

a language other than the mother tongue). This may apply to a number of occupational demands: to be able to function as a teacher, interpreter, translator, diplomat, holder of scholarships abroad, student, immigrant adviser, civil servant in a bilingual country (or in a monolingual country that uses a language other than the applicant's mother tongue) or region or office.

When the degree of bilingualism is measured for such purposes as these (to determine, that is, whether or not a person is sufficiently bilingual or monolingual for something), all that is usually produced is a single measurement. But we need also to be able to measure the degree of bilingualism repeatedly over a period of time if the aim of the measurement is to *determine what changes occur as the result of deliberate attempts to influence the situation*. The most common situation will, of course, be the attempt to make an individual bilingual or more bilingual by a process of education and teaching. In such a case, the individual's *degree of bilingualism* (in most cases only the knowledge of the more foreign of the languages) *before, during and after* the process of instruction is measured. And measurement of this kind happens in foreign language teaching in schools or language courses.

8 How is bilingualism measured?

8.1 Why should it be necessary to measure bilingualism?

Even though it will already have emerged several times that I consider attempts to describe bilingualism and bilingual individuals to be more useful than attempts at measuring, and even though I believe that we are not yet able to describe bilingualism in individuals securely enough to be able to base any principle of measurement on sound foundations, I propose nevertheless to describe in this chapter the kinds of measurement that have been attempted. Let me emphasize very strongly, however, that simple measurement in isolation may do more harm than good, and that measurement will always do some violence to the subtle texture of reality.

Just like definitions of bilingualism, measurement can assess features and abilities in individuals or communities. In the case of communities, a measurement can be to count how many individuals are bilingual and to what extent (and in estimating that, individual measurement will be necessary), or to what extent communities function in several languages. Many demographers and political scientists have published detailed discussions relating to the description and measurement of bilingual communities (see e.g. McRae – references in the bibliography). I shall not concern myself here with the measurement of community bilingualism but only of individual bilingualism.

There are many reasons other than research reasons for measuring the bilingualism of individuals. It is important, for instance, to know who is *sufficiently monolingual, that is, little enough bilingual, to need or to be entitled to some degree of language service*. This applies to the placing of minority children in schools, the right to the services of an interpreter, the right to language teaching, the right to demand service in one's mother tongue, and so on. It is also important to know who is *sufficiently bilingual to qualify* for positions where bilingualism is expected (or a good knowledge of

8.2 Different ways of classifying the measurements

Linguistic behaviour is to a high degree unconscious, beyond conscious control of detail. No comparison has been made between the degree of conscious control exercised by monolinguals switching between different social or regional variants, and bilinguals switching between two different languages. But just as switching between two languages is easier to observe than switching between varieties of one language, so it seems natural to assume that bilinguals themselves are perhaps more conscious of their linguistic behaviour than monolinguals. The reports by monolingual individuals as to the linguistic norms they observe and their language use show a very weak correlation with the ways in which they actually in practice use their language (see, for example, Blom & Gumperz, 1972; Labov, 1966, 1972). Individuals tend to think they are reporting their linguistic usage when they are in fact reporting the linguistic norms they think they *ought* to observe in speech (Labov, 1972a; Thelander, 1974: 44, 68; Widmark, 1973: 20), without *consciously* distorting the evidence. Something like this can also be observed in diglossic communities, but on the whole we can assume that linguistic behaviour among bilinguals is more conscious than among monolinguals, *provided* there is no external pressure. The evidence available as to the greater metalinguistic awareness of some bilinguals (see references in the section on bilingualism and intelligence, 9.4.3) points in the same direction. This is why it is assumed that reports by bilinguals

about their own linguistic behaviour can be used, with certain reservations, as a measure of their bilingualism.

In what follows I have chosen to separate the different sorts of measurement of bilingualism into *reported* and *observed* linguistic behaviour. I classify as reported behaviour all information provided by a bilingual speaker herself, either in questionnaire or interview form, as to linguistic use and knowledge: this will consist of various kinds of *background data*, of *self-assessment*, and of *the results of measurements of attitudes*. Among the background data, of course, is the relevant demographic information, which the researcher need not obtain directly from the informant, but which at one time or another the informant has supplied to the authorities or to other researchers. I classify as observed behaviour information gathered from observation of the informant and her linguistic performance. The period of observation may be a short one, as for example during an interview (when it is possible to make some assessment of the linguistic performance of the individual being interviewed), or it may be a long one, for example participant observation where the researcher seeks confirmation of reported linguistic behaviour. The most important form of observed linguistic behaviour consists of different kinds of language test.

The tests used in the measurement of bilingual linguistic proficiency are of two kinds: 1. tests which are also used for measuring the *linguistic proficiency of monolinguals*, and which use the monolingual as a norm. Such tests measure the bilingual individual's proficiency in her two languages separately; and 2. tests which are used only to measure *a bilingual proficiency*. Such tests measure the bilingual's two languages in relation to each other.

8.3 Reported linguistic behaviour

8.3.1 Background data

Information about when, where and to whom the bilingual speaks each of two languages is usually collected by means of interviews or questionnaires about background data.

The first carefully constructed questionnaire focusing on the linguistic background of bilinguals, the Hoffman Bilingual Schedule, was published as early as 1934. For decades it was the model for similar questionnaires aimed at collecting background data, and it is still used (for example by Zirkel, 1972: 70). Hoffman's well-tested questionnaire contained questions about the spoken language, about reading and writing, about the informant's own active language use ("which language do you speak to your mother?"), about the language of the rest of the family ("which language does your father speak to your mother?"), and about activities outside the home

("how often do you listen to talks given in L1?"). Hoffman weighted all these questions equally, and his measure of bilingualism (bilingual score) is the average of the points scored. More recent research has been critical of this kind of questionnaire in some respects. Einar Haugen believes that a more *detailed* differentiation would provide a more reliable picture of the background variables. Haugen would prefer to construct a bilingual profile, or use more precisely limited and defined measurements, each of which would describe *one* aspect of the complex of background factors (1964: 15). Macnamara divides the background questions into five categories, relating to the language use of the different speakers (informant herself, father, mother, siblings and the environment). According to Macnamara, nothing of the reliability is lost if all five categories are combined into a single evaluation of the background factors (Macnamara, 1969: 85-86). A serious criticism levelled against this sort of questionnaire is that it combines such different kinds of information without any proper weighing of the relative importance, attributing equal importance to data which, in my view, are non-comparable. Factors which have been given equal weight (where I think they should not) are for example length of residence, whether or not the mother works outside home, whether or not the child speaks L2 with friends, whether or not the child listens to talks in L2. Very important factors affecting the parents' knowledge of L2, like sex, education and age at the time of migration are not discussed at all, for instance, in Hoffman's scale (see Muncio & Meisaari-Polsa, 1980, describing how these factors differentially affect the knowledge of Swedish among Finns, Poles, Turks and Yugoslavs in Sweden).

We can also criticize the failure to make any adequate distinction between *knowledge* of L2 (or L1) and the *actual use* of it, to distinguish between competence and performance. Such a distinction is not brought out sufficiently clearly if we only ask which language the informant speaks to the father without asking an additional question about how well the father speaks this language. Parents may for reasons of prestige and for other reasons choose to speak to their children in a language they do not speak particularly well, in order to help their children to learn it (Jaakkola, 1973a). In Sweden it still happens that teachers tell immigrant parents to speak Swedish to their children, even if their Swedish should be very poor and thus provide a very poor model for their children (Häkkinen, 1974; Jauho & Loikkanen, 1974: 63). And even academic researchers have been known to give such advice! Macnamara gives an example of social pressure to exaggerate the use of L1: the Irish government used to pay £10 per year per child to the parents of children who spoke Irish at home. The temptation in this case to say one did when one did not speak Irish at home must have been great (Macnamara, 1969: 86). The opposite case is also found, for instance in the Finnish-speaking districts of the Tornedä Valley in Sweden:¹ speakers report that they speak L1 less and L2 more at home than is really the case, or they may entirely suppress the fact that they use L1, or feel offended at the

suggestion that they speak the majority language of the district (Hansgård, 1972; Loman, 1974a). The status of a language, the attitudes to its speakers, the official attitude, and so on, all greatly affect the reliability of background data. However, in many investigations the background data do show a fairly strong correlation with other measurements of bilingualism (see, for example, Arsenian, 1937; Pintner & Arsenian, 1937; Kolasa, 1954; James, 1947; Sahlman-Karlsson, 1974). If the above caveats are borne in mind as to the possible sources of error, background data can be used as one kind of measurement of bilingualism. The collection of background data is the easiest way of acquiring information about the use of languages. A more reliable way would be a lengthy period of participant observation but few linguists will have the opportunity for this unless the informant is a member of their own (extended) family.

Other common kinds of measurement of bilingualism, by self-assessment or linguistic tests, do not usually measure language use but linguistic proficiency, competence.

8.3.2 Self-assessment

In self-assessment an informant is asked to assess her own knowledge (or use) of L2 or of both languages. One may ask for a fairly *well-differentiated assessment* (see the examples below) or be content with an *overall judgement*. One may distinguish between *primary* and *secondary skills*, or between *receptive* and *productive skills*. Each language may be treated *separately* or questions may relate only to the more foreign of the two. The *languages* can be investigated *in relation to each other*, i.e. questions can be asked about the degree of bilingualism. Here particular scales may be used which start from the notion of a *balance effect* (they may be scales based on the assumption that a speaker knows one of two languages better and therefore the other one less well, or scales which accept the answer "both equally well" as unproblematical, that is, as evidence of the fact that knowing both languages *equally well* means knowing both of them well). Or scales may be used which make no *assumptions about the balance between the languages* in advance, so that all possible combinations are taken into account, even the possibility that the speaker may know both badly.

Self-assessment seems to be a fairly reliable way of measuring bilingualism.² In the investigation carried out by Invandrartredningen (the Swedish Immigrant Survey), in which self-assessment was compared with observed behaviour (the judgement made by the interviewer), the two kinds of measurement were found to differ only in minor ways. "Only in the case of the extreme values on the scale was there a certain discrepancy: interviewers used extreme values more often than interviewees, who tended to prefer the intermediate values" – thus the result is summarized by Municio & Meisaari-Polsa (1980: 18–19).

We should, however, be aware of certain reservations about the reliability of self-assessment, as about the reliability of background data. If one of the languages enjoys much higher prestige, informants tend to exaggerate their knowledge of it, and to minimize or completely suppress their knowledge of the less prestigious language (for examples of this, see Hansgård, 1972: 111; Jaakkola, 1973a, and section 8.3.1). Self-assessment tends to be more realistic if both languages enjoy more or less equal status, if a proficiency in one or the other of the two languages does not lead to positive or negative sanctions, i.e. if the sociolinguistic climate is relaxed. The amount of contact with L2 also influences the assessment so that a speaker who is frequently in a position to hear and speak L2 will have a more realistic notion of her ability than another speaker who more rarely uses L2. It is easier for a Finland Swede to over-estimate her knowledge of Finnish if she lives in Pargas/Parainen (where about two-thirds of the population report themselves as Swedish-speaking) and only rarely comes into contact with Finnish speakers, than if she lives in Kokkola/Gamlakarleby (where less than one-fifth of the population is Swedish-speaking) and uses Finnish daily at work (Sandlund, 1970, 1972). Swedish speakers' assessment of their knowledge of Finnish is more reliable in Finland than Finnish speakers' assessment of their knowledge of Swedish (Lindman & Alho, 1971), because the Swedish speakers probably more often come into contact with their L2, Finnish, than do the Finnish speakers with their L2, Swedish. In the case of children of school age, it has proved that their grade at school in the language in question affects their self-assessment (Macnamara, 1969: 86), and this has to be taken into account when the results are interpreted.

The expectation an individual herself or others have of her knowledge of L2 may also vary. There may be lower expectations as to someone's knowledge of a language if the individual lives in a country or a locality where there are fewer opportunities of using the language in question. A speaker may also come to expect more of herself without being aware of it if she is faced with a subtler and more varied range of the language than before. A speaker may thus think less of her grasp of a language at a later date than she did earlier, even though, objectively speaking, she has learnt more of it in the intervening period, because during that time she may have faced novel situations which revealed that her linguistic grasp was inadequate in some areas.

Friendly native speakers, who with a beginner in their language may be very understanding and loud in their praise, expect more as the stranger comes to be more integrated, and they may occasionally be irritated when even after a couple of years the foreign speaker still does not understand everything or finds it difficult to express herself in writing. Language difficulties may also at the beginning be used as an excuse, but after a while it becomes less easy for native speakers to identify as language difficulties what they take to be simple stupidity, ignorance, defective logic, or an inability to think out properly formulated sentences. And this will, of course,

also affect the bilingual speaker's own estimate of her linguistic ability.

The discrepancy in children between pronunciation (which may be absolutely accurate) and the rest of their proficiency in the language can be just as confusing (see sections 5.2.2.5 and 7.3.5.2). Surface fluency affects children's self-assessment to a very high degree, even leading them to insist that they do not need help with L2 at school. However, it is more usually the case that immigrant children, or at any rate slightly older immigrant youngsters assess their knowledge of L2 as being rather worse than their teachers do: the pupils say that they have language problems whereas the teachers will say that they speak Swedish as well as the average Swedish child would. Even adults who speak with a good accent can in the same way lead native speakers to think that they know the language well, despite the fact that they may themselves say that they have difficulties. The discrepancy prevails above all between speaking and writing. If the speaker's self-assessment is based on both primary and secondary proficiency in L2 (where the former is good and the latter poor) and the observer bases the assessment on primary proficiency only, the correlation between the two is of course poor. This is more often true of speakers from oppressed minorities than from stationary indigenous minorities educated in their own language. An overall judgement should therefore not be used when assessing them, but primary and secondary proficiencies should be kept separate. The expectations that native speakers have of how well foreigners should command their language, also affect the non-native speaker's self-assessment and wish to try to speak the language. People who have been able to force other people to learn their language often have high expectations of foreigners (who often do not dare to try to speak the foreign language, because of negative experiences of the tolerance of the L2-speakers), while people whose languages are seldom learnt by foreigners often show foreigners good-will which is out of proportion to the foreigner's actual knowledge of the language – and this encourages them to learn more. All this also affects the self-assessment.

Many sociologists take the view that the conceptions informants themselves have of their knowledge of a language are more important as objects of study in sociological investigations than the actual state of their knowledge, since their assessment (both of their own linguistic knowledge and of their interlocutors') is decisive for their choice of language and their linguistic behaviour in general (Pipping, 1969: 217; Sandlund, 1970: 10; Broo, 1974: 6).

8.3.3 Measurement or description?

Where background data and self-assessment are used in *descriptions* and not in *measurements* of bilingualism, different kinds of information are often collected which throw light on the choice and use of language by

informants in various situations. Good surveys of the variables that might form part of such a description have been made by William F. Mackey (1970) and Joan Rubin (1970: 525–27). In Rubin's questionnaire the components are organized hierarchically so that later alternatives are determined by earlier choices. In many models the factors affecting language choice are separated into cognitive, affective and action components. When background data and self-assessment are used for the *measurement* of bilingualism rather than for its description or for producing a bilingual profile (Haugen, 1964: 95; Mackey, 1970), the variables have to be very much restricted in number, so that one single measurement can be produced. Just as with the construction of a social index, so with an *index of bilingualism*, it is the most decisive variables which must be focused upon so that they can then be combined into one index. And in both cases the choice of variables is crucial, and will at the same time also be a measurement of the researcher's ideological and political priorities.

Some researchers collect information about only one of the languages in measuring bilingualism, others about both. If we assume that an informant always has at least *one* complete language, L1, then whether or not we believe that complete bilingualism exists, we can be contented to collect information about L2 only. If, on the other hand, we suspect that *both* languages may be more or less incomplete, then we should preferably have information (and especially test results) relating to both languages. An index of bilingualism based on one language can only be used if there are other types of evidence that the informant has a "complete" command of the other language (the language not measured in the index), or at least of those aspects of it which the index measures.

8.3.4 Attitudes

8.3.4.1 Limitations

In studying attitudes towards different languages, we can focus upon monolingual speakers' attitudes to foreign languages as well as upon bilingual speakers' attitudes to their two languages. *Attitudes towards languages* are very often, if not always, impossible to separate completely from *attitudes towards speakers of those languages*. To try to do it is in part artificially to wrench language away from its use. But again, studying attitudes towards language *users* is the topic of other books, books about prejudice, discrimination and racism at the individual level, leading to studies about power relationships and racism at structural and institutional levels which are important prerequisites for understanding language questions. But here I have to limit myself to attitudes of individuals not groups, towards languages not language users.

Many studies have been made of the attitudes of monolingual speakers (i.e. majority speakers – minority speakers are mostly bilingual) towards

foreign *languages* in the attempt to discover what kinds of attitude lead to the best results in foreign language learning.

Most of the investigations which concern themselves with the attitudes of monolingual speakers towards *users of various languages* elucidate these attitudes by asking questions simply about attitudes towards *various languages*. Such studies try to map out the contours of prejudice, racism, ethnocentrism, and so on. Studies of this kind are still relatively rare in Scandinavia.³

8.3.4.2 Instrumental and integrative attitudes

The Canadian psychologist Wallace E. Lambert, studying the role of attitude and motivation in L2 learning, especially (but not only) in bilingual environments where the groups come in contact with each other, has distinguished between what he calls instrumental and integrative motivations. An *instrumental* orientation emphasizes the utilitarian aspects, an *integrative* the desire to learn (and capability of learning) to know the other linguistic group and perhaps to a certain extent to identify with it, become a member of it (Lambert, Gardner, Olton & Tunstall, 1970: 473-74). The integrative motivation seems to be more durable in the long drawn out labour of L2 learning, when intelligence and language aptitude are kept constant (Gardner & Lambert, 1972: 193-94).

In what follows we will discuss some of the attitudinal factors which create good relationships between minority and majority students. The first question is whether or not contact with the speakers of L2 and knowledge of L2 create a more favourable attitude towards it and its speakers. Results are here divided. Gardner & Lambert show that students doing well had parents with positive *attitudes* towards other language groups, but their attitudes did not correlate with their *knowledge* of other languages or their *contacts* across the linguistic border. Arne Trankell (1974: 158-62) indicates the exact opposite, showing a connection between contacts with foreigners and positive attitudes (more generous, less fearful and less discriminatory) towards immigrants. Mirko Takač (1978) again points to very negative attitudes among Swedish-speakers towards immigrants (but less *vice-versa*) in school districts with a great deal of contact. Lambert also discusses the direction of the correlation and assumes that the positive attitude in students which gives rise to an integrative motivation is primary and results in a good command of L2, but he cannot exclude the opposite possibility: that an individual who knows an L2 well will also develop a positive attitude towards its speakers.

It seems that part of the answer lies in the *type of contacts* one has – the amount does not seem to be decisive. In a Canadian experiment English speakers who had learnt French at school were asked to communicate with what they thought were invisible French and English speakers (in reality tape recordings made beforehand) who both launched into informal discussion and also invited a more formal style of response much like the one

the informants had learnt at school during French lessons. The informants were asked to judge both the French and English speaker with regard to various characteristics in both the formal and the informal situation, and to say what kind of impression they thought they themselves had made on the partner. They felt more ill at ease themselves, thought that they had seemed more unfriendly, less intelligent and less self-confident, and gave the partner more unfavourable traits of character in the *informal* discussion in French, while they thought positively of both themselves and the partner in the informal discussion in English (Segalowitz, 1976, reported in Segalowitz & Gabonton, 1977). If the learner of L2 with an integrative attitude tries to come into contact (or is forced into contact, like many minority children, regardless of their attitude) with speakers of L2 before she has a *good* command of the style of informal speech, she may find the experience off-putting and may be discouraged from further attempt at contact. If a minority child in Scandinavia who has just begun to learn the most basic school vocabulary in L2 is placed in a situation where she is forced to communicate informally with Swedish/Norwegian/Danish children, all the negative and defensive psychological mechanisms may swing into operation. The child will feel stupid and helpless, will attribute to the native speaker a lot of negative characteristics, and will experience such a sensation of unease as is likely to influence her response for long afterwards, no matter how encouraging teachers, peers and other partners are. It would often seem better, then, to delay informal contacts until the child has reached the stage of being able to cope with them linguistically as well without any damage to the self-confidence – this is a further argument for not being in too much of a hurry to integrate immigrant children with children who are majority speakers.

Two pedagogical implications seem to follow from a number of studies. It should be clear that negative attitudes and racism cannot be informed away – they spring from economic and political causes, and these basic causes are reflected for instance in majority speakers' negative attitudes towards minority languages and speakers. If one wants, nevertheless, to try to do something at the individual level, contacts between majority and minority speakers should be organized so that both (but especially the one who has to function in an L2) can feel secure and self-confident in a natural way, i.e. be in a position where it is possible to do justice to the knowledge and skills that each brings with her. Again, forcing minority children into situations where they are made to feel incompetent, both linguistically, socially and in terms of content, in majority medium classes, does not seem to be conducive to fostering positive integrative attitudes in minority or majority children. With the differences between prerequisites and results among different children trying to become bilingual in mind (discussed in section 4.2) we could also predict that majority students learning a minority language as their L2 (for instance English children or adults in Britain learning Gujarati) should be able to tolerate more situations where they feel

incompetent in Gujarati, than minority students learning a majority language as their L2 (Gujarati children in Britain learning English), because the whole endeavour is a game for the former and bitter reality for the latter.

When studying a bilingual's attitude towards her two languages it is important to distinguish between a majority group bilingual with a clearly dominant mother tongue and a much less good command of L2 (who can be studied much in the same way as a monolingual learning an L2), and either one with approximately equal command of both, or a minority bilingual. The decisive factor seems to be whether or not we can assume that the speaker has a positive or "neutral" attitude towards L2 (whether it is an enrichment, much like a game, where it is not connected with a central value in the speaker's life). If L1 is a minority language with low status, then a negative attitude towards L2 may be felt by the speaker as defending and preserving her language and identity.

The distinction between instrumental and integrative motivation is particularly useful in situations where a distinctly non-integrative attitude to L2 is judged to make for this preservation of the identity or even the existence of a minority group. Here one might nevertheless expect a strong, instrumentally orientated motivation to learn L2 since the minority speaker needs the majority language as an instrument.

The difference between minority and majority group informants (whether monolingual or bilingual) has not always been sufficiently taken into consideration when results have been interpreted. In Gathbonton (1975; as reported in Segalowitz & Gathbonton, 1977) Francophones hearing French Canadian voices reading English, ascribed pro-Anglophone attitudes to the speakers on the basis of a native or near-native English accent, while voices with a more marked Francophone accent were ascribed pro-Francophone attitudes (which did not show any correlation with the real attitudes of the voices). The more "nationalistically" minded informants chose the most fluent sounding speakers only for positions with connections with Anglophones (i.e. purely instrumentally) while the least "nationalistic" subjects tended to choose the voices with the most native-sounding English accents for *all* leadership roles, regardless of the amount of contact the roles implied with Anglophones. The existence of a certain sociolinguistic pressure is revealed in the study: a minority speaker whose accent in the majority language is "too" genuine might be chosen for roles where the command of L2 is instrumental, but her loyalty to her own group may also be questioned by the group because of the accent. Another study with the same type of experimental design (but with both Francophone and Anglophone voices, with different degrees of accent in their L2) by Genesee & Holobow (1978) had both bilingual (immersion) and monolingual Anglophone children as informants. There are many interesting details in this well-controlled study, but the interesting thing from the point of view of the discussion here is that Genesee & Holobow, who in many respects record results that contradict those of Gathbonton & Segalowitz, do *not* interpret

them as a difference between minority and majority group informants, which to me is the natural interpretation. The majority group, whether bilingual or monolingual, has no reason for feeling linguistically threatened by the minority. It needs to have recourse to no boundary-preserving mechanisms, and an integrative attitude, a desire to identify with the minority group, is not for it in opposition to a desire to identify at the same time with its own majority group. There would be very little risk that positive attitudes towards Francophones and the desire to mix with them socially would lead an Anglophone to abandon her English for French. Nor is there any risk in Canada that the Anglophone population *as a group* might be absorbed into the Francophone group, and cease to exist as an independent entity. Thus for the majority group there is no risk entailed in combining positive attitudes towards *both* linguistic groups. For a minority, on the other hand, such attitudes may objectively bring with them the risk of the dissolution of the minority group because of the relative strength of the majority group.

We also need to discuss to what extent it is *language* and to what extent other factors that influence attitudes. Language functions as a symbol of group membership. Usually a number of other features are associated with that linguistic symbol, even where, to judge from outward appearances, the language difference seems to be the only one separating two groups (as with Finland Swedes and Finnish-speaking Finns). Other associated features are, of course, particularly in evidence where the linguistic border coincides with other borders (ethnic, racial, economic, political, etc.). That language need not be the most important symbol informing attitudes towards other groups is clear from Allardt, Miemois & Starck's thorough study (1979) of the Finland Swedes (912 subjects from Metropolitan Helsinki). When asked which groups they most and least identified with, only 7% of them gave Swedes from Sweden (speaking the same language) as the group most identified with, whereas Finnish-speaking Finns from Helsinki were by far the group most often identified with (28%) – and 34% gave Swedes from Sweden as the group they had the weakest feeling of affinity with (whereas Swedish-speaking Finns from other regions in Finland and Finnish-speaking Finns from the area around Helsinki occupied intermediate positions on the scale). Eric Allardt's interpretation of the answers stresses the *existence of multiple identities*, that an individual feels she belongs to several different groups (without feeling that there need be any kind of conflict between them). He stresses similarly the great *importance of local and regional ties*: an individual feels greater affinity with others *from the same country who speak a different language* than with individuals *from a different country who speak the same language*, to give the two extreme examples.

The same tendency for language as a factor to lose its importance in circumstances where the language question is not felt to be very problematical (as it *could* be if the official status of the two languages were highly unequal) is also clear from other investigations. One of them

(Sandlund, 1971), carried out by Ulf Berg, investigated which conflicts 454 Finland Swedes regarded as the most serious in Finland (conflicts between language groups, urban and rural dwellers, social classes, employers and employees, the last two classified together in the analysis as class conflicts). They were also asked whether they thought that the language conflict was going to grow or diminish in the future. The results indicated that those in more homogeneous Swedish-speaking areas, with a less secure grasp of Finnish, and from the lower social classes tended to emphasize the language conflict and be pessimistic (quite understandably because it is more difficult for them to get work in Finnish-speaking areas because of language if they become unemployed, one reason for Finland Swedes being over-represented among immigrants from Finland to Sweden), while those from bilingual areas or closer to the capital tended to emphasize the class conflict and think that conflict between the language groups was bound to diminish. Bilingualism may thus lead to a lessening of the importance of the language factor itself, provided that the standing of the two languages is similar (as it is in Finland), and one can afford to emphasize other solidarities and conflicts. This represents a more instrumental and relaxed attitude to the languages in question and can be found in situations where the need for boundary-preserving mechanisms is no longer so great and where there is a possibility of attaining some type of instrumental integration (or structural incorporation) into the majority. It need *not* imply any less of an ethnic awareness, as Allardt points out – it merely lessens the risk of serious conflict (Allardt, Miemois & Starck, 1979: 22). A good way to avoid conflict and assist integration in multilingual schools would thus also be to make the status of the minority languages as high as possible, so as to equalize them with the majority language – that is bound to give a basis for positive attitudes in *both* minority and majority students.

8.4 Observed linguistic behaviour

8.4.1 Limitations

Much of the information that can be obtained by collecting background data, reported by a bilingual informant, can also be obtained or confirmed by observation, systematic or unsystematic. The most important way of getting a measurement of bilingualism by “objective” observation, however, consists in the measurements of proficiency by means of *tests*.

The usefulness of language tests depends completely upon what one tries to measure with them, and the purpose for which the results are required. Many linguists still doubt that we know enough about language, language learning, and linguistic proficiency to be able to construct adequate, varied, objective linguistic tests (see, for example, Gumperz,

1972: 204). The value of measurement as such is questioned by some researchers, especially if measurement, the collection of “hard evidence”, is not combined with other, more descriptive kinds of observation: “soft evidence”. In spite of all my reservations about the use of tests in an area about which we know too little to allow any great confidence on our part that they can be made reliable, they merit some discussion here. For more comprehensive kinds of discussion see, for instance, Oller, 1979.

I have already above classified tests used to measure the linguistic proficiency of bilinguals into those which are (or can be) also used for monolinguals and which take monolingual proficiency as the norm, and those which are designed solely for the purpose of measuring bilingual proficiency. Certain tests which fall into the first category may, as an unintended side-effect, also bring forth features which are particularly characteristic of the bilingual, for instance, interference. If, however, the purpose for which the test was designed is to measure something else, for instance certain syntactic constructions, and the measurement of interference was not part of the intention (test material could for instance be designed with the help of contrastive analysis so as to be likely to produce interference), then the test clearly falls into the first category. It is the purpose for which the test was designed (and the linguistic norm envisaged), not the results, which decide the category to which it belongs. It should be said that by the use of monolinguals as a norm, two quite different things may be meant (the two may sometimes be combined in the same test, but sometimes only one kind of use is involved):

- In the first case, the linguistic proficiency aimed at, and which the subjects’ knowledge of a language should approach and be comparable with, is the kind of proficiency that a native speaker of the target language possesses. The *native speaker of the target language* is assumed to be *monolingual*, and it is she who sets the norm for the person whose proficiency in the target language is being tested.

- In the second case, the individual from whose language teaching and learning situation one starts is the monolingual speaker: she is in process of learning the target language as a school subject or in some other pedagogic situation, and is thus being subjected to a conscious process of instruction; she is not learning the language in “natural” circumstances. There is no attempt made to focus upon the kind of circumstances in which a “natural” bilingual would have acquired L2, and to test a language most probably to be acquired in those circumstances. Rather what is being tested is the pedagogically presented *knowledge* about the language that teachers have *tried to impart* quite consciously, by ordinary instruction. Here the *subject being tested* is *monolingual*; and the norm for proficiency in the target language is the kind of command a monolingual speaker might expect to have of a foreign language.

When tests of the first kind are used, one wants to compare the bilingual’s linguistic proficiency with that of a monolingual. The implicit

assumption is thus, that a bilingual can achieve an equal command of one or both languages to that of a monolingual. One is interested in measuring everything the bilingual has *in common* with the monolingual. With tests of the latter kind one is more interested in measuring the *differences* between the monolingual and the bilingual. One is looking for those specific features which are *characteristic of a bilingual* (as opposed to monolingual) *language proficiency*, and that is our main interest here.

8.4.2 Tests which use the monolingual as the norm

Tests of this kind have mainly evolved in three different sorts of context:

1. There are *pedagogically* orientated tests, which measure the prerequisites for participation in a language teaching programme (initial behaviour), or the results of the teaching of a foreign language or the mother tongue (terminal behaviour at the completion of the programme); often these tests start language teaching. In a number of countries tests have been developed for placing purposes. They determine, that is, whether a student is monolingual or bilingual enough to benefit from a certain kind of teaching (see, for example, the discussion and the references in Cummins, 1980b and in Dulay & Burt, 1980). The L1 of the learner is in most cases left out of account completely.
2. There are *linguistically* orientated tests, which earlier often measured simply grammatical competence, as defined by intra-lingual structural criteria.
3. There are *psychologically* orientated tests, which are based on investigations of the verbal factors of intelligence, and which are constructed with the cognitive and sometimes affective processes in mind that are manifested in linguistic behaviour.

Since my intention here is not to discuss the ordinary test *construction* at all, I shall confine myself to a consideration of such aspects of tests as tell us something about them as instruments for the measurement of bilingualism.

Many of the pedagogically orientated tests have been criticized for not being based on any linguistic theory, which could predict or explain the relation between the variables in the test. The variables to be measured have often been intuitively chosen, without reference to any theoretical model of linguistic proficiency (Wrede, 1971: 3); or they have been adopted because the grammar books use the corresponding categories, without any discussion as to whether these categories have any psycholinguistic relevance (Teleman, 1974). Linguists, too, are often guilty of this sort of procedure. Many tests constructed by linguists and psychologists are also used for pedagogical purposes.

The rapid progress in the last few years in both foreign language pedagogy and contrastive analysis, error analysis and especially

interlanguage studies has led increasingly to a tendency to make use of the learner's own production as an own "norm", attempting to establish through which phases the learning develops. In interlanguage studies errors are not seen as negative but rather as necessary positive stages in the process in which the learner constructs her own grammar as it gradually approaches the monolingual norm. Deviations from the monolingual norm are interpreted as indicative of the learner's trying out different rules and strategies in making the best possible use of a limited competence and repertoire (see Færch, Haastруп & Phillipson, 1983; Færch & Kasper, 1982, 1983; Haastруп & Phillipson, 1983). So far it has been mainly the *attitude* (and not the actual test construction) that has changed (and more in the formative than the summative phase – see Bloom, Hastings & Madaus (eds) 1971, especially Cazden), but interlanguage studies have contributed towards improving the theoretical foundation on which pedagogically orientated tests could be based, and some easing of the norms which have been unfair to bilingual speakers. Partly this is because it is linguists, people who know something about *language* and *language* learning (and not educationalists, who know about *language learning* and (language) *teaching*) who have worked with interlanguage. Even *linguistically* orientated tests have hitherto often been constructed by educationalists, psychologists or teachers. Many linguists have had decidedly little interest in testing (partly because they do not as a rule learn anything about it in their studies).⁴

The criticism that is now beginning to be directed, particularly by sociolinguists, at linguistically orientated tests is that they do not take into account how language is used in natural situations, and that tests provide an unnatural social situation, and thus produce distorted results. This should not, of course, be taken to mean that tests involving language used in so-called natural situations are always to be preferred for all purposes to tests in which more formal aspects of linguistic and cognitive competence are measured (see Cummins, 1980b). Quite the opposite – often more formal, CALP-type tests may be more informative when minority children's L2-proficiency is assessed for placement purposes.

Psychologically orientated tests have above all tended to consist of verbal parts of intelligence tests and further developments of these. Here, too, attention has begun to be focused upon making the test situations more meaningful by presenting the subjects with real tasks and with people they really have to communicate with.⁵

Many of the new developments are at a description stage only, and there are very few attempts to quantify. A combination of developmental (cognitive, linguistic) and situational (psychological, social, pedagogical) aspects, with both qualitative, in-depth data based on a longitudinal approach, and more quantifiable cross-section data, seems to be the ideal, but we are still far from it, and many of the attempts made at, especially, quantifying have attracted severe criticism.⁶ Forthcoming reports from the 5-year project on describing communicative competence at OISE (the

Ontario Institute for Studies in Education) Modern Language Center (Merrill Swain, Jim Cummins) will be of interest for the testing of bilinguals.

8.4.3 Tests which are used to measure bilingual linguistic proficiency

8.4.3.1 *What types of test have been used?*

The use of methods of measurement which test each of the bilingual's two languages separately, with monolingual proficiency as the norm, implies the two-code theory (5.2.1), as should be clear. The two-code theory is also clearly behind some of the methods of measurement used to try to isolate the characteristic features of bilingual competence. This reflects the negative attitudes: bilingualism thought of as something undesirable, as a transitional phase to be worked through as quickly as possible, almost as some kind of pathological linguistic condition. Attitudes like this were responsible for the description and enumeration of the characteristics of bilinguals as though they were the symptoms of a disease, a deviation from the healthy norm represented by monolingualism, deviations from two different norm systems. It has been said that the term "interference", for instance, carries with it this impression of bilingualism as something abnormal and unpleasant (Fishman, 1968: 29-30; Haugen, 1970: 1-12). As more and more sociolinguistic investigations made it plain that bilingualism is a permanent feature of many societies (Gumperz, 1964, 1967, 1969, 1971; Fishman, 1966, 1968, 1972c), theories about one single bilingual repertoire/one code developed. The adoption of such theories necessarily entails the development of special methods to measure the *bilingual code*, different from methods used to measure a monolingual code. Many different kinds of test have been used, and different researchers classify them in different ways (for a more detailed survey, see Skutnabb-Kangas 1975a). Linguists have often been particularly interested in measuring interference. Psychologists have developed different kinds of speed test and tests to measure automatic functions: how fluently a speaker speaks, how quickly she switches languages, how flexible she is, and to what extent one of the two languages dominates. Sociologists have examined to what degree and in what situations a bilingual speaker uses both languages. And educationalists have often been interested to measure the size of the repertoire in both languages. Some are trying to measure bilingual proficiency directly, some more indirectly.

If we survey the development in the last few years of tests to measure bilingual proficiency, we find that in fact, for all this research interest, disappointingly little has happened. A number of attempts have been made to construct methods of measurement specifically for bilinguals (for example, Burt *et al.*, 1975), but these seem to represent little improvement on the older tests, and to reflect little in the way of any new theoretical thinking. Many of the new ideas are coming from interlanguage studies and

from studies of testing in immersion programmes. In what follows I propose to discuss briefly some of the central notions that lie behind many of the tests and to draw attention to some difficulties in relation to them.

8.4.3.2 *Dominance versus balance*

The balance between the two languages has been used as a measurement of bilingualism, the assumption being that the more equal the balance between the languages the more bilingual the speaker is. We may doubt this for two reasons at least. Fishman has sharply criticized the use of balance between languages as a measure of bilingualism on the grounds that this defines balanced bilingualism as the ideal (Fishman, 1968: 25-26). In his view, bilingualism as a stabilized phenomenon can only exist where there is functional differentiation between the languages, diglossia. No society, he affirms, needs two languages for the same range of functions. He believes that a bilingual society produces exactly those kinds of bilingual whose one language is dominant in one area and whose other language is dominant in another. Such a society will *not* produce bilinguals whose languages are in balance. So a method of measurement with balance as the ideal is unrealistic. Moreover, balance as a measurement of bilingualism entails the assumption that a speaker whose languages are in balance commands both well. The possibility is ignored that a bilingual might have a poor command of both languages, speaking both equally badly rather than both equally well (see also sections 10.4-5 for a discussion of the quantitative aspects of double semi-bilingualism).

The method used to measure dominance has also been criticized. In principle this involves subtracting the score for the language the informant knows less well from the score for the one she knows better. There may also be additional operations. In this way the attempt is made to eliminate that area of linguistic command which is common to both languages, the cognitive operations which function in the same way in the two languages. The validity of the hypothesis upon which this attempt is based may be questioned. The way the mediating operations function (see Osgood's model in section 5.2.2) is far from clear, even for monolinguals and even less so for bilinguals. We do not know whether it is the influence of the g-factor (g = general) or only the verbal factor that the dominance measurement tries to keep constant through the subtraction operation. It is very difficult as a consequence to say how relevant dominance or balance between the languages is as *measurement of bilingualism*. Finally we may also criticize the measurement of dominance on the grounds that the dominance in any case is largely determined by the *selection of content* used for the test, at any rate for natural bilinguals who will always show a degree of functional differentiation determined either by external circumstances (the opportunities for using the two languages) or by themselves (certain preferences more or less chosen by themselves). Before the dominance measurement can be used as a measurement of bilingualism, the bilingual's linguistic

ability should be tested in a number of different domains to guarantee that the content of the test material is representative at least to a certain extent. As long as the dominance measurement uses material that takes no account of the domains in which a bilingual might be dominant in which language (and variations between speakers are likely to be so great that individual adjustment will have to be made for each speaker), the measurement of dominance that results will be largely a matter of chance.

8.4.3.3 Speed

The use of speed as a measurement of bilingualism (sometimes measuring the verbal speed of reaction in one of the languages only; sometimes as a measure of dominance, the difference in reaction speed between the two languages) has also been criticized. Speed has been chiefly used by psychologists (for example, Ervin-Tripp, Lambert), and criticized by sociologists (for example, Fishman). The tests used in the measurement of speed can be divided, for instance, into those that measure language comprehension and those that measure production. In order to know how quickly comprehension has occurred, the informant has to be asked to react in some way, so that it becomes clear that she has understood. Comprehension tests, and in particular production tests, have at times compared bilinguals' results with those of monolinguals, but it has often been a matter of determining the relative degree of bilingualism, i.e. of comparing the performance in each language (rather than using any external criterion), and thus obtaining a measurement of dominance. In fact, speed seems to be quite reliable as a measurement of dominance.⁷ However, the general objection still applies, that dominance is an uncertain measure of the degree of bilingualism, especially if performance in the two languages is much the same. With production tests, we may also say that relative speed tests measure more the degree of functional differentiation than the degree of bilingualism – the results depend to a large extent on the area from which the test material is drawn.

Besides, there may be changes when children grow older: there may be a generalized speed factor in younger children, which becomes more specified in older children, so that different group factors may account for, for instance, speed in physical and motor tasks as opposed to speed in verbal tasks (see Wissler, 1901; Burt, 1973a). Eysenck also shows that speed manifests itself importantly in information processing, and not in simple experiments with speed of reaction (1973: 247–51).

Fishman regards speed as a very ethnocentric kind of measurement: if the culture in which the bilinguals live does not value speed highly, it is pointless to use speed as a measurement (Fishman, 1968: 24–25; Liungman, 1970: 159). In many cultures slow, thoughtful speech is valued highly, and the kind of verbal speed normally measured in these tests will relate hardly at all to language command. Many of the measurements seem to measure a quite primitive, often almost automatic speed of reaction. Other tests have

measured, for example, how quickly informants are able to obey instructions in both languages, how many words they can remember correctly after seeing (or hearing) them for a limited period of time (or how long it takes them to learn the words by heart in the two languages), how many words they produce within a certain period of time in response to certain stimuli, and so on. Tests like this are often combined with dominance tests. The assumption has been that the more rapidly an individual can function in a language the better command she has of it, and that the language in which she reacts the more swiftly (or in which her production per unit of time is the greater) is the dominant language; a fairly primitive model. It is common for the reaction time of bilinguals to be slightly longer either in both or at least in one of their languages than that for the equivalent monolingual, and this is easy to understand, since the bilingual has more to choose between. This difference seems to apply above all to tests where the stimulus is ambiguous, in that it could belong to either language, or to tests in which the subject does not know in advance in which language the stimulus is going to be. The difference is less noticeable where the subject knows in advance which language is going to be involved. But, as Edith Mägiste says: "this delay is hardly a handicap if it is set in relation to the great gains a command of a new language confers. ... Tenths of seconds, which may be of decisive importance on the athletics field, are of no practical importance in most linguistic contexts" (Mägiste, 1979: 99).

8.4.3.4 Code switching, translation and interference as measures

Many researchers have used as one of the criteria for bilingualism the bilingual's ability to switch quickly or automatically between the two languages (see the definitions of bilingualism in section 4.3). This requirement may apply only to the primary proficiencies, so that one is able easily to *understand* discussions which alternate between two languages and is able oneself freely to switch from *speaking* the one to speaking the other when needed; the requirement may also be extended to the secondary proficiencies of reading and writing. The degree of interference ("instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language, i.e. as a result of language contact" (Weinreich, 1967: 1) has also been used as a measurement of the degree of bilingualism, especially by linguists. The more interference, the less bilingual, in this view – absence of interference has even been required in definitions of bilingualism. I disagree profoundly with this view, which seems to be based on the rigid use of monolingualism as the norm, making bilingualism and bilinguals negatively deviant.

We can discuss code switching and translation as instruments for the measurement of bilingualism starting from the assumption that these skills are usually learnt and practised by bilinguals. If we accept this assumption, then code switching and the ability to translate are good measurements of the degree of bilingualism. We may also start from the assumption that these

are special skills which are acquired by special training; if this is the case, then they might correlate only weakly with other measures of bilingualism, and the ability to switch codes and to translate would not be a reliable measurement of degree of bilingualism. Interference can be discussed likewise: is it bad command of L1 or L2 (or general sloppiness) that causes it (good measure) or is it used by most bilinguals in certain situations (then it would be a bad measure)?

It should be pointed out here that I will discuss code switching, translation and interference only from the point of view of how accurately they measure the degree of bilingualism. There is a huge general theoretical and empirical literature about them for the interested reader, both more purely linguistic and sociolinguistic (see for instance Clyne, Dittmar, Haugen, Hasselmo, Oksaar, Poplack, Sankoff, Vildomec, Weinreich in the bibliography, just to mention a few). Many of the case studies of bilingual children describe interference and code switching in detail, too. I will not discuss where the borderline is between interference and switching, neither structurally/situationally nor historically, nor the fact that the norms of bilingual speech communities are (and should be accepted as) different from the norms about interference (and what constitutes interference in *langue* and *parole*) in monolingual communities (although I have a section on that in the Swedish version of the book).

One of the main functional differences between interference and code switching on one hand, and translation or interpretation on the other hand is that the "same" content is reproduced in the latter but usually not in the former. Bilinguals who commonly use both their languages in speech must have considerable opportunities to practise code switching, except in those rare cases where the bilingual commutes between two wholly or nearly monolingual communities at fairly long intervals. If the community is bilingual, with or without diglossia or monolingual in one of the bilingual's languages and at the same time diglossic so that the bilingual's other language is used only in unofficial contexts (at home, with friends, etc.), then the bilingual must switch codes. How skilful at code switching a speaker becomes seems to a large extent to depend upon how widely separate from each other the situations are where the different languages are used, and upon how many bilingual situations and speakers the individual comes into contact with. Probably the most demanding situation is the one involving many bilingual and multilingual speakers as well as monolinguals, representing a variety of languages, and where a speaker must switch rapidly from one language to another as different individuals are spoken to, and where there are no clear rules as to which language is expected of one in conversation with other bilingual speakers. The situation is one where languages are switched rapidly, and where it is difficult to predict which language one will hear and which one will have to answer in. A test situation which uses code switch as a criterion for the degree of bilingualism is an artificially accelerated version of a very demanding situation of this kind.

For this reason, code switching, especially when linked to demands made on speed, probably is more a measure of the number of times a speaker has had to cope with situations where rapid code switching was required, than being a measure of the degree of bilingualism itself. Code switching tests can also be said to measure learning contexts. If we call to mind for a moment the distinction between compound and co-ordinate bilingualism, then it will be clear that an individual who has learnt the two languages in wholly separate contexts will probably find it more difficult to switch rapidly than someone who has learnt both languages in the same environment, and possibly from the same speakers.

In the same fashion it has been discussed whether translation as a skill should be wholly distinguished from bilingual ability (Halliday, McIntosh & Strevens, 1970: 142), or whether the ability to translate well (simultaneously and successively) is simply evidence of such complete bilingualism that only a few speakers ever achieve it. Many experienced teachers testify to the superiority of translation as a method of measurement. But on the other hand there are many bilinguals who complain that they find translation very difficult. This might be true in particular for co-ordinate bilinguals. There is, of course, a need for translation even in diglossic communities. Only in communities where everyone is bilingual does the need not exist, and according to Fishman, communities of this sort will tend to become monolingual. But even the contact between two monolingual communities necessitates translation. Therefore it seems reasonable to suppose that most naturally bilingual situations should produce bilinguals who are capable of translating, at least to some extent, or at any rate of interpreting between the two languages, with no exact details required. On the other hand, it is also clear that translation and interpretation at a high level are abilities that require special training. Simultaneous and successive interpretation even requires regular practice, since the act of translation is here performed at very high speed, as is not necessarily the case with written translation or less formal interpretation. Here, too, the environment is influential. A young child can be fairly skilful at translation if it has grown up in a bilingual family and gone through the phase of running back and forth between the parents and translating. The teaching at school is also important. If one of the bilingual's languages is taught as a second or foreign language while the other is the language of instruction, the child will probably get a good deal of practice at translation.

If exact and subtle equivalence is not part of the requirement, and if great speed is not expected, then translation may perhaps be used as a measure of bilingualism. But on the one hand, a very skilled bilingual speaker need not be a good translator, especially if speed is required; and on the other hand, an individual with a fairly limited knowledge of an L2 may nevertheless produce an acceptable translation, with the use of dictionaries, and given enough time.

I have described how, in bilingual families, many children both mix languages and have all sorts of interference in the beginning. Most of them can keep their languages apart before school age, and mix them thereafter mostly only when speaking to other bilingual speakers. It is important to investigate those situations in which bilinguals think that it is necessary to mix languages to some extent. According to many researchers a speaker may in certain informal situations make a rather unfavourable impression, be thought of as self-important and remote, or signalling detachment, domination or formality if her speech is "too pure". Many discussions among Swedish-speaking mother tongue teachers of the minority of Finland Swedes suggest that their pupils are interested in observing in which situations a mixture of languages is required and approved of, and in which situations it can create negative feelings in the listener and counteract the speaker's purpose by irritating the listener unnecessarily, for instance in an official situation where the pupil is in a weaker position than an adult partner.

There are contradictory reports about the awareness of bilinguals themselves of using interference, where third or fourth generation immigrants who are clearly dominant in the language of the new country seem to be unable to tell "correctly" to which language certain items belong "originally" (see Hasselmo, 1974). With more recent immigrants, stable indigenous minorities and individuals who have grown up in bilingual families the awareness seems to be greater, and many of them use code switching and interference consciously and elegantly for different stylistic, situational, etc. purposes. When interference signals a weaker command of one of the languages it often seems to be a situation which the individual knows of and may try to remedy (for instance asking about a word she doesn't know – see examples in Saunders 1982, especially the chapter "Departing from normal language choice in the family").

With a bilingual norm it becomes pointless to try to measure the degree of interference (in the monolingual understanding of the term), and to use this as a measure of the degree of bilingualism. If we are to use the monolingual's norm in deciding what is interference (every deviation from the L1 or L2 norm which is due to the speaker's familiarity with another language), then sociolinguistically sensitive bilingual speakers will be classified as less bilingual if they speak naturally, especially if the tester is bilingual, and thus leads the sensitive bilingual being tested to function according to the bilingual norm.⁸ If we use a bilingual norm in deciding what constitutes interference (failure to adhere to the bilingual community's own norms in the choice of language varieties: L1, L2, or L1 + L2), then what we are measuring is not the degree of bilingualism, but sensitivity to the bilingual norm and willingness to adhere to it.

If this argument is accepted, then the only way to use interference as a measure of degree of bilingualism is to explain in the test instruction that the test situation requires the bilingual speaker to speak L1 or L2 in as "pure" a

form as she can, because the communication partner is absolutely monolingual. But there are difficulties about this too. Who will decide, for example, what kinds of interference actually impede intelligibility, and what kinds perhaps have the effect of making the language more varied, enjoyable and creative?⁹

To summarize, in my view use of the amount of interference as a measurement of the degree of bilingualism rests on a negative view both of bilingualism and interference, a one-dimensional and puristic attitude which can only be harmful. To my mind this kind of measurement is at worst actually objectionable and at best problematic in that it tends to penalize linguistic creativity and social sensitivity and to define this as "less bilingual". Measurement of this sort should be used with very great care, if at all.

What I have said holds equally for interference (or transfer, as it is called in language learning studies) as it occurs in learner language, whether it is children or adults who are learning a second language, i.e. in the process of becoming bilingual. There are many different opinions as to how much the learner's deviation from the target language is the result of influence from L1. Burt & Dulay's calculations suggest a figure for syntax as low as 5%, for example, though most other researchers put the figure higher. In spite of the diversity of view here, interlanguage studies have helped to establish a more positive view of interference. The learner's use of the resources of L1 in situations where the norms of the target language are still unclear is seen as one positive communication strategy rather than as an illicit intrusion into the territory of the target language, even if it does not always help (see Haastруп & Phillipson, 1983).

8.5 To test or not to test – that is the question

As I have said before, different attitudes are possible to the business of testing bilinguals.

■ First we may say: tests are unreliable, they give a distorted picture of reality, they reduce the individual, they can never measure or capture what one is capable of doing with language in real situations, and tests can never catch all the important features anyway. The devising or use of tests is something one ought not to get involved with.

Let me give you an extreme example my friends were discussing once when I was in Berlin (West). The experts on immigrant language in Germany had been brought together to decide how immigrant guest workers in West Germany should be tested in German when they applied for German nationality. In the first place, all my friends were absolutely opposed to the idea that a good knowledge of German was necessary for the obtaining of German nationality – they felt that this condition was unjust so long as ordinary guest workers in the country had so little opportunity of

learning the language. Secondly, they thought it impossible to construct and administer a test that was multi-faceted enough to reveal in real terms how the subjects, for example, would manage in German in their place of work. Thirdly, they thought it grossly unfair to put workers through a test in standard German who had learnt their (non-standard) German at work – but of course it would have been impossible to devise tests for all the variety of dialects and sociolects. On the other hand, if they were to refuse to devise a test, then the command of German would be tested by police officers with no linguistic training whatever, and simply in the course of a conversation, or the police might devise a test. It is often a matter of choosing the least bad alternative rather than agreeing only to co-operate if something really satisfactory could be done or refusing to co-operate. If we don't do the testing, someone else might, who might know even less than we do about bilingualism. P.S. My friends chose to devise a test. . . .

■ Second we may say: since there are no really good tests anyway, I will make use of this kind of enquiry, which is not a test, but which will still provide me with some information, with the practical data I want. This is not a test, it is no more than . . . whatever you like to call it. And so we use something we have put together ourselves (without enough scrutiny of it) or have taken over from a colleague, something that fails to meet even the most elementary of the requirements for a satisfactory test or questionnaire. Something that could have been improved had we had any acquaintance at all with research on testing procedures. There are frightening examples of "tests" and "questionnaires" of this kind used in official enquiries. The information they provide is often likely to be more misleading and mischievous than no information at all, since those who are given the results (and unfortunately sometimes even the very people who have conducted the enquiry) believe they derive from the use of solidly respectable tests and testing procedures and provide, at any rate, reliable basic information. Testing in these cases has undoubtedly been a choosing of the greater evil in a situation where there is no really good alternative available.

■ Or we may as a third alternative concentrate so much on refining the instruments to be used for measurement that this task comes to occupy more of our attention than the real problem that measurement was supposed to elucidate. This kind of attitude is well illustrated, to my mind, by the rather alarming way in which the discussion (one might almost say the agitation) about Pertti Toukoma's remarkable investigations of immigrant schoolchildren has proceeded. The attacks made upon him have been full of errors and untruths (see further my critique of Ekstrand's methods in Skutnabb-Kangas, 1979c and 1980a). "These attacks have in general avoided discussing directly the problem he focuses upon, its social origins and consequences, and have instead concentrated on discussing matters of technical pernicketiness. For example, psychological tests used for sociolinguistic and educational purposes have been discussed as though their purpose had been the technical, linguistic one of revealing exactly where and

how immigrant children's linguistic ability is deficient. The problem that should be discussed is *why* immigrant children do so badly, worse than majority children, in different kinds of measurement of linguistic ability and school achievement, and what the long-term consequences of this may be, for them and for our society. By comparison with this, the question of how the different tests are constructed seems quite immaterial: immaterial too because these are tests that are also used (and standardized) for monolingual children. For immigrant children to be able to cope well in schools where abilities of the kind that these tests measure are required, they *must* acquire them, *whatever they are*. By concentrating on a secondary question, by reducing the question about immigrant children's poor school achievement and its causes to a question about testing technique, the real issue is avoided. It may indeed be interesting to discuss technical matters for their own sake too, and we do need discussion about the validity of various methods of data collection, but if these technical discussions are used (by the researchers themselves or by others) to deflect attention from the real problems and to discredit critical research which on the basis of theories about minority education tries to view the problem in its whole social context, highlighting the reasons for and the consequences of the present state of affairs, then we are guilty of just such mistakes . . . as characterize the analytical positivistic philosophy of science attitude at its worst" (Skutnabb-Kangas, 1979b: 172-73).

The reader interested in questions of pedago-political consequences of minority education research and its different philosophy of science paradigms, ideologies and methods can find a detailed account in a book comparing North American and Scandinavian debates (Cummins & Skutnabb-Kangas, in preparation).

The three different attitudes I have described may be seen to run parallel with the three different stages in the development of language testing described by Bernard Spolsky (1976): the pre-scientific; the psychometric – structuralist; and the psycholinguistic-sociolinguistic. The pre-scientific attitude would be the one characterized by an uncritical reliance upon an intuitive view of what are the important variables to be measured, with no attempt made to systematize them or give them a basis in theory about language as a system, language use, or bilingualism. The psychometric-structuralist stage brought technical precision and dexterity, and statistically reliable methods, but with little attention paid to the material in the tests, or the purposes for which the tests were being used. This tendency in its extreme form led to the quite barren discussions in Scandinavia about the concept of semilingualism. At the more psycholinguistically-sociolinguistically orientated stage we begin to see both how difficult testing is (and sometimes to refuse to test for this reason), and also to try to develop ways of describing language usage which might provide a basis for more realistic methods of measurement. We begin, too, to take into account for which purposes and in whose interest tests are used.

There is a further danger worth pointing out here, that in our eagerness now to test what people actually *do* with their languages (instead of or in addition to testing what they *could* do), we only test what people do in concrete everyday communication situations. We construct tests which imitate certain types of ordinary communicative situations: for example, explaining something to another person in a "real" situation (where the other person is dependent on the explanation given – typical examples are tests where two subjects are separated by a screen with the same kind of landscape before them, and one of them is asked to give the other directions as to how to reach a certain goal). Or we test children to see whether they understand the basic vocabulary and syntax used in school textbooks. All this is in reaction against the "unreal" tests which required conscious manipulation of language in a way that is not demanded by everyday face-to-face communication. The danger with this more sociolinguistically orientated kind of testing is that we forget that language is also used for purposes other than everyday communication, and that a child who can manage only this ordinary sort of communication (giving someone directions, for instance) might not in fact have a good enough command of the language to be able to use it for academic work at school. Jim Cummins has warned us particularly of this where language tests are being used for school placement purposes (see section 5.2.2.5 and Cummins 1980b) where often only the child's surface fluency has been tested and not the ability to use the language as a cognitive tool.

Notes to Chapter Eight

1. Here is an example from my own experience. I met a teacher, a Finnish-speaker, from the Torneå Valley, Sweden, at a course for teachers, who at the beginning of the course denied in rather hurt tones that she knew any Finnish. The course was on bilingualism: among the things dealt with were attitudes, the consequences of long-term linguistic oppression, a colonized consciousness and the shame individuals feel at belonging to a linguistically oppressed group. After a few days at the course, she began to speak lively but slightly halting Finnish, and of her own accord took up with me the question of why she had denied she had any Finnish at the beginning.
2. See, for example, Koironen, 1966; Macnamara, 1969: 86; Jaakkola, 1969: 80–86; Lindman & Alho, 1971: 34; Sandlund, 1972: 13–14; Sahlman-Karlsson, 1974: 140.
3. An example is provided by the investigations conducted by Arne Trankell and the IMFO group of the prejudice Swedes feel towards immigrants (see, for example, Landén, 1971; Montelius, 1971; Trankell, 1971, 1973, 1974). Another example, where children's attitudes have been studied, is the research carried out by Mirko Takáč and his group in Gothenburg (see Carlsson & Valík, 1979; Takáč,

1978). There are very many elaborate and less elaborate descriptive reports on attitudes to immigrants in Scandinavia, for the most part concentrating on Sweden. The reports from the Commission on Discrimination (Sweden) summarize a number of them. I refer the reader to them.

4. Though they have been quick enough to criticize, often knowing very little about test construction – see, for example, Bengt Loman's critique of tests used by Pertti Toukomaa in Loman, 1978a: 39–40, where it emerges that Loman does not know that the test manuals normally set out the criteria for scoring, i.e. give criteria for and examples of acceptable and unacceptable answers.
5. See, for example, the tests to measure communication in schizophrenic families developed by Rolv Mikkelsen Blakar and his group in Mossige, Pettersen & Blakar, 1979, these have also been used for measurement of sexual differences (Blakar & Pedersen, 1980a, b).
6. See, for example, the discussion that followed Norbert Dittmar's lecture introducing implicational scales, that is, scales in which the attempt is made on the basis of frequency of occurrence to arrange different rules about verb phrases in order of learning, so that the knowledge of a "later" rule implies an "earlier" rule already (partly learnt (Dittmar, 1978).
7. For example, Lambert, 1955; Lambert, Havelka & Gardner, 1959; Preston & Lambert, 1969; Rao, 1964; Mägiste, 1979.
8. I often find it difficult to trust results about the language use and linguistic competence of bilinguals if the test has been conducted by someone who does *not* speak both languages – and this is to direct fairly fundamental criticism at most Swedish projects on bilingualism, which have been conducted often by Swedish researchers with no knowledge of the immigrant language they were investigating. Here, too, Hansegård is an exception, with his excellent command of Same and Finnish.
9. I believe, for example, that some of the Danicisms that I have started using in my Swedish make it richer – they are immediately intelligible to all Swedes, even though it is quite clear that my use of them depends actually on my familiarity with Danish.