

CODES AND CHOICES

Chapters 2–4 of this book have dealt mainly with the operation of languages at large in society – with scenarios of multilingualism, the nature and repercussions of language shift, the impetus for language maintenance, and the making and unmaking of languages as a result of drastic social change. In this chapter and the next we turn to a more micro level of enquiry, to look at how speakers choose different language resources in different circumstances, including not just what are regarded as separate languages but also varieties within those languages. This leads on in Chapter 6 to considering different facets of the interaction of speakers and audiences in conversation.

5.1 VARIETIES, CODES AND REPERTOIRES

The foundation of all sociolinguistics is that speakers have choices when they talk. Their choices range from the macro to the micro, from the wholesale choice between different languages to scarcely distinguishable alternative pronunciations of the same consonant. A macro choice can be between the Xhosa, Afrikaans or English languages for a South African. A micro choice can be between a glottal stop and a [t] consonant in the middle of the English word *butter* for a Cockney in London. ‘Language choice’ usually refers to the choice between different languages rather than micro choices within languages.

Sociolinguists use the term **variety** to cover all kinds of differences within a single language. A variety is a relatively distinguishable form of a language, often based on geographical or social differences – Chilean versus Castilian Spanish, African American versus European American English. Variety is a less loaded term than ‘dialect’, which lay usage tends to equate with substandard or exotic language forms. It also enables us to encompass concepts such as ‘genres’ or ‘registers’ of a language without having to commit

ourselves to defining them precisely. It remains a serviceable label even though, as we shall also see later, varieties have very porous boundaries.

Both entire languages and varieties within them can be called linguistic codes. The term 'code' is useful in sociolinguistics because it corralles the whole range of language resources that speakers use, whether these are regarded as distinct languages or as varieties of a language. In this book I often use 'code' as a cover term when – as so often – a generalization applies equally to whole languages or to varieties within languages. The word is used particularly in the terms 'code switching' or 'code mixing', where two or more languages or varieties are intertwined (see later in this chapter). Recall also from the discussion in Chapter 1 that although distinguishing languages from varieties is useful for some purposes, there is no absoluteness in that distinction.

The suite of codes which a speaker is able to draw on makes up their linguistic repertoire. The term may be used to describe the linguistic range of an individual or a small group, but not of larger, more diffuse entities such as a nation or ethnic group. If the codes are varieties of the same language – such as dialects of Urdu or Japanese – the speaker's repertoire is said to be 'monolingual'. If the codes are two separate languages, as for Kazakh and Russian in Kazakhstan, the repertoire is bilingual. If three languages or more are involved, such as Mandarin, English and Hokkien in Singapore, the speaker's repertoire is multilingual. However, by focusing on the linguistic resources that a speaker draws on, the concept of repertoire tends to break down attempts to make too-ready distinctions between those codes. Contemporary sociolinguistics particularly uses the idea of repertoire to cover the fluid language mixes used by young speakers in ethnically diverse urban situations and described with terms such as 'translanguaging' and 'metrolingualism' (Busch 2012).

Exercise 5.1 Class language varieties profile

Exercise 2.2 surveyed the repertoires of class members, concentrating on the different languages that they speak. Now consider all the codes in their repertoires – that is including the varieties of languages that they speak as well as the languages. List their languages again, then ask:

- 1 What varieties of those languages do they speak?
- 2 How are these varieties distinguished from each other socially – for example, are they associated with different regions, or social groups, or ethnicities?
- 3 How are they distinguished from each other linguistically? Are they very different? Do the differences lie in the vocabulary, or pronunciation, or grammar?
- 4 What labels are used to name these varieties?
- 5 What are other people's attitudes towards them?
- 6 For some of these, can it be difficult to decide whether they should be described as a language or a variety? Why?

Log the answers in a grid (languages × varieties) to construct a variety profile of the class's languages.

In some situations the resources may configure in unusual ways, with participants happily using different languages from each other in the same interaction. In the rural situation of Misión La Paz in northern Argentina, three indigenous languages are spoken (Campbell and Grondona 2010), and marriages are typically between speakers of different languages. Each person in a household speaks in their own language, and everyone understands, but replies in their own (different) language, creating a linguistic situation that the researchers consider unique.

All languages include a range of varieties, and all normal adult speakers can control more than one variety of their language. Among the most striking types of variety are the so-called 'mother-in-law' codes which are a feature of many Australian aboriginal languages, but are also known in North American and African Bantu languages. In the Dyirbal language of northern Queensland (see Example 4.3 in previous chapter), the everyday speech style *Guwal* is replaced by an 'avoidance' style, *Jalnguy*, when certain opposite-sex relatives, especially in-laws, are present. Avoidance speech is used because the everyday language is regarded as taboo in the presence of those relatives (Dixon 1983). The avoidance code tends to maintain the phonology and syntax of the everyday language but uses alternative lexical items.

5.2 THE SPEECH COMMUNITY

You cannot go far in sociolinguistics – or even linguistics itself – without meeting the notion of 'speech community', which now becomes important to our discussion. It has been applied to all scales of human groupings, from an entity as small as a jury or a single tribal longhouse, to entire major cities and even to all women. Not surprisingly, it has also been defined by many sociolinguists and in different ways – Exercise 5.2 shows some of these. The temptation is to jettison a term which has such a broad and inconclusive range of definitions. Nevertheless, many research articles begin by laying out their authors' idea of what a speech community is for their particular study. Although nobody has been able to define it conclusively, it remains a widespread term which sociolinguistics seems to need. Alternative concepts have also been suggested, particularly social networks and community of practice, which we deal with in Chapter 8.

The notion of the speech community goes back at least as far as Leonard Bloomfield, whose 1933 book *Language* was the foundational text of American linguistics. The concept was picked up into sociolinguistics in the 1960s by three of the field's founders, Hymes, Gumperz and Labov. Each had his own inflection on the concept, reflecting his particular research interests – 'multilingualism for Gumperz, linguistic evaluation and style-shifting for Labov, ways of speaking and communicative competence for Hymes', as Peter Patrick lucidly summarizes it in an overview article (2002: 575).

Perhaps the most basic and intractable issue is whether the speech community is based on social or linguistic criteria – or both. Gumperz's was the original definition and was largely social (1962). It assumed some cohesion among members as well as a shared linguistic repertoire, although not a single language. Later he added the linguistic criterion that the speech community is differentiated from adjacent groupings by its language usage (Gumperz 1968).

Exercise 5.2 Defining 'speech community'

Here are some of the definitions that sociolinguists have given of 'speech community'. Clarify what each of them means, compare their similarities and differences, evaluate their strengths and weaknesses, and arrive at your own definition.

- Bloomfield 1933:

A group of people who use the same system of speech-signals is a *speech-community*. Obviously the value of language depends upon people's using it in the same way ... A speech-community is a group of people who interact by means of speech. (pp. 29 and 42)

- Gumperz 1962:

A social group which may be either monolingual or multilingual, held together by frequency of social interaction patterns and set off from the surrounding areas by weaknesses in the lines of communication. (p. 31)

- Gumperz 1968:

Any human aggregate characterized by regular and frequent interaction by means of a shared body of verbal signs and set off from similar aggregates by significant differences in language usage. (p. 381)

- Hymes 1974:

A social rather than a linguistic entity. One starts with a social group and considers the entire organization of linguistic means within it. (p. 47)

- Hymes 1972:

A community sharing rules for the conduct and interpretation of speech, and rules for the interpretation of at least one linguistic variety. (p. 54)

- Labov 1972b:

The speech community is ... defined ... by participation in a set of shared norms. (pp. 120–1)

- Labov 1966:

New York City is a single speech community, united by a common set of evaluative norms, though divergent in the application of these norms. (p. 355)

- Kachru 2001:

A speech community crosses political boundaries ... does not necessarily represent one religion or culture ... comprise[s] idiolects and dialects. (p. 105)

- Bolinger 1975:

There is no limit to the number and variety of speech communities that are to be found in a society. (p. 333)

Hymes (1962) uses the term but initially does not define it, preferring 'the speech economy of a community'. He leaned explicitly towards the speech community being socially not linguistically based (1974): the researcher begins with a social grouping and then examines their linguistic repertoire. However, elsewhere Hymes implies the speech community is defined by language-related criteria (1972).

For Labov, shared linguistic norms were central (1972b), although his work was based in a geographically defined area – the Lower East Side of New York City. He found that community members all evaluate linguistic features similarly and behave accordingly in using them (see my examination of Labov's New York study in Chapter 7.1).

It seems, then, that for these three scholars, both social and linguistic criteria are needed, but that generally the social comes first. What is true is that in actual research practice, sociolinguists tend to start from a social basis, then add linguistic characteristics. They identify some grouping – a gang, profession, village, chatroom, neighbourhood, nation – and research its linguistic repertoire and behaviour. But in looser discussion about large-scale language commonalities, they will be more liable to use shared linguistic repertoire as a criterion for treating something as a speech community – the francophone community of Canada, the German-speaking community in Australia, or even 'the speech-community which consists of all English-speaking people' (Bloomfield 1933: 42). In these cases, the constituency of people involved is so scattered and diverse that most social criteria for recognizing them as a community would fail (although see the very broad definition by Kachru 2001, Exercise 5.2). But overall, Patrick's summarization seems apt, that the speech community is 'a socially-based unit of linguistic analysis' (2002: 577).

The speech community concept must allow for difference and divergence as well as for commonality or it becomes inoperable (see the rider in Labov's 1966 definition). The knotty question remains of how much difference is tolerable before the boundaries dissolve (cf. Bolinger's definition, 1975). A second and linked issue is how different speech communities relate to each other. Patrick (2002) suggests that there are at least two kinds of relationships among speech communities:

- Nesting, where one community sits inside another. A New York youth gang may be a speech community in its own right, but it also nests within a community definable as all such gangs in New York, which again is nested within the speech community of the city as a whole.
- Overlapping, where different speech communities partially cross with each other. In contemporary London, immigrant youths from 20 or more ethnic backgrounds are members both of their separate groups of origin and of wider, pan-ethnic networks, as studies of multicultural English there show (Cheshire et al. 2011).

Unlike many early sociolinguistic concepts, the idea of speech community has not progressed much since it was proposed. It has been criticized for its indefinability, deconstructed and set aside on theoretical grounds (e.g. Rampton 2009), but still often comes back into discussions as a working concept which scholars cannot do without. It has also been challenged on ideological grounds. Labov's definition of shared norms in New York City has been regarded as assuming the existence of agreement between /disparately placed social groups. Milroy and Milroy (1992) argue that a model of society

Exercise 5.3 Speech communities near you

What speech community or communities do you individually belong to?

- 1 What are the social characteristics of your speech community/ies? What are the linguistic characteristics?
- 2 Do you find that the primary criteria are social or linguistic?
- 3 How much diversity is there within your speech community/ies?
- 4 If you belong to more than one speech community, what different kinds of configurations are there? – mono/multilingual, large/small, permanent or shorter-term, based on geography or age, etc.
- 5 How do these speech communities relate to each other? Are they, for example, overlapping or nested?
- 6 Pool the speech community profiles of all class members, and evaluate the conceptions and definitions they lead to.
- 7 Is this sociolinguistics class itself a speech community? Why/not?
- 8 Consider the question: What *isn't* a speech community?
- 9 In the light of the above discussions, revisit the definitions in Exercise 5.2 – what definition of 'speech community' emerges from your consideration?

that takes account of conflict between groups is more appropriate than a consensus model. Given this level of debate, focus has shifted to alternative concepts that have been proposed for dealing with the nature of linguistic community. These include the Milroys' Social Network model (e.g. Lesley Milroy 1987), and the Community of Practice framework associated in sociolinguistics with Penelope Eckert (e.g. 2000). We return to these approaches in Chapter 8.

5.3 DIGLOSSIA

The concept of diglossia has to do with the way in which different language codes tend to be stratified in a society. The term dates from the earliest days of sociolinguistics. It derived from the French *diglossie* (itself taken from the Greek for 'bilingualism'), was adopted into English by the American Charles Ferguson, and later extended by Joshua Fishman and Ralph Fasold.

In 1959 Ferguson published an article with the one-word title 'Diglossia'. It became the most cited and influential single paper in the history of sociolinguistics. Even older bibliographies (e.g. Fernández 1993) show literally thousands of publications that reference the concept, which has been applied to language situations all over the world. Central to diglossia is the concept of differential language functions that we first met in Chapter 3.2 in relation to multilingualism. Ferguson observed that there exist societies which have two related varieties of a language that are used in quite different sets of functions. These varieties and functions can be stratified into 'Low' and 'High' and are in complementary distribution. The Low variety is used in everyday functions such as at

home, in the market and in conversation among friends. The High language is used in prestige functions such as education, media and government. The two languages are clearly related linguistically, but rather distantly.

Classic diglossia

Ferguson offered an extended definition and four defining cases of diglossia:

- Greek has Demotiki as the Low, everyday, local form, alongside Katharevousa, Classical Greek, which is the High, prestige, universal form.
- The Caribbean nation of Haiti has the local Haitian Creole that has grown up there, based partly in French and partly in African languages, alongside standard European French.
- In Switzerland there is Schwyzertütsch, Swiss German, which is a cluster of local dialects. Superposed is the pan-Germanic standard, Hochdeutsch or High German.
- Across the Arab world, varieties of colloquial Arabic are spoken alongside Classical Arabic. Colloquial Arabic is not itself a single variety but differs greatly from place to place, e.g. between Jordan and Morocco. Classical Arabic is the language of the Koran and is constant. Despite differences among varieties of Colloquial Arabic, Colloquial and Classical Arabic stand everywhere in a constant relation to each other.

Defining classic diglossia

Diglossia is a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation.

(Ferguson 1959: 336)

Characteristics:

- the High (H) and Low (L) languages are related
- there are complementary sets of functions or domains for H and L
- H has overt prestige
- H has an admired literary heritage
- L is acquired naturally from the start
- H is book-learned later
- H is standardized in grammars/dictionaries
- H is more linguistically complex than L

In all four of these cases, the two codes are distinct enough to jeopardize comprehension of one code for a listener who knows only the other. They tend to be labelled as the same language but are divergent enough to have characteristics of different languages.

Ferguson laid out characteristics he saw in diglossic situations (Example 5.1). The H language has overt prestige, it is openly regarded as the good language. The L language may be regarded – including by its own speakers – as a lesser language. Communities may even deny the existence of the L language. In some areas H has a more complex grammar than L. It has a literary heritage of classically established works which form a cultural deposit and serve as standard-setters for the language as a whole. Often the epitome of the literary heritage is the society's primary religious text. In Arabic, the standard is the Koran itself. In German the standard was the Lutherbibel, reformer Martin Luther's original translation of the Bible (1534).

The H language possesses codifying prescriptive texts such as dictionaries and grammars. The L language may not have these, or may acquire them much later, and its spelling may not be stable. Thus in Switzerland there may be a full slate of reference works for High German, but relatively little on Schwyzertütsch. And whereas children grow up naturally with L and learn it as their mother tongue, H is generally learned as a subject through school. It is superposed and often external. Its home base may be

Exercise 5.4 Diglossia in the twenty-first century

Towards the end of his 1959 article, Ferguson hazarded a prediction for each of the four cases in two centuries' time (to the year 2150):

Swiss German	Relative stability
Arabic	Slow development towards several standard languages based on regional L varieties
Haitian Creole	Slow development towards a unified standard based on the L language of the capital, Port-au-Prince
Greek	Unified standard based on the L language of Athens.

Half a century since Ferguson's work, how much has changed, and is the change in the direction he expected?

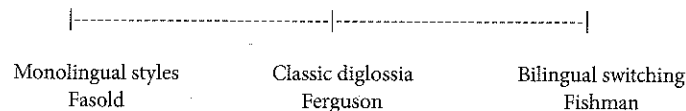
- 1 Search out more recent information on one of the four cases (e.g. Frangoudaki 2002 for Greek).
- 2 Compare it to Ferguson's original presentation.
- 3 Assuming the validity of Ferguson's original description for this case, does diglossia still operate in the same way there nowadays? What if anything has changed and how?
- 4 Is the situation still diglossic in Ferguson's sense?
- 5 Compare the outcome after some 50 years with Ferguson's long-term prognosis above. Is the case developing in the direction Ferguson envisaged or not?
- 6 Do any differences in the current situation challenge the basic concept of diglossia?

geographically outside the society itself (e.g. in France not Haiti) or be identified with an earlier historical period (e.g. classical Greece). Diglossia may also be stable over a very long period. In Switzerland, Hochdeutsch and Schwyzertütsch have co-existed in a diglossic relationship for centuries – although in the period since Ferguson's article, Schwyzertütsch has made inroads into prestige domains.

Fishman and Fasold extend diglossia

In the 1960s, Fishman brought together diglossia with his own sociology-of-language concepts. He noted (1967) that the stratifying of the H and L functions also occurs in societies where unrelated languages co-exist. That is, languages which are linguistically quite unrelated may still be in a socially diglossic relationship of H and L to each other. One of his examples was Spanish and Guaraní in Paraguay. These undoubtedly stratify socially in a High/Low relationship, although one is a native South American language and the other European. Fishman therefore argued for extension of the concept of diglossia to bilingual situations where any two languages are stratified.

Later still the American sociolinguist Ralph Fasold (1984) applied the term to the stratification that also occurs within single languages. Even when the linguistic codes involved are closely similar varieties of the same language, there may be a diglossic High-Low relationship between them. A typical example is that of a standard version of many of the major European languages as used in public speaking, alongside the vernacular variety used in conversation. Fasold therefore proposed extending the concept of diglossia yet again to include all situations where language codes stratify socially, regardless of the linguistic relationship between the codes. Thus repertoires which are monolingual (Fasold), bilingual (Fishman) or somewhere in between (Ferguson) could all be labelled diglossic if they fit the stratifying criteria, forming a continuum like this:



My view of these extensions is that, while I agree on the commonalities across the three cases, it makes for an unhelpful dilution of Ferguson's originally much more sharp and specific concept. I prefer to reserve the label 'diglossia' for the classic in-between language situation that Ferguson defined. Where parallel processes are evident in fully monolingual or bilingual situations, I would refer to those as 'in a diglossic relationship' rather than actual 'diglossia'.

Diglossia: caveats and critiques

Diglossia has been a very productive concept for sociolinguistics. Ferguson's original article is full of lucid and insightful generalizations and has been republished many times (e.g. in Coupland and Jaworski 2009a, volume 4). However, both the

concept of diglossia, and its definition and application, raise serious issues and have been challenged:

- Scholars who are specialists in the four defining cases have questioned Ferguson's characterizations of them (e.g. Kaye 1972 asked how stable Arabic diglossia is).
- The situations in the four cases have changed over time (which Ferguson assumed). His description is now more than 50 years old and may no longer apply (see Exercise 5.4).
- Much more basically, the dichotomizing involved in diglossia is questionable (Dorian 2002). Sociolinguistic situations are scarcely ever clear-cut in this way, and the labeling involved in diglossia implies too static and monolithic a situation to suit the on-the-ground reality.
- The model presents the 'High/Low' terminology as a very uncontested, commonsense view, but the labels are not neutral (Exercise 5.5).
- Overall the description of diglossia reflects an orientation to these sociolinguistic situations which reflects and serves the status quo (Williams 1992). The siting and valuing of the language codes is uncoupled from the social position and evaluation of their speakers, which treats a very political situation as apolitical.

These are real and principled objections of a kind which are addressed in some of the attached exercises and which we will return to later in discussing language ideologies

Exercise 5.5 Apply and critique diglossia

Apply the concept of diglossia in your society:

- 1 Are any linguistic codes in your society in a diglossic relationship?
- 2 Which of Ferguson's defining characteristics do they meet?
- 3 Which kind of diglossia is this? Or is more than one kind involved? Or does it not really fit any of these three?
 - Ferguson's classic diglossia
 - Fishman's bilingual diglossia, or
 - Fasold's variety-based diglossia.

Now critique diglossia as a concept:

- 4 How much does 'diglossia' further our understanding of how languages relate in your society?
- 5 What is gained and what is lost through Fishman's and Fasold's extensions to Ferguson's original concept?
- 6 Some subsequent researchers have found Ferguson's 'High' and 'Low' labels 'unfortunate' (Williams 1992: 95) and 'invidious' (Dorian 2002: 64). Why? Can you suggest alternative labels that do not create such problems?
- 7 Evaluate the characteristics and definition that Ferguson offered (Example 5.1). Remembering the critiques in the text of the status-quo stance which is implied by diglossia, what alternative ways are there of thinking about the relationship between these kinds of language codes in a society?

(Chapter 10). Despite the problems, I myself find that 'diglossia' captures a basic sociolinguistic generalization about a near-universal relationship between linguistic codes in societies. However, we need to be constantly aware of and question the social and political positioning and power that underlie such a situation, and not just adopt descriptive phrasings that endorse linguistic inequities.

Diglossia continues to be frequently referenced and forms the basis of some sociolinguistic work in the twenty-first century (e.g. Snow 2010), but probably less than formerly. In origin it was a very structuralist/functionalist concept, in accordance with American social science of its time, and is less congenial to the fluidity of postmodern approaches. The issues of language choice which the concepts of repertoires, domains and diglossia seek to address underlie most of the matter in the rest of this chapter – and indeed in later chapters in the book. Understanding the social embedding and significance of language choices is arguably the key question in sociolinguistics, and we will address it in a number of ways as the book develops, including in relation to the concept of 'style' (Chapter 11) which overlaps with Fasold's monolingual diglossia.

5.4 CODE SWITCHING

Code switching is one of the choices available to bilingual and multilingual speakers. It occurs when speakers switch backwards and forwards between distinct codes in their repertoire, often within the same sentence or utterance. It is a complex and skilful type of language choice, and involves the accomplished handling of two or more languages simultaneously – structurally, psychologically and socially. Code switches often carry a lot of social meaning, and attempting to interpret and explain them can be a challenge.

Code switching is simultaneously one of sociolinguistics' most interesting and most demanding areas because of the complexity of the linguistic structures it creates. The challenges begin with radical disagreement over what counts as code switching. Definitions will often differ even between different authors in the same publication. Milroy and Muysken (1995: 12) record that a major European research collaboration and its ensuing publications had to abandon the attempt to find consensus on code switching terminology. One distinction that many researchers agree on is between inter- and intra-sentential switching – what occurs between sentences versus within the same sentence. Switching inside the one sentence is also sometimes distinguished as 'code mixing.' The quantitative sociolinguist Shana Poplack who studied Spanish–English switching in New York (1980) recognized a third category, 'tag-switching', when the switch occurs on a sentence tag, like *you know* in English. A basic issue is what stretch of language should be taken into account – the sentence, the utterance, the turn, the conversation?

Much of the research on code switching focuses on the structural linguistics of switches. One approach comes from the American contact linguist, Carol Myers-Scotton. The Matrix Language Frame model was put forward in her 1993(a) book *Duelling Languages*, and developed in later publications particularly with Janice Jake (e.g. 2009). Myers-Scotton's main contention is that in code switching, one or other language will always be dominant. This is called the 'Matrix Language', and it sets the structural frame of a code-switched sentence: the order of elements will be that of the matrix language, which also provides all the necessary structural material. The other

Code switching in Kenya

This example comes from two students in Nairobi discussing their school work in Swahili and English (Myers-Scotton 1993a: 79). The hyphens identify bound morphemes, those that cannot stand alone.

Leo siku-*come* na *books* zangu
 leo si- -ku- -*come* na *books* z- angu
 today 1S/NEG PAST/NEG come with books CL my
 10

'Today I didn't come with my books.'

Myers-Scotton treats Swahili here as the matrix language, and English as the embedded language. The speaker's first word is in Swahili, then there are switches back and forth on each of the remaining four words of the sentence. English provides two content words – the main verb *come* and noun *books*. The structural material comes from Swahili: the first person singular pronoun *si-*, past tense marker *-ku-*, negation marking, preposition *na*, and the prefix *z-* on the possessive *angu* (identifying the English noun *books* as belonging to class 10 out of the approximately 15 noun classes of Swahili).

language is the 'Embedded Language', which provides only content material. Although widely used, the Matrix Language Frame model has received many challenges from other scholars (e.g. Auer 2007). They point out that the matrix language can change from sentence to sentence, and it may not be easy to decide which language is in fact the matrix or how all morphemes should be classified. Example 5.2 shows a sentence with four code switches, drawn from Myers-Scotton's extensive Swahili-English research in Kenya. Extended switching among several languages simultaneously is by no means unusual, as shown for example by Migge's research (2007) on the languages of the rainforest of Suriname and Guyane in the north of South America.

5.5 THE SOCIOLINGUISTICS OF CODE SWITCHING

Code switching is often socially meaningful, and many sociolinguists have examined this works. It also triggers strong reactions from audiences. It tends to be denigrated by popular opinion from inside as well as outside the speech communities where it occurs, and is often regarded as a corrupted semi-language. This gives rise to pejorative labels such as Spanglish (or Espanglés), Franglais or Chinglish, which represent various mixes of other languages with English. However, the research shows that code switching is a routine behaviour in all bilingual and multilingual communities. Exercise 5.6 shows an extended passage of switching whose topic, ironically, is the practice of switching itself.

Exercise 5.6 Code switching attitudes

Here is a Panjabi-English bilingual from a study by Romaine (1995: 122) discussing code switching:

I mean I'm guilty in that sense *ke ziada wsi* English *I bolde fer ode nal eda hunda ke twhadi jeri zaban ē, na?* Odec *her ik sentence ic je do tin* English *de word honde ...* but I think that's wrong. I mean, *mə khəd čana me ke, na, jədo Panjabi bolda ē, pure Panjabi bola wsi* mix *kərde rēne ā.* I mean, unconsciously, sub-consciously, *kəri jane ē,* you know, *pər* I wish, you know *ke mə pure Panjabi bol sēka.*

I mean I'm guilty as well in the sense that we speak English more and more, and then what happens is that when you speak your own language, you get two or three English words in each sentence ... but I think that's wrong. I mean, I myself would like to speak pure Panjabi whenever I speak Panjabi. We keep mixing. I mean unconsciously, subconsciously, we keep doing it, you know, but I wish, you know, that I could speak pure Panjabi.

- In the passage identify the three kinds of code switching mentioned earlier in the text:
 - inter-sentential (at sentence or clause boundary)
 - intra-sentential (within the sentence)
 - tag switching (such as *I mean*).
- Tease out the description that the speaker gives of code switching, and the attitudes he expresses. Evaluate them.
- Consider the speaker's expressed attitudes to code switching in the light of the amount he uses in the excerpt itself.

Gumperz: interactional code switching

The way code switches operate in conversation was a focus of the American anthropological linguist John Gumperz, one of the founding figures of sociolinguistics. Gumperz was the originator of 'interactional sociolinguistics', which we will deal with in more detail in Chapter 6.6. He studied bilingual situations in places as diverse as India, Austria, Norway and California, summarizing much of his work in a 1982 book. His interest was in the operation of switches in the flow of conversation regardless of whether the codes involved were distinct languages or related varieties of the same language.

Gumperz noted that the conversational code switches he observed occurred in interactions that were as fluent and unitary as monolingual exchanges – except that they involved two languages not one. The switches might be salient to outsiders or to linguists, but the participants scarcely noticed them, being immersed in the meaning of the communication not its linguistic form. Seeking to identify the interactional triggers for switching, Gumperz (1982a) found that they include:

- introducing direct quotation or reported speech
- picking out a specific addressee

- interjections
- reiterations
- qualifying messages.

Take this English/Hindi switch by a bilingual university student in Delhi:

I went to Agra, *to maine apne bhaiko bola ki*, if you come to Delhi you must
(then I said to my brother that) buy some lunch
(Gumperz 1982a: 76)

The switch introduces a quotation – a common use, Gumperz found. A message in one code may also be repeated or modified in the other, to clarify or amplify or emphasize what has been said. This switch involves a repetition by a Chicano professional in California who generally speaks Spanish at home and English elsewhere:

I got to thinking *vacilando el punto ese* you know? I got to thinking well this and
(mulling over that point) that reason.
(Gumperz 1982a: 78)

Gumperz suggested that in bilingual situations there tends to be a *we*-code and a *they*-code – one of the languages will generally be associated with the minority and its ingroup life, and the other with the outgroup, that is the wider society (similarly to diglossia). To investigate what code switches mean in their speech community, Gumperz played back recordings of switches and asked members to interpret them. Listeners reported that flipping the order of the languages in a code switch could make a difference because of the *they/we*-code reversal. A Puerto Rican mother was heard calling to her child in a New York street:

Ven acá. *Ven acá.* Come here, you.
(Come here. Come here.)
(Gumperz 1982a: 92)

The listeners interpreted the switch to the *they*-code, English, as a warning to the child. But when the order was reversed and English came first followed by the *we*-code, Spanish, it sounded like an appeal:

Come here. Come here. *Ven acá.*

One of Gumperz's most fruitful insights was the distinction between situational and metaphorical code switching, based on research in the settlement of Hemnesberget in northern Norway (Blom and Gumperz 1972). In situational switching, there is a regular association between language and social situation, and the switch lasts as long as the situation lasts. The researchers observed that the entry of outsiders to a local group in Hemnesberget triggered switches from Ranamål, the dialect, to standard speech, Bokmål. A move from business to personal subjects also caused a switch from standard to dialect. These situational switches reflect accepted norms of what is appropriate language choice for certain audiences or topics. Metaphorical switches trade on such

regular associations, actively using language to inject the flavour of one setting into another, alien context. So in an otherwise standard language conversation, local forms were introduced to provide anecdotal colour. Conversely, standard forms surfaced as a claim to intellectual authority during a conversation conducted in dialect. Gumperz's situational/metaphorical distinction has been widely applied in the study of sociolinguistic style, which we look at more closely in Chapter 11.

Myers-Scotton: the Markedness Model

Myers-Scotton has developed an approach to the social operation of code switching which has parallels to Gumperz's. Her Markedness Model – detailed especially in her 1993(b) book *Social Motivations for Codeswitching* – classifies switches into either marked or unmarked choices. Unmarked choices are the default choices that tend to be expected, usual or frequent. They contrast with the 'marked' choices, which are more unexpected, unusual or rare. Markedness can be hard to tie down and assign, but the concept does capture something important about the social significance of language choices.

The Markedness Model maintains that many interactions carry the expectation of certain 'rights and obligations' which mean that one of the code choices is unmarked. Other choices in the situation are more marked because they do not meet those rights and obligations. Many of Myers-Scotton's examples (1988, 1993b) come from trilingual exchanges in Kenya, where English, the national lingua franca Swahili and various local languages interact. Swahili would normally be the unmarked choice for addressing another, apparently local African. If the person turns out to be from the same ethnic group, then the shared ethnic language is expected. But if the setting is a white-collar office in the capital, Nairobi, the unmarked choice is English. Note the similarity again to the approach through domains and diglossia.

The model stresses the proactive position of speakers, negotiating strategic choices based on their own interactional goals, the social situations they are in, and the wider sociolinguistic ecology surrounding them. Markedness works because societies have norms which – although they do not control the speaker's choices – frame the interactional and social consequences of those choices. All choices indicate a particular interpersonal balance between the participants. Marked switches usually redefine the identity of a speaker and her social distance from the addressee. They may, for example, exclude outsiders from an ingroup by switching to a language the outsiders do not understand.

As well as its similarity to Blom and Gumperz's situational/metaphorical distinction, the Markedness Model parallels other theories of sociolinguistic styling that we shall deal with in Chapter 11.3. Myers-Scotton argues that there are situations where code switching itself can operate as the unmarked choice. In these cases continuing code switching is the norm, with no specific social meaning adhering to the individual switches. It is the interplay of two languages that indicates the speakers acknowledge affiliations to, for example, both national and local identities. So young men in Harare, the capital of Zimbabwe, may keep on switching between English and Shona, the national language, because they want to affiliate with the prestige of English as well as the indigenous value of Shona (Myers-Scotton 1993b: 123).

Auer: code switching as practice

Gumperz's focus on how code switching is embedded in the flow of bilingual interaction, and what it means to community members, has been picked up by scholars such as the German conversation analyst, Peter Auer. Conversation Analysis is an approach to the study of (usually monolingual) conversation that I will not have space to discuss in its own right in this book. Suffice to say that it focuses on the orderly minutiae of everyday conversations through close analysis of, for example, how turns are taken.

Auer's original research was conducted with a group of bilingual Italian immigrant children in the German city of Konstanz on Lake Geneva. Example 5.3 comes from this study. Auer found that their language alternations were inter-sentential – between rather than within sentences – and proposed a four-part classification of how they operated (Auer 1988), which I diagram like this:

	Transfer	Code switch
Participant related	x	x
Discourse related	x	x

First, the alternations were either transfers (which were mostly lexical) or code switches. Transfers tended to be followed by a return to the previous language, while switches

Wait, Pino

This example is from Auer (1988: 197). 'm' is the project fieldworker, with two of the boys taking part in the study.

((m. has taken Luziano and Pino in his car to his house. The car has stopped, the three are about to get out.))

70:	06	m:	là là si apre, là sotto
	07	Lz.:	ah là
71:	01		Pino -- wilscht rau:s -- wart mal
	02		wart mal Pino

Translation (Italian in lower case, German in CAPITALS)

70:	06	m:	here here you can open it, down there
	07	Lz.:	oh there
71:	01		PINO -- SO YOU WANT TO GET OUT -- WAIT,
	02		WAIT PINO

Auer writes that Luziano's switch in line 71:01 is part of effecting a change in the 'participant constellation'. His Italian *ah là* was addressed to the fieldworker, acknowledging his instruction about how to open the car door. Luziano then switches to German to give orders to the younger boy, Pino.

usually stayed with the new language choice. For these students, German was dominant. Almost all the code switches were from Italian to German, and the transfers inserted German words into Italian discourse. Secondly, the basis of the alternations was either participant related (e.g. due to a speaker's language preference or capability) or discourse related, that is signalling what the speaker is doing in the interaction such as changing the topic. Code switching is therefore what Gumperz called a 'contextualization cue' (1982a: 131) – one of the means by which speakers signal, and listeners interpret, their interaction and its content.

The conversational take on code switching proceeds from radically different assumptions than structural approaches such as Myers-Scotton's Matrix Language Frame model. It takes bilingualism to be not a capability but a behaviour, and being bilingual as something that speakers do rather than are (Auer 1988). The unit to be analysed is the discourse itself, and code switches can be understood only in the light of the choices that have preceded and followed. This has the advantage of emphasizing that in code switching one choice leads to another.

Auer questions whether bilingual conversation can be reduced to 'two-sided monolingual talk' (2007: 321), partly because of the difficulty of telling where one language stops and the other starts in mixed discourse. What is more, speakers can intentionally use the ambiguity of an item from related languages to trigger a switch. This example from Kathryn Woolard (1999) presents the standard opening gambit used by the Catalan comedian Eugenio in the 1980s, a period when the choice between Spanish and Catalan was particularly salient:

el **saben** *aquel ...*
 him know this-one ...
 'do you know this one ...'

The first word is Catalan and the third is Castilian Spanish, but *saben* 'know' is either – or both. Woolard describes *saben* as 'bivalent' – that is, belonging to both codes. Its hybridity is the source of the humour and makes it impossible to pin down precisely where the transition between the two languages occurs. Woolard's later work looks at whole texts that are bivalent: they could be read as either Spanish or Latin, intentionally composed as a strategy for enhancing the standing of Spanish from the sixteenth century (Woolard and Genovese 2007). Recall Heller's stress (in Chapter 2) on the fluidity of languages and their boundaries.

5.6 THE CASE OF OBERWART

Susan Gal's study of the bilingual community of Oberwart is one of the classics of research into individual and community code choice. The American anthropological linguist conducted her doctoral research in this town on the Austria-Hungary political border in the 1970s. It also lies on the German-Hungarian language border and has been bilingual for centuries. The name is the German translation of the Hungarian *Felsőőr*, meaning 'upper sentry'. It shifted politically between Austria and Hungary as state boundaries changed but has remained part of Austria since 1921.

'Peasant men can't get wives'

Since the Second World War Oberwart has received many German-speaking immigrants and undergone expansion, urbanization and industrialization, in a pattern typical of post-war Europe. This has turned it from a rural, farming-focused community into a more city-like environment. Gal's study (published 1979) offers a rich and winsome characterization of the social traits of the town and its peoples, telling a story of social division that is manifested in history, architecture, urban planning, lifestyles – and language. Although the bilinguals of Oberwart control a range of styles in both their languages, and use this range for various social functions, their main linguistic choice is between their two languages: 'The choice *between* languages is more salient linguistically and more important socially than style differences within each language' (Gal 1979: 97).

In 1970s Oberwart, Hungarian symbolizes peasant status and is deprecated because peasant status itself is no longer respected. Gal characterizes the label 'peasant' as a 'native cultural category' – it is a label that the people use to describe themselves or others, *Bauer* in German. Young people want the newly available status of being employed workers, not peasants, and the world of work speaks German. For this reason, German carries more prestige than Hungarian – 'you can't go far without German' is a frequent local saying that applies geographically and socioeconomically. In diglossic terms, German is the H language and Hungarian the L. One consequence of the growing social and linguistic divide is, in the lucid title of a 1978 article by Gal, 'Peasant men can't get wives.' The young women of Oberwart do not want to 'shovel cow manure', and this affects their marriage decisions. Another consequence is that children of a marriage between a monolingual German and a bilingual learn only German, not Hungarian.

Language choice in Oberwart–Felsőőr

In this context the choice between the two languages is socially meaningful. Table 5.1 shows the choices that Gal observed for 14 individual bilingual women (in the rows) in their interactions with 11 classes of interlocutors (columns) – in her 1978 article, Gal concentrated particularly on women and their role in the language shift. The table is an implicational scale, a technique not much used nowadays but which in skilled hands such as Gal's offers an enlightening way of displaying individual sociolinguistic behaviour on a continuum (1979: 19). Each cell represents one set of speaker × interlocutor relationships and the language choice favoured for them. For example, the cell where row 12 intersects with column 9 represents the language choice of a 64-year-old woman when talking to her children – in Hungarian. 'Implicational' means that the occurrence of the languages in the cells is expected to pattern so that if H (Hungarian), for example, occurs in a particular cell, all cells below and to the left of that will also have H (as happens for the cell at row 12 × column 9). And if G (German) occurs in a cell, all the cells above and to the right should also show G (as for row 2 × column 5). As a result, one corner of the table will have only Hungarian (here, the bottom left) and the opposite corner only German (top right). If the scale were perfect, the Gs and the Hs would meet in between. Both might occur in the same cells where they meet, but in a perfect scale

Table 5.1 Women's choice of German or Hungarian when speaking to different interlocutors, according to researcher's observation

Speaker		Interlocutors										
No.	Age	1	2	3	4	5	6	7	8	9	10	11
1	25	H	HG	HG	G	G	G	G	G	G	–	
2	15	H	HG		HG	G	G	G			–	
3	23	–	H	HG	HG	–	HG	G		HG	–	
4	27	–	H		HG	G	G	–			–	
5	39		H		HG	–	HG	G	G	G	–	
6	52	H	H	–	HG	HG	–	–	HG	G	–	G
7	17	H	H		HG	–	HG	–			–	
8	59	H		H	H	H	H	H	H	HG	–	HG
9	43	H			H	–	–	–	HG	HG	–	
10	54	H		H	H	H	H	–	H	HG	–	–
11	63	H				H	H	H	H	H	–	HG
12	64	H				H	–	–	H	H	–	HG
13	66	H				H	H	–		H	–	HG
14	71	H				H		H		H	–	H

Interlocutors

- | | |
|-------------------------------------|---------------------------------------|
| 1 God | 7 salespeople |
| 2 grandparents and their generation | 8 spouse |
| 3 black market clients | 9 children and their generation |
| 4 parents and their generation | 10 government officials |
| 5 peers | 11 grandchildren and their generation |
| 6 siblings | |

Notes:

Empty cells = no data, not applicable

Dash – = not enough data

Source: Adapted from Gal (1979), table 4.1

they would never cross over. The scale is ordered to produce the best fit, that is with the least number of crossover cells showing an H in G 'territory' (as in row 3 × column 9) or vice versa. The number of cells that are out of order tells how good the fit is. In Table 5.1, the fit is good, with only three deviant cells (empty cells do not count).

Looking at the ages shown in the table, notice that the scaling has produced an approximate ordering from younger to older speakers from top to bottom. And the order across the page for interlocutors who are relatives is clearly from the oldest interlocutors on the left (column 2) to youngest on the right (11). The speakers are arranged this way to produce the most consistent pattern of Gs and Hs, not directly because of their age. This means that the age ordering displayed in the table is actually a finding from the data not a prior decision. That is, the table gives the evidence that the younger you are, the more German you are likely to use. And the younger your addressee is, the more German is likely to be used to them – and vice versa for Hungarian.

As well as the age pattern, there is also a gradient from left to right across the table from more intimate contexts to more public ones. This horizontal ordering can be

reduced, Gal writes (1979: 126), to a single association – that of ‘urban’ or ‘Austrian’, as opposed to peasant/Hungarian – although this correlates closely with age. Gal’s conclusion – and one that we will have reason to return to when we deal with language style in Chapter 11 – is that ‘only the identity of the participants determines language choice’ (1979: 120). In many sociolinguistic situations, such as this, who is the speaker and who the addressee is the overwhelmingly important factor in code choice.

It is also an issue whether Oberwart represents a case of diglossia (see Exercise 5.7). In terms of its social standing, Hungarian has many of the characteristics of a High language – dictionaries, grammars, literature, etc. (Gal addresses the sociopolitics of Hungarian in later work, Gal and Woolard 2001). Within Hungary, it operates as an H language to minority languages which are Ls. Can it simultaneously be an L language to German as H in Oberwart? The situation reminds us that the social positioning of languages is not immutable across space or time. The same language may be regarded as prestigious in one place and denigrated in another. In Mexico, Spanish is the H language to Mayan languages as L; while next door in the United States, Spanish becomes the L language to English as H.

The social valuing of languages can differ over time as well as between places. The use of Hungarian in church (interlocutor 1) remains as a reminder of the former sociolinguistic norms of this society. Until after the First World War, Oberwart was part of Hungary, and Hungarian rather than German was the national language.

Exercise 5.7 Interpreting language choice in Oberwart

Identify and explain the language choice patterns in Table 5.1, for example:

- 1 ‘God’ as an addressee refers to language use in a public church service (Gal 1979: 125). Everyone says they use Hungarian in church. Why do you think this is?
- 2 Explain the language used to black market clients (3). These are people, Gal writes (1979: 126), who come to the speaker’s home to provide a service, such as repairs or renovation, which they are not licensed to perform.
- 3 Look across the age/generation columns for relatives, from grandparents to grandchildren (2, 4, 6, 8, 9, 11). Examine the patterns of choice, and suggest reasons for what you find. For example, explain why language usage to grandchildren (11) is the most bilingual of all the columns.
- 4 The table shows that language choice is strongly age graded according to both speakers and interlocutors. What do you think these gradients mean for the futures of the two languages?
- 5 Do you assess that Oberwart is a case of diglossia (in Fishman’s definition)? Evaluate the relationship of Hungarian and German against the criteria in 5.3. Is it possible for Hungarian to be a ‘Low’ language in Austria and a ‘High’ language in Hungary?
- 6 What other patterns do you observe in the table, and what are the reasons for them?

It operated as the superposed language in education, bureaucracy, media and church until 1921 (Gal 1979: 161), when the area was transferred to Austria, and German became the national language. The religious standing of Hungarian is the residue of the language’s former prestige. Chapter 10 will focus on such language ideologies, including Gal’s later work.

Finally, Gal notes that in this speech community, code switching is rare (1979: 118), in contrast to the extreme level of switching shown in Exercise 5.6, for instance. Code switching is not rife in all bilingual situations.

5.7 RESEARCH ACTIVITY OBSERVATION VERSUS SELF-REPORT

The standard method of gathering language choice information in the sociological and social psychological traditions is, as we saw in Chapter 2, the survey questionnaire. The data in Table 5.1 however, are the result of Gal’s **observation** of people’s language choice in Oberwart. This involved a year spent living in the community, for example sitting in farm kitchens noting the language behaviours that were going on. Her long-term involvement with the community enabled this and is typical of an anthropological research approach, as we shall see in the next chapter. However, in the middle of that year of participant observation, Gal also conducted a survey with these same people, using a questionnaire to have them report on their own usage (1979: 98). These findings are shown in Table 5.2.

Table 5.2 Women’s choice of German or Hungarian when speaking to different interlocutors, according to self-reported questionnaire

Speaker		Interlocutors										
No.	Age	1	2	3	4	5	6	7	8	9	10	11
1	25	H	HG	HG	HG	G	G	G	G	G	G	G
2	15	H	HG		G	G	G	G				G
3	23	H	H		H	HG	H	G			HG	G
4	27	H	H		HG	G	G	G				G
5	39	H	H		H	HG	HG	G	G	G	G	G
6	52	H	H	H	HG	H	H	H	HG	G	G	G
7	17	H	H		H	HG	G	G				G
8	59	H	H	H	H	H	H	H	H	H	H	H
9	43	H	H		HG	HG		G	HG	HG	G	
10	54	H	H	H	H	H	H	H	H	H	H	H
11	63	H	H		H	H	H	H	H	H	H	H
12	64	H	H		H	H	H	H	H	H	H	H
13	66	H	H		H	H	H	H	H	H	H	H
14	71	H	H		H	H	H	H	H	H	H	H

Source: Adapted from Gal (1979), table 4.3

Gal's use of both these methods offers the opportunity for a rare research activity. We can 'triangulate' the two approaches, that is, compare findings about the same issue gathered by different means. We focus here on identifying and understanding any differences between the results of direct observation and of self-report. Gal does not make this comparison herself, but I have extracted from her tables those 14 speakers who could be clearly identified in both (by means of their unique age). Although self-report is the main source of language choice information, researchers know that as a method it must be handled with care and reservation. Recall Figure 3.1: this showed a steep rise in Māori fluency in a short time, which I interpreted as representing increased motivation to report fluency more than an increase in fluency itself. The comparison in Oberwart enables us to calibrate self-report against what the researcher has actually observed.

The most obvious difference between the two tables shows up at a glance: the self-report table has many more filled cells than the observation table. You can count them and calculate the number of cells that are filled in each table out of the maximum possible total of 154 cells (14 speakers × 11 interlocutors). The reason for the difference is also obvious but worth making explicit: whereas you can ask informants about what they do in any situation, you are unlikely to be able to observe everyone in all situations. Thus, Gal did not observe all 14 speakers in church, but all of them reported on their church usage. There are also some kinds of situations (e.g. with government officials or black market clients) which will routinely occur behind closed doors. Other situations may no longer be observable – a researcher can ask someone how they used to talk to their grandparents when they were alive, but cannot observe that for herself. The coverage of the questionnaire was therefore much more complete than was observation.

The key question to ask about the two sets of findings now is this:

- Are there differences between the information Gal obtained about language choice through direct observation (Table 5.1) compared to the choices speakers reported for themselves when interviewed (Table 5.2)? If so, what do they mean, and how can they be explained?

Exercise 5.8 addresses this overall issue through a number of specific sub-questions, after working through a method for comparing the two tables.

Once the comparison exercise has been completed and your conclusions drawn, we can close by comparing the pros and cons of observation and self-report. The central issue between them is a trade-off between time and reliability. It is much quicker to administer a questionnaire than it is to sit and observe. We also presume that observation is more reliable than self-report in the sense that people may offer inaccurate reports, either because they are unaware of their behaviour or are misleading the researcher. But remember that observation is itself partial, in both senses of the word – it is incomplete and represents one observer's viewpoint.

The comparison we have run here shows differences between the two methods – but more striking is their close fit. About 85 per cent of the cells that are filled in both tables are the same through both methods. When you factor in the amount of time the two methods take – a year devoted to the community versus the much shorter and easier time for doing a survey – it is understandable that researchers tend towards using

Exercise 5.8 Comparing observation and self-report

To make a detailed comparison of the two tables, use Table 5.2 (self-report) as the basis. Ignore cells that are empty in one or other table, mark only cells that are filled in both tables. Go carefully along each row in Table 5.2, marking any cell that is different from Table 5.1. For example, speaker 1 × interlocutor 4 drops H to go from HG in Table 5.2 to G only in 5.1, and speaker 10 × interlocutor 9 adds G to the existing H to create an HG cell. Although this exercise looks complicated, it is in practice quite straightforward if not rushed, and can readily be done in class by individuals or groups.

This comparison should identify just 12 cells that are different between the two tables. Look at these differences, and address the following questions:

- 1 Which language was observed more than it was self-reported? That is, how many Gs does your comparison between Table 5.2 and 5.1 add, and how many Hs? Is there a pattern? If so, what do you think is the explanation?
- 2 Conversely, which language was observed less than it was self-reported? That is, how many Gs were lost in your comparison, and how many Hs? Is there a pattern? Explain.
- 3 Was more or less mixed-language use observed than self-reported? That is, does your comparison find more HG cells in Table 5.1 than in 5.2, or fewer? You can work this out by calculating the number of HG cells out of all filled cells. Does this mean people's reports are more variable or their actual behaviour? Why do you think this is?
- 4 Taking these comparisons, relate them to the interlocutors. Are there any interlocutors for whom there are a lot of differences between observation and self-report? Explain.
- 5 Overall, what is the key generalization that can be made on language choice in Oberwart from the above comparisons between observation and self-report? And what do you take to be the reasons behind the difference?
- 6 What can this comparison tell us about observation and self-report as research methods, and the relationship between the findings that result from them?

self-report. It is rapid and, at least in Gal's study, apparently quite accurate. It offers a ready means of first approach to language choice in a speech community. What it does miss out on – and many researchers would argue this is central to understanding, not peripheral – is the texture of ethnographic context and detail that participant observation garners, and that Gal presents so richly in her book.

5.8 SUMMARY

- Linguistic choice is foundational to all sociolinguistics. 'Language choice' usually refers to the option between different languages rather than between varieties – a term sociolinguists use rather than 'dialect' to cover all kinds of differences within a single

language. Both varieties and whole languages can be covered by the term 'code'. The range of codes which speakers are able to draw on makes up their linguistic repertoire.

- The speech community is a contested notion, with three founding sociolinguists proposing different definitions. At one end the term has encompassed small, specific groups; at the other, enormously broad categories like 'English speakers'. It remains a widespread concept which sociolinguistics seems to need, and has been best defined by Peter Patrick as a socially based unit of linguistic analysis.
- Ferguson's concept of diglossia describes situations where two linguistically related codes are used in different sets of social functions, stratified between 'Low' and 'High'. In classic diglossia, the two codes are related linguistically, as for Arabic, Greek, Swiss German and Haitian Creole. Fishman extended the notion of stratification to clearly different languages, and Fasold extended it further to varieties in monolingual situations. Diglossia has been a very productive notion, although it can be criticized for its status-quo implications.
- Code switching occurs when speakers switch back and forth between distinct codes in their repertoire, often within the same sentence or utterance, and often carrying social meaning. Myers-Scotton originated the Markedness Model to describe the sociolinguistics of code switching, and argues that in some situations code switching itself is the socially unmarked choice. Gumperz proposed an interactional approach, distinguishing a *we*-code from a *they*-code, and situational from metaphorical switching. Auer sees code switching as organized conversational practice, which is either participant-related or discourse-related.
- Gal's study of the Austrian town of Oberwart shows how the choice between German and Hungarian there patterns according to the age and urbanness of speakers, and the effect of different interlocutors. Her use of both observation and a self-report questionnaire enables us to calibrate these two methods, showing that self-report gives broadly similar findings to observation.

5.9 FURTHER READING

Early work by Gumperz (1962), Hymes (1962) and especially Labov (1972b) dealt with the notion of the speech community – Labov's view is critiqued in Milroy and Milroy (1992). Peter Patrick's overview article (2002) offers the best available coverage; see also Rampton (2009), which deconstructs the concept.

Ferguson's original article on diglossia (1959) has been republished often, including in Li Wei's reader (2000) alongside Fishman's 1967 paper that extends the concept. Ferguson became founding Director of the Center for Applied Linguistics (CAL) in Washington, DC, under whose umbrella some of the most influential early sociolinguistic research was conducted or published (e.g. Labov 1966; Shuy, Wolfram and Riley 1968). CAL remains an important focus for socio- and applied linguistic research in the twenty-first century, with a useful website. Hudson (1992) is an accessible bibliography with 1,000 references to diglossia-based work up till that time. Issue 157 of the *International Journal of the Sociology of Language* (2002) is devoted to diglossia, with a long introductory overview by Hudson (2002). Glyn Williams's detailed evaluation of diglossia comes as part of his more general sociological critique of sociolinguistics (1992). The book is a

useful exposé of the ideological underpinnings of concepts and approaches that many sociolinguists have taken for granted.

In code switching, the most accessible presentation of Myers-Scotton's approach is in her two 1993 books, one dealing with structural and the other sociolinguistic aspects. Later developments are found in e.g. Myers-Scotton and Jake (2009). Gumperz (1982a) is the main text on interactional work, plus his 1982b edited collection. Peter Auer's deconstruction of code switching and multilingualism is encapsulated in his 2007 article, and Auer (1988) summarizes his research on Italian/German alternation in Konstanz. Gafaranga (2007) offers a considered overview of the three approaches to code switching distinguished earlier, building mainly on Auer to stress code switching as orderly alternation. Gardner-Chloros's 2009 text is a useful introduction. Collections include Auer (1998), Milroy and Muysken (1995), Bullock and Toribio (2009), and two edited by Monica Heller (1988, 2007a). Volume 3 of Li Wei (2010) republishes leading papers on the sociolinguistics of code switching, with the linguistics covered in volume 1. The collections on overall bilingualism referenced in earlier chapters, such as Li Wei (2000) and Bhatia and Ritchie (2004), carry chapters on code switching. Many papers on the topic appear in the main journals of sociolinguistics and bilingualism (e.g. De Fina 2007).

Gal's Oberwart research is detailed in her 1979 book, with a distillation in the 1978 *Language in Society* article. As we shall see in Chapter 10, her subsequent work has been foundational in the investigation of language ideologies (e.g. Gal and Irvine 1995, Gal and Woolard 2001).

REFERENCES

- Auer, Peter, 1988. 'A conversation analytic approach to code-switching and transfer'. In Monica Heller (ed.), *Codeswitching: Anthropological and Sociological Perspectives*. Berlin: Mouton de Gruyter. 187–213.
- Auer, Peter (ed.), 1998. *Code-switching in Conversation: Language, Interaction and Identity*. London: Routledge.
- Auer, Peter, 2007. 'The monolingual bias in bilingualism research, or: why bilingual talk is (still) a challenge for linguistics'. In Monica Heller (ed.), *Bilingualism: A Social Approach*. Basingstoke, UK: Palgrave Macmillan. 319–39.
- Bhatia, Tej K. and William C. Ritchie (eds), 2004. *The Handbook of Bilingualism*. Malden, MA: Blackwell Publishing.
- Blom, Jan-Petter and John J. Gumperz, 1972. 'Social meaning in linguistic structure: code-switching in Norway'. In John J. Gumperz and Dell Hymes (eds), *Directions in Sociolinguistics*. New York: Holt, Rinehart & Winston. 407–34.
- Bloomfield, Leonard, 1933. *Language*. New York: Henry Holt & Co.
- Bolinger, Dwight, 1975. *Aspects of Language* (2nd edn). New York: Harcourt Brace Jovanovich.
- Bullock, Barbara E. and Almeida Jacqueline Toribio (eds), 2009. *The Cambridge Handbook of Linguistic Code-switching*. Cambridge: Cambridge University Press.
- Busch, Brigitta, 2012. 'The linguistic repertoire revisited'. *Applied Linguistics* 33: 503–23.
- Campbell, Lyle and Verónica Grondona, 2010. 'Who speaks what to whom? Multilingualism and language choice in Misión La Paz'. *Language in Society* 39: 617–46.
- Cheshire, Jenny, Paul Kerswill, Sue Fox and Eivind Torgersen, 2011. 'Contact, the feature pool and the speech community: the emergence of Multicultural London English'. *Journal of Sociolinguistics* 15: 151–96.

- Coupland, Nikolas and Adam Jaworski (eds), 2009a. *Sociolinguistics (Critical Concepts in Linguistics)*, 6 vols. London: Routledge.
- De Fina, Anna, 2007. 'Code-switching and the construction of ethnic identity in a community of practice'. *Language in Society* 36: 371–92.
- Dixon, R.M.W., 1983. *Searching for Aboriginal Languages: Memoirs of a Field Worker*. St Lucia, Queensland: University of Queensland Press.
- Dorian, Nancy C., 2002. 'Diglossia and the simplification of linguistic space'. *International Journal of the Sociology of Language* 157: 63–9.
- Eckert, Penelope, 2000. *Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Malden, MA: Blackwell Publishers.
- Fasold, Ralph, 1984. *The Sociolinguistics of Society*. Oxford, UK: Basil Blackwell.
- Ferguson, Charles A., 1959. 'Diglossia'. *Word* 15: 325–40.
- Fernández, Mauro, 1993. *Diglossia: A Comprehensive Bibliography, 1960–1990*. Amsterdam: John Benjamins.
- Fishman, Joshua A., 1967. 'Bilingualism with and without diglossia; diglossia with and without bilingualism'. *Journal of Social Issues* 23: 29–38.
- Frangoudaki, Anna, 2002. 'Greek societal bilingualism of more than a century'. *International Journal of the Sociology of Language* 157: 101–7.
- Gafaranga, Joseph, 2007. *Talk in Two Languages*. Basingstoke, UK: Palgrave Macmillan.
- Gal, Susan, 1978. 'Peasant men can't get wives: language change and sex roles in a bilingual community'. *Language in Society* 7: 1–16.
- Gal, Susan, 1979. *Language Shift: Social Determinants of Linguistic Change in Bilingual Austria*. New York: Academic Press.
- Gal, Susan and Judith T. Irvine, 1995. 'The boundaries of languages and disciplines: how ideologies construct difference'. *Social Research* 62: 967–1001.
- Gal, Susan and Kathryn A. Woolard (eds), 2001. *Languages and Publics: The Making of Authority*. Manchester, UK: St Jerome Publishing.
- Gardner-Chloros, Penelope, 2009. *Code-switching*. Cambridge: Cambridge University Press.
- Gumperz, John J., 1962. 'Types of linguistic communities'. *Anthropological Linguistics* 4: 28–40.
- Gumperz, John J., 1968. 'The speech community'. In David L. Sills (ed.), *International Encyclopedia of the Social Sciences*. New York: Macmillan and Free Press. 381–6.
- Gumperz, John J., 1982a. *Discourse Strategies*. Cambridge: Cambridge University Press.
- Gumperz, John J. (ed.), 1982b. *Language and Social Identity*. Cambridge: Cambridge University Press.
- Heller, Monica (ed.), 1988. *Codeswitching: Anthropological and Sociological Perspectives*. Berlin: Mouton de Gruyter.
- Heller, Monica (ed.), 2007a. *Bilingualism: A Social Approach*. Basingstoke, UK: Palgrave Macmillan.
- Hudson, Alan, 1992. 'Diglossia: a bibliographic review'. *Language in Society* 21: 611–74.
- Hudson, Alan, 2002. 'Outline of a theory of diglossia'. *International Journal of the Sociology of Language* 157: 1–48.
- Hymes, Dell H., 1962. 'The ethnography of speaking'. In Thomas Gladwin and William C. Sturtevant (eds), *Anthropology and Human Behavior*. Washington, DC: Anthropological Society of Washington. 13–53.
- Hymes, Dell, 1972. 'Models of the interaction of language and social life'. In John J. Gumperz and Dell Hymes (eds), *Directions in Sociolinguistics*. New York: Holt, Rinehart & Winston. 35–71.
- Hymes, Dell, 1974. *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphia, PA: University of Pennsylvania Press.
- Kachru, Braj B., 2001. 'Speech community'. In Rajend Mesthrie (ed.), *Concise Encyclopedia of Sociolinguistics*. Oxford, UK: Elsevier. 105–7.

- Kaye, Alan S., 1972. 'Remarks on diglossia in Arabic: well-defined vs. ill-defined'. *Linguistics* 81: 32–48.
- Labov, William, 1966. *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, William, 1972b. *Sociolinguistic Patterns*. Philadelphia, PA: University of Pennsylvania Press.
- Li Wei (ed.), 2000. *The Bilingualism Reader*. London: Routledge.
- Li Wei (ed.), 2010. *Bilingualism and Multilingualism: Critical Concepts in Linguistics* (4 vols). London: Routledge.
- Milroy, Lesley, 1987. *Language and Social Networks* (2nd edn). Oxford, UK: Basil Blackwell.
- Milroy, Lesley and James Milroy, 1992. 'Social network and social class: toward an integrated sociolinguistic model'. *Language in Society* 21: 1–26.
- Milroy, Lesley and Pieter Muysken (eds), 1995. *One Speaker, Two Languages: Cross-disciplinary Perspectives on Code-switching*. Cambridge: Cambridge University Press.
- Migge, Bettina, 2007. 'Code-switching and social identities in the Eastern Maroon community of Suriname and French Guiana'. *Journal of Sociolinguistics* 11: 53–73.
- Myers-Scotton, Carol, 1988. 'Code-switching as indexical of social negotiations'. In Monica Heller (ed.), *Codeswitching: Anthropological and Sociological Perspectives*. Berlin: Mouton de Gruyter. 151–86.
- Myers-Scotton, Carol, 1993a. *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Oxford University Press.
- Myers-Scotton, Carol, 1993b. *Social Motivations for Codeswitching: Evidence from Africa*. Oxford: Oxford University Press.
- Myers-Scotton, Carol and Janice Jake, 2009. 'A universal model of code-switching and bilingual language processing and production'. In Barbara E. Bullock and Almeida Jacqueline Toribio (eds), *The Cambridge Handbook of Linguistic Code-switching*. Cambridge: Cambridge University Press. 336–57.
- Patrick, Peter L., 2002. 'The speech community'. In J.K. Chambers, Peter Trudgill and Natalie Schilling-Estes (eds), *The Handbook of Language Variation and Change*. Malden, MA: Blackwell Publishing. 573–97.
- Poplack, Shana, 1980. 'Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: toward a typology of code-switching'. *Linguistics* 18: 581–618.
- Rampton, Ben, 2009. 'Speech community and beyond'. In Nikolas Coupland and Adam Jaworski (eds), *The New Sociolinguistics Reader*. Basingstoke, UK: Palgrave Macmillan. 694–713.
- Romaine, Suzanne, 1995. *Bilingualism* (2nd edn). Oxford: Blackwell Publishers.
- Shuy, Roger W., Walt Wolfram and William K. Riley, 1968. *A Study of Social Dialects in Detroit*. Washington, DC: Educational Resources Information Center.
- Snow, Don, 2010. 'Hong Kong and modern diglossia'. *International Journal of the Sociology of Language* 206: 155–79.
- Williams, Glyn, 1992. *Sociolinguistics: A Sociological Critique*. London: Routledge.
- Woolard, Kathryn A., 1999. 'Simultaneity and bivalency as strategies in bilingualism'. *Journal of Linguistic Anthropology* 8: 3–29.
- Woolard, Kathryn A., and E. Nicholas Genovese, 2007. 'Strategic bivalency in Latin and Spanish in early modern Spain'. *Language in Society* 36: 487–509.