the findings of this study is that their foreign accent ratings based on speech rate perceptions were inaccurate, as they were unable to distinguish natives from non-natives. The findings of this study have clearly shown that it was challenging for native Turkish-speaking listeners to detect accentedness correctly (the accent language Turkish) when they were unfamiliar with Finnish as the target language.

In contrast, if listeners have excellent familiarity with the target language (it is their L1) and no familiarity with the accent language, as the FLs had in this study, they have their L1 language source, hence both segmental and non-segmental information from their native language, available to them. Thus, the FL had multiple sources of information on which to base their foreign accent evaluation. The acoustic analysis of speech rate by Praat and correlation analyses, however, showed that FLs did not make use of speech rate. All in all, the findings of this study that the TLs relied on non-segmental information such as speech rate agrees with the findings of Riney et al. (2005), Major (2007), Weber & Pöllmann (2010), and Bond et al. (2008). Riney et al.'s (2005: 441) acoustic and

auditory analyses showed that their untrained native listeners relied more on segmentals (vowels and consonants, especially /r/ and /l/), whereas untrained, non-native listeners of American English (native speakers of Japanese) relied more on non-segmental parameters (meaning everything else, including intonation, fluency, sentence duration, and speech rate) to make perceptual judgments. One must keep in mind, though, that the non-native listeners in the study by Riney et al. (2005) were familiar with English, the target language, whereas TLs in this study were unfamiliar with the target language.

The findings given in this study suggest that the commonly accepted view in literature observing that a listener's native speaker status has a strong positive effect on their foreign accent detection success is correct. However, this same view has been challenged by previous research such as Major (2007), Bond et al. (2008), and Weber & Pöllmann (2010). Of course, this raises the further question of why these empirical findings seem to point in the other direction as the everyday observation, and this might be a topic for further research. All in all, the findings of this study suggest that markers of non-nativeness are language-specific because in this study only the Finnish listeners who have formed a native perception of the language could identify a speaker's native speaker status, whereas the Turkish listeners failed to identify non-nativeness. However, the findings also showed that the Turkish listeners made use of some language-independent general markers of non-nativeness such as speech rate to identify non-nativeness.

Abbreviations

L1	first language
L2	second language
GLMM	generalized linear mixed model
ICC	intra-class correlation
LMM	linear mixed model
M	mean
NS	native speaker
NNS	non-native speaker
FL	Finnish listener
TL	Turkish listener
SPSS	statistical package for the social sciences

Appendix

The spontaneous speech instructions

Discuss one of these subjects (or make a subject up yourself). Your reply should be only 1-minute long (reply in Finnish).
A. Describe your weekend or your daily routine: What do you usually do, when, with whom, for how long, what is interesting about it, etc.?
B. Describe one significant experience in your life: Who was included? How old were you then? How did this affect you?
C. Describe a person in your life who means a lot to you: How do you know this person? Why is she/he important in your life?

References

Anderson-Hsieh, Janet & Johnson, Ruth & Koehler, Kenneth. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language learning* 42(4). 529–555.

Bond, Zinny. September 5th 2016. (Personal communication.)

Bond, Zinny S. & Stockmal, Verna. & Markus, Dace. 2008. A note on native and non-native accentedness judgments. *Ohio university working papers in linguistics and language teaching*. 1–8.

Bunta, Ferenc & Major, Roy. 2004. An optimality theoretic account of Hungarian ESL learners' acquisition of/e/and/æ. *IRAL: International Review of Applied Linguistics in Language Teaching* 42. 277–299.

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences.* (2nd ed.) Hillsdale (NJ): Lawrence Erlbaum.

Derwing, Tracey M., & Munro, Murray J. 1997. Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition* 19(01), 1–16.

Dewaele, Jean-Marc. 1996. Les phénomènes d'hésitation dans l'interlangue française: analyse de la variation interstylistique et interindividuelle [The phenomena of hesitation in the French interlanguage: analysis of the interstylistic and interindividual variation]. *Rassegna Italiana da Linguistica Applicata* 28. 87–103.

Esling, John H. 2000. Crosslinguistic aspects of voice quality. In Raymond D. Kent & Martin J. Ball (eds.). *Voice quality measurement*, 25–35. San Diego: Singular Publishing Group.

Flege, James E. 1988. Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America* 84(1). 70–79.

—— 1981. The phonological basis of foreign accent: A hypothesis. *Tesol Quarterly* 15(4). 443–455.

Flege, James E. & Munro, Murray J., & MacKay, Ian R. A. 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16. 1–26.

González-Bueno, Manuela. 1997. Voice-onset-time in the perception of foreign accent by native listeners of Spanish. *IRAL-International Review of Applied Linguistics in Language Teaching* 35(4). 251–268.

IBM Corp. Released 2016. *IBM SPSS statistics for Windows, version 24.0*. Armonk, NY: IBM Corp.

Iversen, Gudmund R. 1991. *Contextual analysis.* Sage university paper series on quantitative applications in the social sciences, No. 7. Beverly Hills, CA: Sage Publications.

Jilka, Matthias. 2000. Testing the contribution of prosody to the perception of foreign accent. *New Sounds 2000: Proceedings of the fourth international symposium on the acquisition of second-language speech,* 199–207. Klagenfurt, Austria: University of Klagenfurt.

Kang, Okim. 2008. Ratings of L2 oral performance in English: Relative impact of rater characteristics and acoustic measures of accentedness. *Spaan fellow working papers in second or foreign language assessment* 6. 181–205.

Karlsson, Fred. 2008. *Finnish: An essential grammar.* London: Routledge.

Lennon, Paul. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning* 40. 387–417.

MacKay, Ian R. A. & James E. Flege & Satomi Imai. 2006. Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics* 27(02). 157–183.

Madsen, Henrik & Thyregod, Poul. 2010. *Introduction to general and generalized linear models*. Boca Raton: Chapman & Hall/CRC Press.

Magen, Harriet S. 1998. The perception of foreign-accented speech. *Journal of Phonetics* 26(4). 381–400.

Major, Roy C. 1987. English voiceless stop production by speakers of Brazilian Portuguese. *Journal of Phonetics* 15(2). 197–202.

—— 2001. *Foreign accent: The ontogeny and phylogeny of second language phonology*. New York: Routledge.

—— 2007. Identifying a foreign accent in an unfamiliar language. *Studies in Second Language Acquisition* 29(04). 539–556.

—— August 25th 2016. (Personal communication.)

Moyer, Alene. 2013. *Foreign accent: The phenomenon of non-native speech.* Cambridge: Cambridge University Press.

Munro, Murray J. & Derwing, Tracey M. 1995. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and speech* 38(3). 289–306.

—— 1998. The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning* 48(2). 159–182.

—— 2001. Modeling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate. *Studies in second language acquisition* 23(04). 451–468.

Munro, Murray J. & Derwing, Tracey M. & Flege, James E. 1999. Canadians in Alabama: A perceptual study of dialect acquisition in adults. *Journal of Phonetics* 27(4). 385–403.

Munro, Murray J. & Derwing, Tracey M. & Burgess, Clifford S. 2010. Detection of nonnative speaker status from content-masked speech. *Speech Communication* 52(7–8). 626–637.

Munro, Murray J. & Derwing, Tracey M. & Morton, Susan L. 2006. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition* 28(01). 111–131.

Nelder, John A. & Wedderburn, Robert W. M. 1972. Generalized linear models. *Journal of the Royal Statistical Society* 135(3). Series A (General). 370–384.

Nunnally, Jum & Bernstein, Ira. 1994. *Psychometric theory* (3rd ed.). New York: New York: McGraw-Hill.

O'Dwyer, L. M. & Parker, C. E. 2014. *A primer for analyzing nested data: multilevel modeling in SPSS using an example from a REL study*. Washington, DC: Regional Educational Laboratory Northeast & Islands.

Riney, Timothy J.& Takada, Mari & Ota, Mitsuhiko. 2000. Segmentals and global foreign accent: The Japanese flap in EFL. *Tesol Quarterly* 34(4). 711–737.

Riney, Timothy J. & Takagi, Naoyuki. 1999. Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning* 49(2). 275–302.

Riney, Timothy J. & Takagi, Naoyuki & Inutsuka, Kumiko. 2005. Phonetic parameters and perceptual judgments of accent in English by American and Japanese listeners. *TESOL Quarterly* 39(3). 441–466.

Sawilowsky, Shlomo S. 2003. A different future for social and behavioral science research. *Journal of Modern Applied Statistical Methods* 2(1). 128–132.

Schmid, Monika S. & Hopp, Holger. 2014. Comparing foreign accent in L1 attrition and L2 acquisition: Range and rater effects. *Language Testing* 31(3). 367–388.

Scovel, Thomas. 1995. Differentiation, recognition, and identification in the discrimination of foreign accents. In J Archibald. (ed.), *Phonological acquisition and phonological theory,* 169–195.

Shrout, Patrick E. & Fleiss, Joseph L. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin* 86(2). 420–428.

Toivola, Minnaleena. 2011. *Vieraan aksentin arviointi ja mittaaminen suomessa* [The assessment and measurement of foreign accent in Finnish]. Helsinki: Helsingin yliopisto.

Trofimovich, Pavel & Baker, Wendy. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28(1). 1–30.

Weber, Andrea & Pöllmann, Katja. 2010. Identifying foreign speakers with an unfamiliar accent or in an unfamiliar language. *New sounds 2010: Proceedings of the sixth international symposium on the acquisition of second language speech*, 536–541. Poznan: Adam Mickiewicz University.

Contact information:

Melike Uzal
Faculty of Arts
Department of Modern Languages
Speech Sciences
P.O. Box 9, FI-00014
University of Helsinki, Finland.
e-mail: melike(dot)uzal(at)helsinki(dot)fi

Erkki Komulainen
University of Helsinki
e-mail: erkki(dot)komulainen(at)helsinki(dot)fi

Mehmet Akif Kılıç
Istanbul Medeniyet University
e-mail: makilic(at)yahoo(dot)com

Olli Aaltonen
University of Helsinki
e-mail: olli(dot)aaltonen(at)helsinki(dot)fi

Fixed and NOT free: Revisiting the order of the main clausal constituents in Finnish Sign Language from a corpus perspective

Tommi Jantunen
University of Jyväskylä

Abstract

This squib investigates the order of the main clausal constituents in verbal-centered clauses in Finnish Sign Language (FinSL). With the help of the frequencies calculated from narratives included in the recently compiled FinSL corpus, the study suggests that the order of the main constituents in FinSL clauses is more fixed than has been claimed in a previous study. With intransitive and transitive clauses with a Type 1 or Type 2 verbal predicate, the study shows that their internal structures strongly favor the orders SV and AVP, respectively, although both S/A and P core arguments are often left lexically unexpressed. Concerning Type 3 verbal predicates, the study shows that they most typically form simple sentences on their own and that if they appear with any additional nominal material, this material tends to precede the verbal.

Keywords: constituent order, intransitive clause, transitive clause, Finnish Sign Language

1 Introduction

Previous work (Jantunen 2008) on the order of the main constituents in Finnish Sign Language (FinSL) clauses has targeted transitive structures and suggested that they are organized according to the "fixed and free" principle: the A argument (the primary core argument of a transitive clause) always precedes the V (the verbal), and the order of all three main constituents is either AVP, APV or PAV (P refers to the secondary core argument). Empirically, this result was mainly based on elicited data which – as the author admitted at the time – was too narrow to allow for any

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

numerical generalizations about the typicality of the three orders. Now, for the first time, the recently compiled corpus of FinSL makes possible the investigation of the constituent order with reference to frequencies. Consequently, in this squib we will revisit the prevailing "fixed and free" view of FinSL constituent order and specify and, if needed, correct it with the help of numerical information. Moreover, as the corpus now makes possible the investigation of the constituent order in other kinds of structures, we will extend the discussion to cover also intransitive clauses.

2 Research material and its processing

All the results in the present paper concerning the order of the main clausal constituents are based on a sample of video data extracted from a larger body of material constituting the FinSL corpus (Jantunen et al. 2016; Salonen et al. 2016). In practice, this sample refers to the ten signed re-tellings of the stories *Snowman* (5 re-tellings) and *Frog, where are you?* (5 re-tellings) elicited in 2013 from ten native FinSL signers (6 female, 4 male; ages between 20 and 60 years) with the help of text-less picture books by Briggs (1978) and Mayer (1969), respectively. In the recordings, the signers worked in pairs in a dialogue setting in which the recording set-up consisted of 6 cameras directed toward the signers from different angles. The task of the signers was to look at the book, memorize the story, then put the book away and tell the story to the addressee. The combined duration of the signed stories in the sample is 37 minutes and 56 seconds.

The sample has been annotated on various levels (see Jantunen et al. 2016) in ELAN (Crasborn & Sloetjes 2008; Max Planck Institute for Psycholinguistics[1]) by three annotators with native competence in FinSL. Of the different annotations, the most crucial ones for the present study are those that identify the signs and the structures of the verbal-centered clauses. The sample includes altogether 3379 sign tokens and 933 structurally annotated verbal-centered clauses.

Each sign has been annotated for a gloss and a lexico-grammatical category. In terms of the lexico-grammatical category, a sign can be a nominal (marked with the prefix *n* in front of the sign gloss) or a verbal (*v*), or be overtly unspecified in terms of category (*x*). The category has been

[1] The Language Archive, Nijmegen, the Netherlands: see http://tla.mpi.nl/tools/tla-tools/elan.

indicated with all signs except with those that are glossed as pointings or ad hoc gestures: for the purpose of future research, these have been treated as semantically and formally independent units and annotated without any information prefixed to their glosses. In the sample, the number of nominal signs is 743, the number of verbal signs is 1300 and the number of signs marked as unspecified is 852. Thus, as the total number of signs in the sample marked for some category is 2895, the total number of signs identified as pointings and ad hoc gestures is 484.

In FinSL research, the lexico-grammatical categories nominal and verbal have been defined by semantic and grammatical criteria (Jantunen 2010) and they can both be further divided into subclasses. Of these, the three subclasses of verbal signs – Type 1, 2 and 3 verbals (resembling the plain, indicating and depictive verbs of Liddell 2003, respectively; see Jantunen 2010 for a full discussion of the differences) – are the most researched ones and crucial also for the present study. Type 1 verbals (e.g. THINK, PLAY) are formationally the most fixed type, with a relatively straightforward form–meaning connection. Type 2 and 3 verbals, on the other hand, are groups of signs which include gradient features (i.e. a gestural component) as part of their structure. In Type 2 verbals (e.g. LOOK-AT, TAKE) – as, for example, in pointings – the gestural component is manifested through the directionality of the movement of the hand: in order to understand the meaning of these signs, the addressee must make a semantic association between the morphological content of the sign and the location toward which the sign is directed (see Liddell 2003). In Type 3 verbals (e.g. CL-V-fall-down ‘a two-legged animate object falls down from a high place’), the gestural component refers to the gestural properties inherently present in the placement, orientation and movement parameters of the sign structure. Together, these features enable Type 3 verbals to iconically depict events involving autonomous or caused motion taking place in different locations. Type 3 verbals also contain a handshape (e.g. a whole-entity or a handling handshape) that is analyzed as a nominal classifier (CL) morpheme, which refers to the entities present in the event (e.g. V, the two-finger “victory” whole-entity handshape, refers to the animate two-legged objects such as ‘the boy’). In the annotation of verbal signs, Type 3 verbals have been distinguished from the other types (with the prefix *k* in the gloss).

The 933 verbal-centered clauses that have been annotated for their syntactic structure constitute 63% of all of the clauses produced by the ten signers (n = 1477). In the annotation work, the notion of clause was understood in the basic sense of Van Valin & LaPolla's (1997) *Role and Reference Grammar* (RRG), and all verbal-centered clauses were identified by looking for verbal signs that function as a predicate (V) on the clause level. Once the predicate was identified, its immediate context was then analyzed and annotated in terms of possible core arguments: S, the single core argument of an intransitive clause (with semantically monovalent verbal predicates), and A and P, the primary and secondary core arguments of a transitive clause (predicates with valency more than one), respectively.[2] The analysis also indicated the cases where core arguments were not expressed lexically. This was done by placing the core argument symbol(s) in parentheses (for more on argument ellipsis, see Jantunen 2013).

The prevailing theory of verbal-centered clauses in FinSL, presented first in Jantunen (2008) and followed in this work, treats clauses formed around Type 1 and 2 verbals differently from clauses formed around Type 3 verbals. In the prototype of the clause with a Type 1 or 2 verbal as its predicate, the predicate and the core argument(s) are all free lexical or semi-lexical units (e.g. lexical nominals, pointings). In the prototype of the clause with a Type 3 verbal as its predicate, on the other hand, the core arguments are analyzed as being fused into the predicate (cf. the phenomenon of head-marking; Nichols 1986; Jantunen 2008). In practice, the core arguments are represented by the entity-referring classifier handshape(s) of the Type 3 verbal. A consequence of this analysis is that Type 3 verbals are well-formed clauses – and simple sentences – on their own without any additional lexical material. A Type 3 verbal can concatenate with nominal material in a sentence but this material is not counted as core internal: following the RRG view of the clause, it is analyzed either as a clause-internal periphery (e.g. material with adverbial or oblique function) or clause-external left or right-detached frame-setting material (e.g. a topic) (Jantunen 2013; see §4).

[2] Practically all transitive clauses in the sample had a bivalent verbal predicate. The total number of trivalent predicates in the sample was 13. The third core argument was annotated with the symbol E.

In the annotation of the verbal-centered clauses, a distinction was made between clauses that have a Type 1 or 2 verbal as their predicate and those that have a Type 3 verbal as their predicate. In the former case, the core argument symbols were written in upper case letters disconnected from the predicate symbol (e.g. *S V*, *A V P*), whereas in the latter case the symbols were written in lower case together with the predicate symbol (e.g. *sV*, *aVp*). Making the distinction is important because the investigation of the main constituent order makes sense only with respect to clauses that have a Type 1 or 2 verbal as their predicate and free units as their core arguments.

A part of the above described data is openly accessible in a slightly reduced format via the LAT online service of the FIN-CLARIN's Language Bank of Finland[3] (*Kielipankki* in Finnish).

3 The order of the main constituents in clauses with a Type 1 and 2 predicate

The sample of 933 verbal-centered clauses comprises 712 clauses (76%) that have a Type 1 or 2 verbal predicate and 221 clauses (24%) that have a Type 3 verbal predicate. In this section (3), we will deal with the clauses with a Type 1 or 2 predicate (for structures built around Type 3 predicates, see §4).

3.1 Intransitive clauses

The sample contains 331 intransitive clauses that have either a Type 1 or Type 2 verbal as their predicate. However, in only 119 of these clauses (36%) is the S argument expressed overtly (i.e. the clauses are structurally full), whereas 212 clauses (64%) have no overtly expressed S (the transitivity status of structurally incomplete clauses has been decided on the basis of context and the semantic valency of the predicate). Consequently, the data suggests a strong trend towards omission of the S argument in intransitive FinSL clauses with a Type 1 or Type 2 predicate (Jantunen 2013).

[3] See http://urn.fi/urn:nbn:fi:lb-1001100113005.

According to the annotated data, the preferred constituent order of FinSL intransitive clauses is SV. In clauses that consisted only of S and V and no other material (n = 34), the order was SV in all cases. Two examples of such clauses are given in (1). In all of the numbered examples that follow, the first row always presents the glosses, the second row the syntactic annotation and the third row the translation into English.

(1) a. n_BOY v_SLEEP
S V
'The boy sleeps.'

b. n_SNOWMAN v_WAKE-UP
S V
'The snowman wakes up.'

In clauses that also had other material before, between or after the main constituents, the order was SV in 116 cases, that is, in 97% of the analyzable clauses. Two examples of these are given in (2).

(2) a. n_MORNING n_BOY v_WAKE-UP
S V
'The boy woke up in the morning.'

b. x_BECAUSE n_SNOWMAN x_SELF v_MELT-DOWN
S V
'...because the snowman himself would melt down.'

The three clauses with VS order in the data were all existential or locative intransitive expressions (e.g. x_THERE [v_EXIST n_BOY] 'The boy is there'; for existential and locative expressions in FinSL, see De Weerdt 2016).

3.2 Transitive clauses

The total number of transitive clauses with a Type 1 or Type 2 verbal predicate in the syntactically annotated sample is 381. However, as was the case with intransitive clauses, only in a small subset of transitive clauses were all the nominal main constituents (A and P) expressed overtly: the number of such clauses was 56 (15%). Conversely, altogether 325

transitive clauses (85%) had no overtly expressed A and/or P core argument(s). Consequently, the omission of the lexical core argument(s) is a strong trend also in transitive FinSL clauses with a Type 1 and 2 predicate (Jantunen 2013).[4]

A closer inspection of argument omission in transitive clauses reveals that the phenomenon has most effect on the A argument: 79% of A arguments (n = 299) were omitted in the data. However, 39% of P arguments (n = 150) were omitted too. It is also important to note that 33% of the transitive clauses (n = 126) consisted only of a verbal predicate without any major nominal elements.

Concerning the order of A, P and V in the 56 syntactically complete transitive clauses, the preferred order in the sample was clearly AVP: 48 instances (86%) of the clauses had this particular order. The orders APV and PAV were also identified in the data (Jantunen 2008). However, the total number of APV orders was five and that of PAV orders only three. Examples of all three orders are given in (3).

(3) a. n_BOY v_MAKE n_SNOWMAN
A V P
'The boy makes a snowman.'

b. n_BOY n_SNOWMAN v_WAVE
A P V
'The boy waves to the snowman.'

c. POINT:that n_BOY v_REMEMBER
P A V
'That (is what) the boy remembers.'

The dominance of the order AVP over the two other orders (APV and PAV) is further strengthened when we look at the relative order of V and P in clauses in which they are present. For this purpose, in order to guarantee comparability, we used a reduced sample of 216 clauses from the total of 231 clauses with V and P. In this comparison, V precedes P in 186 cases (86%; see 4a–b) while P comes before V in only 30 cases (14%; see 4c). For example (in 4b, the slash marks a pause that detaches the frame-setting topic from the rest of the comment clause from which the A argument has

[4] Of the 13 clauses with a trivalent predicate, only 6 had an overt third argument (E).

been omitted; the clause in 4c is also structurally incomplete in terms of its A argument):

(4) a. POINT:you v_SEE n_MY-OWN n_WORLD
A V [P]
'You (will) see my world.'

b. n_WINDOW / v_LOOK-AT n_SNOWMAN
V P
'In the window, [the boy] looks at the snowman.'

c. n_FRONT-DOOR v_OPEN-DOOR
P V
'[The boy] opens the front door.'

In the data, A always precedes V, which is in accordance with the original claim made by Jantunen (2008) that the VA order is not found in declarative FinSL clauses. It should be emphasized that the annotators of the data were instructed not to respect this claim and to use the order VA if they encountered it in the data.[5]

3.3 Interim discussion

Previous work on FinSL has argued (Jantunen 2013) that, because of the frequent omission of core arguments (attested also in other sign languages, e.g. Wulf et al. 2002), FinSL can be considered a discourse-oriented language. The essence of the argument is based on McShane (2005), according to whom ellipsis is very frequent in discourse-oriented languages. The numbers presented in the present work provide further evidence for the argument concerning FinSL: of all the clauses with a Type 1 or 2 verbal predicate, the clear majority (n = 537; 75%) are syntactically incomplete.

The most frequent and unmarked constituent order of simple declarative transitive clauses has been used to identify the so-called basic order of elements in the world's languages. From the present data, this

[5] In clauses with a trivalent predicate, E came after V. The order between P and E could not be determined because of the very low number of tokens.

order in FinSL would be AVP. In general, the order AVP or equivalent has been found to have a fairly dominant role in all sign languages, though the order APV has been shown to exist as well (e.g. Liddell 2003 for American Sign Language; Johnston & Schembri 2007 for Australian Sign Language; and Kimmelman 2012 for Russian Sign Language, to name just a few). The order AVP, together with APV, dominates also outside the domain of sign languages: according to the *WALS* data provided by Dryer & Haspelmath (2013), these two orders are the ones most commonly found in the world's spoken languages.

4 The structure of simple sentences forming around a Type 3 verbal predicate

As explained in §2, Type 3 verbals are here treated as well-formed clauses on their own. The analysis derives on the one hand from the fact that the classifier handshape or handshapes included in these verbals can be analyzed as nominal core arguments of the predicate in the same way as certain bound morphemes fused into the predicates can be analyzed as core arguments in strong head-marking languages such as Navajo and Trotzil (Nichols 1986; Jantunen 2008). On the other hand, the analysis of Type 3 verbals as full clauses derives from the fact that the meaning of Type 3 verbals is typically very clause-like, that is, it covers the whole event. Due to the semantics of classifiers, however, there is often some vagueness in the meaning of such verbals/clauses. For this reason, Type 3 verbals may be preceded by a nominal phrase whose function is to set an interpretative framework for the main predication (here, the Type 3 verbal). In such structures, the Type 3 verbal is always sentence final (Jantunen 2008). However, although sentence-internal material typically precedes Type 3 verbals, these verbals may also be followed by peripheral (e.g. adverbial or oblique-like) material that describes the location or even the goal of the depicted motion event. Such material is counted as clause-internal in the present framework.

The two examples in (5) demonstrate two different structures involving Type 3 verbals. In (5a), the Type 3 verbal forms the clause (and, consequently, a simple sentence) on its own. In (5b), the same Type 3 verbal occurs with additional clause-external material (a frame-setting topic on the left) and peripheral material (the goal setting oblique nominal on the

right) which, together with the verbal, constitute a simple sentence. In (5b), the single quotation mark (') indicates a non-durational break in prosody (typically involving at least an eye blink).

(5) a. v_k_CL-V-fall-downwards
sV
'A two-legged animate object falls down from a high place.'

b. n_BOY / v_k_CL-V-fall-downwards ' n_RIVER
TOP sV periphery
'The boy falls down from a high place into the river.'

Table 1 summarizes the statistics concerning the order of elements in simple sentences containing a clausal Type 3 verbal predicate (n = 221) in the sample data.

Table 1. The order of elements in simple sentences formed around a Type 3 verbal predicate in the data

Sentence structure	**n**	**%**
only a Type-3-verbal	139	63
element + Type-3-verbal	71	32
Type-3-verbal + element	11	5
element + Type-3-verbal + element	9	4

Table 1 shows that, in the data, Type 3 verbals most typically form a sentence on their own, without any additional lexical material (63%). If a Type 3 verbal is combined with lexical material within a sentence, then this material precedes the verbal (32%). Material following the verbal is very infrequent in the data, as are structurally full manifestations of simple sentences with Type 3 verbal predicates.

Prototypically, an element or elements surrounding Type 3 verbals are separated from the verbal by prosodic breaks. In general, breaks before the Type 3 verbal tend to be more prominent (e.g. a pause) than breaks after the verbal (e.g. often only an eye blink). However, in the data, there are also constituent borders inside the sentences containing Type 3 verbals which cannot be assigned any prosodic break. This is at least partly due to coarticulation and to the fact that prosody is constantly used to bind together larger bits of discourse in FinSL (Jantunen 2016).

Of the 221 Type 3 verbals, 54% have been annotated as intransitive and 46% as transitive. However, it must be noted that the analysis of Type

3 verbals in terms of their transitivity was considered to be very difficult by the annotators: the verbals were not easily categorized in these terms. This, it is assumed, is because of the gestural component of the verbals, which causes both their structure and meaning to appear as highly gradient. More work on the transitivity of Type 3 verbals is called for.

5 Conclusions

The study reported in this squib has investigated, for the first time with the help of corpus frequencies, the order of the main clausal constituents in verbal-centered clauses in FinSL. In terms of intransitive and transitive clauses with a Type 1 and 2 verbal predicate, the study has shown that their internal structures strongly favor the orders SV and AVP, respectively, although other structures do exist and core arguments are often left lexically unexpressed. Concerning Type 3 verbal predicates, the study has shown that they normally form simple sentences on their own and that if they appear with any additional nominal material, this material tends to precede the verbal. We conclude that, on the basis of the frequencies drawn from the present data, the order of the main constituents in FinSL clauses appears to be more fixed than was previously thought and *not* as "free" as was characterized by Jantunen (2008).

Acknowledgments

The author wishes to thank the two anonymous reviewers for their comments as well as Eleanor Underwood for checking the English of the text. The study has been supported financially by the Academy of Finland under grants 269089 and 273408.

References

Briggs, Raymond. 1978. *The snowman*. London: Hamish Hamilton.

Crasborn, Onno & Sloetjes, Han. 2008. Enhanced ELAN functionality for sign language corpora. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, 39–43. Paris: ELRA.

De Weerdt, Danny. 2016. Existential sentences in Flemish Sign Language and Finnish Sign Language. *SKY Journal of Linguistics* 29. 7–38.

Dryer, Matthew S. & Haspelmath, Martin (eds.). 2013. *WALS: The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info) (Accessed 2016-05-04.)

Jantunen, Tommi. 2008. Fixed and free: Order of the verbal predicate and its core arguments in declarative transitive clauses in Finnish Sign Language. *SKY Journal of Linguistics* 21. 83–123.

—— 2010. Suomalaisen viittomakielen pääsanaluokat [The main parts of speech in FinSL]. In Jantunen, Tommi (ed.), *Näkökulmia viittomaan ja viittomistoon* [Perspectives on sign and lexicon], 57–78. Jyväskylä: University of Jyväskylä.

—— 2013. Ellipsis in Finnish Sign Language. *Nordic Journal of Linguistics* 36. 303–332.

—— 2016. Clausal coordination in Finnish Sign Language. *Studies in Language* 40. 204–234.

Jantunen, Tommi & Pippuri, Outi & Wainio, Tuija & Puupponen, Anna & Laaksonen, Jorma. 2016. Annotated video corpus of FinSL with Kinect and computer-vision data. In *Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*, 93–100. Paris: ELRA.

Johnston, Trevor & Schembri, Adam. 2007. *Australian Sign Language: An introduction to sign language linguistics*. Cambridge: Cambridge University Press.

Kimmelman, Vadim. 2012. Word order in Russian Sign Language: An extended report. *Linguistics in Amsterdam* 5. 1–56.

Liddell, Scott. 2003. *Grammar, gesture, and meaning in American Sign Language*. Cambridge: Cambridge University Press.

Mayer, Mercer. 1969. *Frog, where are you?* New York: Dial Press.

McShane, Marjorie J. 2005. *Theory of ellipsis*. Cary, NC: Oxford University Press.

Nichols, Johanna. 1986. Head-marking and dependent-marking grammar. *Language* 62. 56–119.

Salonen, Juhana & Takkinen, Ritva & Puupponen, Anna & Pippuri, Outi & Nieminen, Henri. 2016. Creating the corpus of Finland's sign languages. In *Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*, 179–184. Paris: ELRA.

Van Valin, Robert D. & LaPolla, Randy J. 1997. *Syntax: Structure, meaning and function*. Cambridge: Cambridge University Press.

Wulf, Alyssa & Dudis, Paul & Bailey, Robert & Lucas, Ceil. 2002. Variable subject presence in ASL narratives. *Sign Language Studies* 3. 54–76.

Contact information:

Tommi Jantunen
Finnish Sign Language,
Department of Language and Communication Studies
P.O. Box 35 (F)
40014 University of Jyväskylä
Finland
e-mail: tommi(dot)j(dot)jantunen(at)jyu(dot)fi

Ethnographic chats: A best of both method for ethnography

Charlotte Selleck
University of the West of England

Abstract

Conventionally, ethnographic methods in sociolinguistics aim to discover how language works as "situated social practice and how it is tied to social organisation" (Heller 2011: 10). Within this, ethnography has viewed participant observation as central and essential. More recently ethnographers have moved to combine this with more structured, researcher-facilitated question-based tools such as ethnographic interviews (Sherman Heyl 2001) and focus groups (Suter 2000). This article reports on another creative method, aiming to bring together the strengths of both these approaches to access school-age young people's orientations to language education policies. There were three main motivations: firstly, to minimise the distracting influence of the researcher's presence, secondly, to aid in empowering participants, encouraging them into an active role in the research process and thirdly, to avoid favorability bias in participant responses. On the latter point, to truly value the voice of participants you have to find ways to move beyond the "right answer", which often requires pushing methodological boundaries. I developed a new protocol, ethnographic chats,[1] which I found offered the best of both from existing approaches: a compromise between the immersive depth of participant observation and the greater thematic precision of focus groups or ethnographic interviews. The method was characterised by specific procedural and interactional characteristics of frame and genre, which differentiate it in specific ways from ethnographic interview and focus group methods. Rich data emerged from this process, which would not otherwise have been available. I conclude by outlining the potential for ethnographic chats in other social and geographical contexts.

Keywords: ethnography, research methods, bilingual education, Wales

[1] Note that the term chat is not referring to online, but to face-to-face communication.

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

1 Ethnography – the tradition

Hammersley (2006: 3) suggests that like many other methodological terms in the social sciences, ethnography does not form "part of a clear and systematic taxonomy". It is used in different ways and to describe various related approaches. Nevertheless, these different ethnographic approaches share many common features. Principally, ethnography refers to a form of social and educational research that is committed to "the first-hand experience and exploration of a particular social or cultural setting" (Atkinson et al. 2007: 4). The nature of ethnographic research means that "no homogeneous units or specific characteristics of culture are defined a priori, but rather those groups and processes recognised by native participants are discovered and studied in their terms during the research" (Gregory 1983: 366). Malinowski (1922: 8–9) talks of "foreshadowed problems", rather than fixed research questions; and his anthropological linguistic research was foundational for ethnography.

Instead of going into the field with fixed ideas, ethnography is concerned with producing descriptions and explanations of particular phenomena, with the process and inquiry becoming progressively more focused. More than any other research method, ethnography requires the researcher to follow themes wherever they lead; it is a generative process,[2] requiring flexible adaptation.

The term ethnography refers primarily to a "particular method or set of methods" (Hammersley & Atkinson 1993: 1) characteristically involving the researcher participating, overtly or covertly in people's daily lives: watching what happens; listening to what is said; asking questions (through informal or formal interviews); and collecting whatever data is available to shed light on the focus of the research. In other words, ethnography, as a method of social research, seeks to capture and understand the meanings and dynamics in particular cultural settings using a range of systematic data-collection techniques.

2 Focused discussions

Whilst the mainstay of ethnography is participant observation (Hymes 1972), ethnographers often combine this with more structured question-

[2] One where new ideas and representations are constantly emerging and where existing understanding is continuously questioned and challenged.

based ethnographic methods, as well as audio recordings and visual materials, including photography, film and video. These more structured question-based methods can conventionally range from an opportunistic conversation, where questions arise on the spur of the moment and where accounts of these passing and fleeting conversations are captured in field notes (Roberts et al. 2001), to in-depth, one-to-one interviews (Sherman Heyl 2001) that are formally arranged, recorded and transcribed. O'Reilly (2012: 136–138) also talks of "group interviews" arguing that they are akin to focus groups in allowing for multiple views to be garnered. Suter (2000) advocated for the use of focus groups in an ethnographic approach where topics of inquiry do not provide ample opportunities for observation.

Focus groups share many common features with less structured interviews, but still revolve around a discussion being guided, monitored and recorded by a researcher. They also still sit some way apart from the more immersive experience of participant observation.[3] In my research, I sought to bring these two elements together. My research aimed to assess students' orientations to the consequences of language education policies. In this context, the traditional format of initiation/response sequences was felt to be inconsistent with the ethnographic priority that "no homogenous units or specific characteristics of culture are defined a priori" but rather "those groups and processes recognised by native participants are discovered and studied in their terms during the research" (Gregory 1983: 366).

Additionally, my aim was to empower research participants, to give them a voice and to allow them to become an active part of the research process. In order to fully realise this aim, researchers often need to work in new or creative ways in order to push methodological boundaries. In light of this, I built on existing methods to develop a refined ethnographic protocol.

3 The ethnographic chat

My research began as a conventional ethnography. My sites spanned two schools and one youth club. I observed activities both inside and outside the classroom. My observations were recorded in 27 sets of field notes

[3] Although note that Bloor et al. (2001: 5–6) argue that a focus group methods can, if managed appropriately, "yield up as much rich data […] as long periods of ethnographic fieldwork".

representing approximately 110 hours of fieldwork (in all three sites). An analysis of my initial field notes was undertaken in order to formulate research questions. Participant observation was working well, but when it came to complementing this with something more targeted, conventional researcher-facilitated tools like focus groups seemed ill-suited to capturing rich ethnographic insights. Furthermore, as noted above, I sought to empower my participants, to encourage active participation in the research. This similarly required some innovation.

Meanwhile, working as a non-Welsh speaking researcher in a bilingual (Welsh and English) community raised other practical concerns. I wanted to ensure that I was able to offer the participants a choice as to which language(s) to use during the research process but would have been unable to do this with a researcher-facilitated approach such as an interview or a focus group. The development of the ethnographic chat helped with this as well.[4]

Open-ended prompts were written to be used as the basis for the ethnographic chats, a sample of which is shown below, in Figure 1.

Discuss what you think your school thinks about language.
Trafodwch beth rydych yn meddwl bod eich ysgol yn meddwl am iaith.

Discuss how you think they would describe the perfect student.
Trafodwch sut byddai'r ysgol yn disgrifio'r disgybl perffaith, yn eich barn chi.

Discuss and describe whether there are Welsh or English students or staff at Ysgol Arnant / Ysgol Ardwyn.
Trafod a disgrifio a oes fyfyrwyr Cymraeg neu Saesneg neu staff yn Ysgol Arnant / Ysgol Ardwyn.

Figure 1. Sample of prompts

The prompts were written in both Welsh and English. Participants were given the choice as to which language(s) to use, and were explicitly told they could use both. Prompts were pragmatically realised as open-ended

[4] It is worth pointing out that my own position as a non-Welsh speaking researcher also had benefits in that participants felt the need to fully explain and justify their experiences, views, and ideologies (as opposed to implying and assuming knowledge on my part). Furthermore, my "outsider" status afforded me analytical distance on the research and emergent data. Winchatz (2006) also notes that the researchers own language skills (or lack of them) do function as fruitful ways to reach emic interpretations and are not always a hindrance.

"topics" rather than specific questions. This allowed and encouraged participants to have open and apparently frank conversations. Primarily the prompt-based chats were deployed to elicit evaluative discourse and key ideological stances as well as an analysis of reported language practice. That chat data was not therefore treated as a potential proxy for direct observation. Whilst the students generally proved to have a shared understanding of everyday experiences, some disagreement did emerge. A multiplicity of views was garnered but with consensual stances predominating. I had limited involvement in these chats, which proved crucial (discussed further below).

But these "chats" were not simply thrust at these young people out of nowhere. In the tradition of ethnography, I had previously spent several months living and working in the community, carrying out participant observations. Approximately forty visits of varying length were made to the community; and my time at the schools was spent observing classrooms, assemblies, break times, lunchtimes, school shows, sporting fixtures, and parents' evenings. I also observed and participated in community events such as local fêtes and cultural festivals.

On the basis of initial observations (as recorded in field notes), approximately twenty students were chosen as principal participants (key informants) in each school. Selecting key informants for ethnography should not be thought of as a sampling procedure based on empiricist principles of representativeness. That said, careful consideration was given to ensure, where practically possible, that a broad spectrum of experiences was reflected in the research, and in light of this a range of language abilities, language preferences, medium of instruction, ages, and genders were taken into consideration. Key informants were chosen on the basis of initial observations (see Selleck 2013: 55–60, for further details of participation selection). It was these key participants who went on to be involved in the ethnographic chats. I was well known to these students and had built good working relationships with them.

A group of 4–5 students (aged between 11 and 18), all key informants and part of an established friendship group,[5] were asked to take part in the ethnographic chats. The format of the sessions was consistent throughout. Participants scheduled the chats themselves, at a mutually convenient time

[5] Gamson (1992) in his "peer group conversations" minimized the researchers role and brought together groups of acquaintances. Likewise, Press & Cole (1999) in their "ethnographic focus groups" also gathered their insights from conversations with groups of friends who met in a home environment.

and location. This made for a relaxed, informal environment with students partaking in seemingly unrelated activities such as eating their lunch and listening to music. Allowing participants to do other things whilst discussing a series of prompts allowed for the kind of blending of approaches identified earlier, namely informal participation and more formal interviewing. These "other" activities often became relevant to the emerging data, for example they led me to previously unknown students, teachers, places, and activities; they allowed me to see the school context through the eyes of the students themselves.

Whilst the majority of the ethnographic chats were held during break and lunchtimes within the school day, some occurred after school at my third main research site, a local youth club. Other more ancillary sites included participants' homes, or other community spaces such as the local library. This flexibility was built into the research design not only to encourage a sense of ownership and control amongst my participants, but also to limit the impact of the research process on students' day-to-day lives.

Once the participants had agreed a time and location for the ethnographic chat, I would briefly meet them to give them the prompts. Students would be asked to elect a member of the group to lead the chat (by reading the prompts). Whilst students were encouraged to talk freely, the lead student was asked to occasionally bring the group back to the prompts. In practice, the discussion that led on from each prompt would at some point naturally wane and the lead participant would read the next prompt.

The chats were recorded using a voice recording app on a mobile phone, normally belonging to one of the participants, in order to minimise conspicuousness, and maximise flexibility in terms of location and timing. The recorded chat was then sent over to me and permanently deleted from the participant's phone. On reflection, using my own phone, or other recording device, may have given greater data security, and lessened the risk of accidental leaks contravening their consent.

As I have discussed, ethnography conventionally meshes observational data with more focussed, question-based methods. What then is distinctive about ethnographic chats? Ethnographic chats were developed by drawing on established methods such as the ethnographic interview (Spradley 1979), semi-structured interviews and focus groups, all of which are traditionally researcher-facilitated. The chats employed here were characterised by specific procedural and interactional characteristics of

frame and genre, which differentiated them from both the ethnographic interview and the focus group. I expand on these differences below.

4 Researcher involvement

In designing ethnographic chats, I sought to unite the best of informal participation and formal recordings. In other words, to blur the boundary between the two. The first point of departure from a more traditional researcher-facilitated approach was the level of involvement from the researcher. Once prompts had been given to the students, I had little or no involvement, choosing instead to leave the room/space.[6] Therefore, follow-up questions were initiated by the students themselves and in this sense the ethnographic chats resembled a conversation in that students were free to bring in new topics, and to signal a change of topic.[7] Extract 1 gives an indication of how this worked in practice.

(1) English-medium school, sixth-form (aged 16–18)[8]

```
Will:       ok (.) so shall we talk about what we think our school
            thinks about language?
David:      you can speak whatever you want
Will:       yeah
David:      sometimes you can speak to a teacher and you won't
            understand (.)can say that I don't know what you mean(.)
            but they won't
Will:       they won't tell you off
David:      won't tell you off or anything
Researcher: so I guess (.) would we say then that they (.) the
            teachers are quite laid back (.) very laid back
```

[6] In some of the more public spaces this was not always possible. In these cases I would move away from the discussion and engage in other activities (such as reading a book, listening to music or working on my computer). It is also worth noting that, given my long-term engagement with the community, on some occasions students sought me out to ask me a question and I would therefore, at times, become briefly involved in the chats.

[7] Sahlstein (2004) put forward the notion of a "couple interview" where two adults, in a relationship, came together to discuss a series of written prompts without the presence of the researcher. Crucially, participants were instructed to stick closely to the pre-written questions and were not given the freedom to introduce new topics. The ethnographic chats put forward in this text differ in that participants were allowed and encouraged to introduce new themes and topics for discussion.

[8] See Appendix A for transcription conventions.

David: yeah (.)just because it's a bilingual school they have to promote both languages
Alice: yeah it's up to you (.) but I don't know how much they actually want to promote Welsh or whether they have to (.) I don't know whether there's some sort of financial gain or something?
David: what do you mean?
Alice: well like (.) if they had a real choice would they just do everything in English(.)which is obviously the most important language in the world (.) it sometimes feels that they are doing the whole Welsh thing to please someone else
Will: I'm not sure I agree
Chloe: loads of the teachers here are first language Welsh speakers so of course they feel it's important (.) some of them are really passionate about the subject and the language

So whilst I was able to maintain a focus consistent with my research themes through the use of prompts (lines 1–2 of the above extract), their loose structure (consistent with ethnographic principles) allowed discussions to flow and develop. Two follow-up questions were posed, one by Alice and one by David (lines 14–17 and line 18). Participants were free to explore the topic in whatever depth they chose, without checking or clarifying from me. Additionally, as seen in the above example, participants were able to build alignments and dissociations (e.g. line 24) with each other relative to the topic of the prompt. Overall, the chats resembled both the purposeful questions of ethnographic interviews and the emergent questions of a conversation. Meanwhile the greater distance between myself, as the researcher, and the participants gave them more autonomy and freedom to speak, and de-emphasised my role.

(2) Year 10 (aged 14 and 15), Welsh-medium School

Megan: OK so let's talk about what our school thinks about language (1.0) well Ysgol Arnant is a Welsh school and if you speak English they'll (.)the teachers (.) be like "speak Welsh" (.) "*siarad Cymreig*"[9]
Harri: yeah we're not supposed to speak English at all (.) we speak more Welsh than English
Ffion: "speak Welsh"
Harri: yeah but we can speak it outside of class (.) well I do

[9] Translates as 'speak Welsh'.

9 Megan: but why do you? (.) is it just to piss the teachers off?
10 Harri: I guess (.) but like I'll do what I want in my own
11 time (.) none of their bloody business
12 Ffion: but that's why you're always in trouble with the
13 gogs (.)(*laughter*)
14 Harri: well the goggy teachers should just fuck off and
15 realise that there's more important things in life (.)
16 people in the world are starving and they're worrying
17 about a little old language
18 Megan: bit harsh innit?
19 Harri: yeah probably (.) I'm just a bit sensitive at the moment
20 because I feel like I'm always in trouble

The above extract illustrates that the prompts allowed for an analysis of reported language practices and discursive understanding of these practices, while allowing for a degree of naturally occurring speech. This in turn enabled analysis of "ideologies in action" (Jaffe 1999), what young people actually do, conversationally, in ways that sometimes allow ideological values to leak through. In the above extract we see quite clearly that Megan, Harri and Ffion begin by articulating the more official school policy (that of separate bilingualism – see Selleck 2013) (lines 1–7). They identify that the school constructs and implements linguistic norms, understood as part of the school's political and nationalist mission, embedded within a minority struggle for power. The group dynamics allowed for a snowballing effect, with one observation initiating a chain of additional comments. From line 8, the topic shifts slightly and we see discussion turn to why one student (Harri) fails to conform to the school's expectations with regards to language use and language choice. Here we see the girls' own ideological values coming to the fore. The ethnographic chats allowed young people to express themselves using their own informal shorthand and in-jokes, without concern for my comprehension.

Based on my earlier in-depth ethnographic observations, I felt assured these insights would not have arisen with more explicit involvement from me; but nor would I have gained these insights from entirely undirected observation alone. Ethnographic chats provided the best of both.

5 Conclusions

An ethnographic chat may be a different, and in some contexts, better way to combine participant observation with more structured recordings. They allow for an element of structure without compromising participants'

freedom to elaborate on topics of interest to them. I feel there is clear potential for the use of ethnographic chats in other contexts. O'Rouke's (2011: 332) research with Irish undergraduate students, for example, used focus groups in order to access "collective discourse practices with a high degree of spontaneity". Rich insights certainly emerge from this; however, the discussions were facilitated by an Irish language tutor from the University who had previously taught many of the participants. Did students feel they could talk openly, without judgment, given the presence of their tutor? Or was there an element of favourability bias? O'Rouke (2011: 333) also notes that these focus group discussions were conducted in Irish. Again, participants may have felt some constraint on their choice of language (or indeed the choice to code-switch) given the presence of their Irish language tutor. It is impossible to say; but that is precisely my point. Developing the ethnographic chat enabled me to find new spaces to experiment in, to allow new insights to arise; and I think the same approach could allow other researchers similar new perspectives.

To reprise my overarching theme, ethnographic chats offered me the best of both worlds, and I think they could do the same for others. They combine the strengths of other qualitative methods; the open and enquiring questions of an ethnographic interview, the overlapping contributions of a focus groups but crucially without the potentially diluting or distracting influence of the researcher. An added benefit is in enabling researchers without competency in the community language(s) to work alone without an interpreter, while offering genuine linguistic choice to research participants.

Last, but by no means least, ethnographic chats empowered my participants to decide where to take the research. They were able to introduce new topics, and through their discourse, also introduced me to new participants. I was ultimately able to hold on to the research aims and expectations through the use of prompts; but the participants had a form of ownership not otherwise available.

Let me close by re-emphasising the wider context of ethnographic chats. Ethnographic chat data enriched my understanding of key topics, but this was underpinned by a much longer and more traditional ethnographic process of participant observation and careful collation of field notes. Nevertheless, I do see potential for certain principles of ethnographic chats to be adapted in less immersive research contexts. For example, a focus group could begin by de-emphasising the role of the researcher, assigning one member as chair and giving prompts for key themes. This would

probably require at least some prior warming up of the participants, getting them used to the format; but it could be done, and may well provide more transparent insights than researcher-led focus groups. As I say, my own research insights relied equally on longer term observation, but I see scope for elements of ethnographic chats to be useful in other contexts.

Transcription key

Name:	the research participant's pseudonym name
(.)	an untimed, short pause
(3.0)	a timed pause, in seconds
Speech	transcribed speech
[text]	clarification
(*text*)	commentary
"speech"	voiced speech
Cymraeg	text in Welsh

References

Atkinson, Paul & Coffey, Amanda & Delamont, Sara & Lofland, John & Lofland, Lyn. 2007. *Handbook of ethnography*. Los Angeles: Sage.

Bloor, Michael & Frankland, Jane & Thomas, Michelle & Robson, Kate. 2001. *Focus* groups in social research. London: Sage.

Gamson, William. 1992. *Talking politics*. Cambridge: Cambridge University Press.

Gregory, Kathleen. 1983. Native-view paradigms: Multiple cultures and culture conflicts in organisations. *Administrative Science Quarterly* 28. 359–376.

Hammersley, Martyn. & Atkinson, Paul. 1993. *Ethnography: Principles in practice*. London: Routledge.

Hammersley, Martyn. 2006. Ethnography: Problems and prospects. *Ethnography and Education* 1. 3–14.

Heller, Monica. 2011. *Paths to post-nationalism: A critical ethnography of language and identity.* New York: Oxford University Press.

Hymes, Dell. 1972. On communicative competence. In Pride, John & Holmes, Janet (eds.), *Sociolinguistics: Selected readings*, 269–293. Harmondsworth: Penguin.

Jaffe, Alexandre. 1999. *Ideologies in action: Language politics on Corsica*. Berlin: Mouton de Gruyter.

Malinowski, Bronislaw. 1922. *Argonauts of the Western Pacific: An account of native enterprise and adventure in the Archipelagoes of Melanesian New Guinea.* London: Routledge.

O'Reilly, Karen. 2012. *Ethnographic methods*. London: Routledge.

O'Rouke, Bernadette. 2011. Whose language is it? Struggles for language ownership in an Irish language classroom. *Journal of Language, Identity, and Education* 10. 327–345.

Roberts, Celia & Byram, Michael & Barro, Ana & Jordan, Shirley & Street, Brian. 2001. *Language learners as ethnographers*. Clevedon: Multilingual Matters.

Press, Andrea. & Cole, Elizabeth. 1999. *Speaking of abortion: Television and authority in the lives of women*. Chicago: The University of Chicago Press.

Sahlstein, Erin. 2004. Relating at a distance: Negotiating being together and being apart in long-distance relationships. *Journal of Social and Personal Relationships* 21(5). 689–710.

Spradley, James. 1979. *The ethnographic interview*. Belmont: Wadsworth.

Suter, Elizabeth. 2000. Focus groups in ethnography of communication: Expanding topics of inquiry beyond participant observation. *The Qualitative Report* 5(1). 1–14.

Selleck, Charlotte. 2013. *A comparative ethnographic study of students' experiences and perceptions of language ideologies in bilingual Welsh/English education: Inclusive policy and exclusionary practice*. Cardiff: Cardiff University. (Doctoral Dissertation.)

Sherman Heyl, Barbara. 2001. Ethnographic interviewing. In Atkinson, Paul & Coffey, Amanda & Delamont, Sara & Lofland, John & Lofland, Lyn (eds.), *Handbook of ethnography*, 369–383. Los Angeles: Sage.

Winchatz, Michaela. 2006. Fieldworker or foreigner? Ethnographic interviewing in nonnative languages. *Field Methods* 18(1). 83–97.

Contact information:

Charlotte Selleck
University of the West of England,
Frenchay Campus
Coldharbour Lane
Bristol BS16 1QY
United Kingdom
e-mail: Charlotte(dot)Selleck(at)uwe(dot)ac(dot)uk

Mikhail Mikhailov & Robert Cooper. 2016. *Corpus linguistics for translation and contrastive studies: A guide for research.* Routledge Corpus Linguistics Guides. London & New York: Routledge. Pp. 234.

Reviewed by Olli O. Silvennoinen

Corpus linguistics is by now an established method even in fields that require comparable and/or parallel data on multiple languages, such as translation studies, contrastive linguistics and language typology. Despite this, introductions to corpus linguistics are heavily biased towards monolingual corpus linguistics in general and English corpus linguistics in particular. The present volume is a successful attempt to close this gap.

The book is published in a series of guidebooks on corpus linguistics. This shows in the contents and design of the book, which are more practical than in most introductions to corpus linguistics (e.g., Biber et al. 1998; McEnery & Hardie 2011), though not as hands-on as Gries (2009), for instance. The book mostly teaches by example, with many of the chapters consisting predominantly of case studies. The case studies are largely based on the authors' own research, which is reflected in the fact that the vast majority of them contrast Finnish with either English or Russian (Mikhailov is a professor of Russian-Finnish translation at the University of Tampere, and Cooper has worked at the English department of the same university). However, they are generally written in a way that should make them easy to understand even for readers who do not know any Finnish or Russian. In addition to the seven chapters, the book contains a foreword, final remarks, a glossary, two appendices and an index. References are given at the end of each chapter rather than at the end of the book.

The authors cite Teubert's (1996) classification of multilingual data in corpus linguistics into parallel, comparable and translation corpora (p. 5). Parallel corpora are corpora that consist of source texts with their translations that have been aligned at the level of words, sentences, paragraphs or whole texts. Comparable corpora are datasets in different languages that are extralinguistically similar (e.g., novels in English and French). Translation corpora include translated texts; their purpose is usually to study the properties of translations (*translationese*) in comparison with non-translated texts. Mikhailov and Cooper explicitly state that their focus will lie on parallel corpora. Given the technical and practical challenges involved in compiling and using parallel corpora, this

is an understandable choice that indeed makes the book stand out from other introductions to corpus linguistics. As a result, however, the book does not quite cover all the ground that its title might lead one to assume. For instance, the "combined corpus approach" (Mortier & Degand 2009), which makes use of both parallel and comparable data, might have been useful for those readers embarking on research projects in contrastive linguistics. Similarly, translation studies also uses corpus linguistics to uncover features of translationese or translation universals using translation corpora (see papers in, e.g., Mauranen & Kujamäki 2004). While these issues are touched upon, the use of comparable and translation corpora in translation studies and contrastive linguistics is largely outside the scope of the book.

However, what the book loses in coverage, it gains in coherence. The steps of each analysis are presented clearly, and as a result the reader gets a well-rounded picture of what parallel corpora are about in translation studies and contrastive linguistics and what is special in them when compared to traditional monolingual corpora.

Chapter 1 ("Parallel text corpora: a general overview") covers basic issues in corpus linguistics and the use of parallel corpora. It considers such issues as the different types of corpus and the problems in using a corpus. The authors also devote a section to the use of corpora in translation and conclude that building parallel corpora is beyond what most translators would want to invest their time in. The chapter presupposes some corpus-linguistic terminology, which may not be familiar to absolute beginners in the field, such as *type*/*token ratio*, *collocate* and *concordance*.

Chapter 2 ("Designing and compiling a parallel corpus") walks the reader through the stages of compiling a parallel corpus, from planning the corpus design and inputting the texts, through aligning and annotating the texts, to storing the corpus. It ends with a discussion of copyright issues relevant to corpus linguistics. This is a highly useful chapter that contains a lot of essential information and food for thought even for those who use a ready-made corpus.

Chapters 3, 4 and 5 are the heart of the book. They offer lucid and interesting examples of the basic ways of doing research in corpus linguistics. The chapters are ordered from the most elementary to the most advanced. Chapter 3 ("Using parallel corpora: basic search procedures") covers the very basics of corpus-linguistic research, with sub-sections devoted to conducting corpus searches, concordances, frequency lists and collocations. As well as showing how to perform these methods with

parallel corpora and what they are used for, the chapter introduces basic concepts in corpus linguistics, such as *precision* and *recall*, *n-grams* and *KWIC*. On reading the chapter, even a novice should have a good idea of what the basic procedures are and why they matter. The only problem that I would like to raise is that precision is defined in a misleading way: the definition given by the authors suggests that precision is the proportion of false positives, while actually it is the proportion of true positives.

Whereas Chapter 3 covers ground that is common to all textbooks of corpus linguistics, Chapter 4 ("Processing search results") moves to issues that are specific to parallel corpora, taking the basic search procedures one step further. It consists of four sub-sections, each of which is structured around a case study. The first sub-section concerns comparing translation equivalents in parallel concordances, probably the first thing most people would use parallel corpora for. The case study concerns the Russian adverb *pravda* 'actually, really' and its translation equivalents in Finnish. The reader is shown the process of querying the data, removing noise and categorising the tokens. The section also considers the possible effect of translators' preferences for given equivalents. The only problem with the section is that the Finnish translation equivalents are not translated into English in the running text, which may make the discussion somewhat hard-going for readers who are not proficient in Finnish and/or Russian (a rare problem in the book, which generally manages to convey the meanings of Finnish and Russian data quite well). The second sub-section shows how a similar study may be done using frequency lists as the starting-point. The case study for this section concerns the English verbs *say* and *tell* and their Finnish translations *sanoa* and *kertoa*. The authors show that genre-based translation preferences can be discovered using frequency lists rather than concordances.

The third sub-section of Chapter 4 moves on to a more fine-grained analysis by considering collocations. This time, the case study is on the English adjective *clear* and its Finnish equivalents *kirkas*, *selkeä* and *selvä*. The head nouns of the Finnish adjectives are categorised according to semantic domain, which reveals patterns of usage that are not often captured even in monolingual dictionaries. The fourth and last sub-section concerns the seemingly incongruous topic of parallel corpora in monolingual studies. Using English *before* as illustration, the section shows how the French translation of the word may be used for teasing apart locative uses from temporal ones since French makes a lexical distinction between the two (*devant* for locatives, *avant* for temporals). The use of

locative *before* instead of *in front of* appears to be highly context-sensitive, as body parts (*before my eyes*) and archaic genres such as legislative texts (*before the jury*) strongly favour its use. Appropriately enough for a textbook, this highlights the fact that the selection of data is of paramount importance in corpus linguistics. Legislative texts are often used in parallel corpus studies because of their easy availability even though they might not represent modern written language very well.

Chapter 5 ("Using parallel corpora: more advanced search procedures") moves into statistical analyses common in corpus linguistics. The chapter opens with a general discussion of whether a researcher should use statistical techniques or not, and how to go about them if one does. Various options of treating quantitative data are presented and evaluated, but the authors advocate using either desktop database software (e.g. Microsoft Access) or statistical programme packages (e.g. SPSS, R). After the generalities, most of the chapter is structured around concrete research problems and case studies exemplifying how they should be solved, as in Chapter 4. The first of these problems is checking the reliability of corpus data. The case study concerns the representation of various time periods in the literary Russian-to-Finnish part of the ParRus corpus. The second quantitative theme in Chapter 5 is measures of central tendency, to which the authors dedicate three case studies. After a quick revision of measures of central tendency, range and distribution, the section moves to case studies on sentence length in Finnish translations of Russian short stories, the dispersion of common words in the TamBiC corpus, and lexical richness in Russian novels and their Finnish translations.

The chapter then has a brief interlude on the chi-square test of independence. While often used and beginner-friendly, the appropriateness of this test in corpus linguistics has been called to question because it assumes that the observations are independent of one another, which is seldom the case in corpus data (Kilgarriff 2005; Lijffijt et al. 2016). The discussion of statistical significance testing paves the way to a more in-depth discussion of collocations. This time the definition of collocation is statistical. Two case studies are offered on the English adjective *necessary* and its translations in Finnish, one using concordances, the other so-called trans-collocations between Russian and Finnish. Trans-collocates are "collocational relationships between the aligned sentences" (p. 131). For instance, the word *bird* would have its translation equivalent as its best trans-collocate, followed by domain-specific words such as 'fly' and 'cage'.

The last technique introduced in Chapter 5 is keyword analysis, which shows what lexical items are over-represented in a given dataset when compared to a reference corpus. The case study in this section examines how well Finnish translations of Bulgakov manage to convey the author's voice. The study usefully highlights the caveats of doing this type of analysis on morphologically rich languages.

Chapter 6 ("Applications of parallel corpora") catalogues various fields of research in which parallel corpora may be of use and provides further case studies. Each section concludes with a list of sample research questions. The chapter opens with a short section on parallel corpora as dictionaries. This is followed by parallel corpora in lexicography. The case study in this section is on the Russian word *prichina* 'reason, cause' and its equivalents in Finnish. Through the example, the authors show that a very large corpus is necessary for lexicographic purposes if one wishes to go beyond the one-word level and consider the phraseologies of words. Since parallel corpora are seldom very large, the authors conclude that they cannot be the sole method for compiling a bilingual dictionary. The same applies for the topic of the following section, terminology. Here, the authors begin by introducing linguistic "laws" such as homonymy and polysemy. While potentially useful, the exposition could at times be clearer; for instance, I did not understand why *recorder* (the musical instrument) and *recorder* (an electrical appliance that records sound) are homonyms but *party* (a festive gathering of people) and *party* (political grouping) are polysemes. The case study in this section is on the terminology of the paint and varnish industry in Finnish and Russian.

The subsequent section treats morphology and syntax through the example of the Finnish present perfect translated using the English simple past. Then it is the turn of pragmatics, which is illustrated through Finnish translations of the English discourse particle *yes*. Finally, the authors exemplify translation studies by considering the sentence positions of English *however* and its Finnish equivalent *kuitenkin*. These three case studies are somewhat similar, which highlights the porousness of the boundary between the fields in question: it is not clear why the study on the translations of *yes* is a matter of pragmatics but that on *however*/*kuitenkin* an exercise in translation studies, for instance. Indeed, it might be better to conduct cross-linguistic studies of many pragmatic phenomena using comparable corpora in lieu of or in addition to parallel corpora.

Chapter 7 ("A survey of available parallel corpora") is basically a list of parallel corpora that currently exist. The chapter includes the basic

characteristics of the corpora, such as the languages involved, size, genres included and the compilers. Such a list is obviously useful, though likely to become outdated fast.[1]

The book concludes with short "final remarks", in which the authors detail their approach to writing the book as well as motivate their choice of using examples from Finnish and Russian, even though a large share of their potential audience does not know these languages. Much of this could already have been said in the preface. The final remarks are followed by a useful glossary of corpus-linguistic terms and then by two appendices, one containing a list of MA theses written at the University of Tampere and the other giving sample programmes in PHP.

The book fulfils its function as a textbook for post-graduates and beginning researchers very well. One of its virtues is that in spite of its practical orientation, it does not lose sight of the theoretical significance of parallel corpora. It is always clear why a given feature of a corpus software is worth using. The procedures are clearly explained and motivated, and there is a clear progression from basic techniques to methodologically more advanced analyses, which build on previously covered material. It is also commendable that the book consistently guides the reader to more advanced sources on the topics covered. The book is mostly well edited, although there are a few solecisms here and there that do not detract from the content, however.

On the whole, Mikhailov and Cooper have produced an introduction to parallel corpora that is clearly written and pedagogically effective. It is required reading for everyone using or compiling parallel corpora in translation studies and contrastive linguistics, but it is useful also for students and researchers in adjacent fields such as linguistic typology and applied linguistics.

References

Biber, Douglas & Conrad, Susan & Reppen, Randi. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

[1] In English corpus linguistics, an example of such a list in the form of an electronic database that is regularly updated is the Corpus Resource Database (CoRD), maintained by the Research Unit for Variation, Contacts and Change in English (VARIENG). See http://www.helsinki.fi/varieng/CoRD/.

Gries, Stefan Th. 2009. *Quantitative corpus linguistics with R: A practical introduction*. New York: Routledge.

Kilgarriff, Adam. 2005. Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory* 1(2). 263–276. doi:10.1515/cllt.2005.1.2.263.

Lijffijt, Jefrey & Nevalainen, Terttu & Säily, Tanja & Papapetrou, Panagiotis & Puolamäki, Kai & Mannila, Heikki. 2016. Significance testing of word frequencies in corpora. *Digital Scholarship in the Humanities* 31(2). 374–397. doi:10.1093/llc/fqu064.

Mauranen, Anna & Kujamäki, Pekka (eds.). 2004. *Translation universals: Do they exist?* Amsterdam: John Benjamins.

McEnery, Tony & Hardie, Andrew. 2011. *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511981395.

Mortier, Liesbeth & Degand, Liesbeth. 2009. Adversative discourse markers in contrast: The need for a combined corpus approach. *International Journal of Corpus Linguistics* 14(3). 338–366. doi:10.1075/ijcl.14.3.03mor.

Teubert, Wolfgang. 1996. Comparable or parallel corpora? *International Journal of Lexicography* 9(3). 238–264. doi:10.1093/ijl/9.3.238.

Contact information:

Olli Silvennoinen
Department of Modern Language, General Linguistics
P.O. Box 24
00014 University of Helsinki
Finland
e-mail: olli(dot)silvennoinen(at)helsinki(dot)fi

SKY Journal of Linguistics

Volume 30:2017

Contents:

Publisher:
Suomen kielitieteellinen yhdistys
Språkvetenskapliga Föreningen i Finland
The Linguistic Association of Finland