The native / non-native speaker distinction and 
the diversity of linguistic profiles of young people 
in Swedish multilingual urban contexts

Kari Fraurud & Sally Boyd 
Stockholm University & Göteborg University

Abstract: The notion of ‘native speaker’ (NS) has played a central role in all areas of linguistics, but it is also perennially questioned. This paper aims to contribute to the discussion of the usefulness of the binary distinction between native and non-native speakers (NNS) by exploring a relatively large body of empirical data collected in a study of language and language use among young people in contemporary multilingual urban settings in Sweden. Data about linguistic background and practices from 222 informants were analyzed by means of so called ‘linguistic profiling’ – here involving a number of variables reflecting various nativeness criteria. The resulting complex and varied linguistic profiles display the great diversity among informants. This diversity is presumably not unique to these 222 young people, but can also be expected to be found in other similar contexts. The application of a binary NS/NNS distinction in such contexts will either – if a single criterion is used – result in a categorization of informants into two widely heterogeneous groups, or – if multiple criteria are combined and only clear cases considered – result in the exclusion of a considerable number of language users from the object of study. These observations should also have implications for the study of language variation and change in multilingual contexts more generally.

1. Introduction

Language contact in many contexts around the globe has led to a revival of the perennial discussion of concepts such as ‘native / non-native speaker’ and ‘first / second language’. In this paper, we would like to contribute to this discussion by contrasting the first of these binary distinctions with a ‘linguistic profile’ analysis of data from 222 young people living in multilingual settings in Sweden’s three largest cities. While our conclusions relate to this particular group of informants, we believe that several of the observations made in this paper are also applicable in studies of young people in other contemporary European multilingual settings and, more generally, in any study of language variation and change involving multilingualism.

---


Our main point is that it is not only the case that the native/non-native speaker distinction turns out to be difficult to apply in some, more or less exceptional, individual cases. Rather, the distinction appears to be fundamentally unsuited to capture the linguistic background, practices and proficiencies of an important number of young people today. In our study, an overwhelming majority of the 222 informants did not fit neatly into any of the categories ‘native’ or ‘non-native’ as defined in the literature. There is still a need for characterizing and generalizing over informants, however, and as an answer to this need we have developed a method of analysis that we call ‘linguistic profiling’, further described below.

The data we use has been collected and analyzed within the project ‘Language and language use among young people in multilingual urban settings’. The overarching goal of the project is to describe, analyze and compare the everyday language and language use of young people in multilingual areas of Göteborg, Malmö and Stockholm. The project is being carried out by a team of senior researchers and graduate students, who are studying various phonological, grammatical and pragmatic features of the varieties used by young people with different linguistic backgrounds in these settings. The first step in the fieldwork was a series of interviews with 222 young people (age 16-21, median 17) in our target informant group. In an earlier paper, we looked at the concept of ‘second language’ and ‘second language speaker’ (Boyd & Fraurud 2004) in relation to this informant group; in this paper, we examine critically the related concept of ‘native speaker’ (NS) to assess its usefulness in studies such as ours. We begin by briefly looking at this concept in various branches of linguistics in order to see which criteria are normally used to identify native speakers. Then, using these criteria as a point of departure, we look at the diversity of backgrounds represented in our informant group, and propose, rather than using one or two of the suggested criteria, combining several criteria to produce a multitude of ‘linguistic profiles’. In the final part of the paper, we suggest ways in which a linguistic profile analysis can give a more accurate and detailed picture of the diversity of linguistic backgrounds in settings such as the one we have studied.

2. Definitions and uses of ‘native speaker’ (NS) within linguistics

In an anthology from the early 1980’s, A Festschrift for Native Speaker, which examined the concept of native speaker, the editor Florian Coulmas (1981:1) calls

---

3 Cf. http://hum.gu.se/institutioner/svenska-spraket/verk/projekt/pag/sprakbruk_eng/. We would like to acknowledge the practical assistance and helpful comments of our colleagues on the project, as well as those of our colleagues at our respective departments. We are also grateful for the financial support of The Bank of Sweden Tercentenary Foundation, which has financed the research project as a whole.

4 The informant group is not a random sample of young people living in these areas. Rather, schools and classes in these schools were selected in order to fulfill various criteria of representativity and comparability among the three cities. All the students in the chosen classes were invited to participate in the study, and all but a few accepted.
this concept “a common reference point for all branches of linguistics” [...] “of fundamental importance to the field.” Despite the centrality of the concept however, it turns out, not surprisingly, that the concept is conceived of and used somewhat differently by different linguists. There is, on the one hand, the idealized native speaker as a theoretical construct in, e.g., generative grammar. Chomsky’s view is that (1963: 3): “Linguistic theory is concerned primarily with the ideal speaker-listener, in a completely homogeneous speech community, who knows its language perfectly”. On the other hand, more empirically oriented linguists (some of whom may actually also need an idealized native speaker for their theories) are faced with methodological problems regarding whom they should choose as their source of empirical data about a particular language. Others are in need of a native speaker as a norm (a control group) against which to measure non-native speakers of the language. Our paper is oriented toward the question of the usefulness of the concept of NS in empirical work, i.e. the second conceptualization, in particular in studies of language and language use in contemporary multilingual contexts.

In his introduction to the *Festschrift*, Coulmas (1981) makes a conceptual analysis which takes up several of the central criteria for a native speaker of a certain language. The criterion he considers to be most necessary and common to all conceptualizations of NS is that of language acquisition in early infancy, in conjunction with primary socialization (Coulmas 1981: 4). This criterion is closely related to most meanings of the term ‘native’ in the collocation ‘native speaker’.

In the same volume, Thomas Ballmer (1981) presents a typology of native speakers, where he enumerates a number of additional criteria. In the passage quoted below, he summarizes the practical problems of finding a suitable native speaker for typologically oriented or other studies within the broad field of descriptive linguistics. He draws the conclusion (1981: 57) that a good candidate can actually be difficult to find:

“A native speaker is old enough to know the language and not so old as to have forgotten it. He is healthy in every relevant respect, thus especially neither blind, deaf, handicapped, paralysed, nor does he lisp, stutter or have a cleft palate. He is monolingual, he lives in his birthplace, his family, especially his mother, speaks natively his natural language L, the place where he lives is strictly monolingual: there is no standard speech/dialect split and there are not other competing languages. [...] The native speaker is educated enough to enter the experiment, but not so educated as to call into question its outcome”.

Taking Coulmas (1981) and Ballmer (1981) as points of departure, we can say that these sources bring forward several important criteria for NSs, primarily thought of as adequate representatives for a community of speakers of a language. Also within sociolinguistics, the concept of native speaker has been needed for practical reasons, as a necessary criterion for informants. In his dissertation, Labov (1966) describes how he selected informants for his New York City study, based on an earlier random sample carried out on the Lower East Side. However,
Labov excluded almost half of the persons selected for the earlier study because they were born abroad, had not come to the neighborhood until after the age of eight, or had “marked foreign characteristics” in their English (1966: 175). Of the original sample of 553 individuals, Labov’s sample comprised only 312, or 56%. His criteria for selecting informants thus included birthplace, monolingualism and sedentaryness, while age of (dialect) acquisition is operationalized in terms of age of arrival in New York City.

Moving on to the area of second language acquisition (SLA) research, the concept of native speaker here builds to a high degree on the Chomskyan conception that universal grammar and the language acquisition device (LAD) is not available to (at least late) second language learners. The definition of the field is in fact dependent on a conception of second language learning or acquisition as different from first language acquisition. Usually, the distinction is conceived of in terms of age of acquisition or age of onset. In addition, native speakers are often selected to act as standards of comparisons to persons learning a second language; they are conceived of both as necessary sources of input and as targets for the second language learning process. This role accentuates native speaker proficiency as an important further criterion. Proficiency is either implicitly assumed to be entailed by other criteria, or explicitly specified (as in Davies 2003, 2004) to include intuitions about one’s own idiolect and its relation to standard grammar, one’s fluency, communicative competence in production and comprehension, creative writing ability, etc.

In a paper outlining future directions for research in the field, Michael Long (1993: 205) wrote about experimental studies comparing near-native speakers and native speakers:

“NS (native speaker) subjects clearly need to be more comparable to NNS (non-native speaker) comparison groups than has sometimes been the case in the SLA literature in the past. The most crucial requirement [is] that NSs speak the L2 variety or second dialect that was the acquisition target for the NNSs […] In addition NS controls should be comparable in age, sex, education and social class […] Ideally, they should also be monolinguals, since there is increasing experimental and anecdotal evidence that learning additional languages can sometimes affect first language abilities in as yet poorly understood ways”.

Here, then, we see some similarities to Ballmer’s conception, quoted above, in that Long considers the native speakers’ age, sex and education to have a potential effect on their native speaker competence. Crucially, they should also be monolinguals, because multilingualism can have an effect on first language abilities (see e.g. Jaspaert et al. 1986, Cook 2003).

Also in foreign language research and teaching, the concept of native speaker has routinely been taken for granted as a norm for students in the classroom. In response to arguments against the feasibility of such a norm for all learners (e.g.

---

5 As evidenced in, e.g., the vast literature on ultimate attainment, opinions differ widely as regards at what age this happens (cf. Hyltenstam & Abrahamssson 2003).
The native / non-native speaker distinction and the diversity of linguistic profiles

Kramsch 1997/2003), several researchers defend the need for the native speaker norm in foreign language classrooms (e.g. Koike & Liskin-Gasparro 2003: 263):

“[I]f the native speaker is not to be considered the model for learners to emulate, then who should provide that linguistic model? […] [Without the native speaker,] the teachers and learners of foreign language are left (1) without a target language norm, and (2) with an unrealizable dream of becoming native-like speakers. These two notions render any efforts in the FL classroom unproductive”.

In this passage, the authors bring out another proposed quality of the native speaker, that of identity. For many, the native speaker should identify her/himself and be identifiable (or at least able to pass, cf. Piller 2002) as a full-fledged member of the speech community by other native speakers.

Drawing on these and other sources, we summarize our brief survey with the following list of common criteria for the concept of native speaker. The first criterion, the essential one according to most definitions, is that of early exposure during primary socialization, the ‘age of onset’ criterion. Other criteria vary in importance in the various sources or are assumed to be entailed by age of onset. They include continued use of the language throughout the lifespan; a particular (usually high) level of proficiency in the language; an adequate, but not too high level of education (or, similarly, class); monolingualism (albeit to different degrees); and a number of criteria that can be subsumed under the heading of context and that assume a homogeneous environment without influences from other varieties or languages (see further below). A final criterion is having an identity as a native speaker.

Like many others, we perceive a number of ideological, theoretical and methodological problems associated with the use of the binary NS/NNS distinction, and with the NS criteria and their operationalization. In this paper, however, we will confine our discussion to some issues raised by our empirical material.

3. Applying the concept of native speaker in a contemporary multilingual context

No one acquainted with current multilingual settings can have failed to meet individuals whose linguistic background and practices challenge the binary NS/NNS distinction (for a particularly insightful analysis, see. Leung et al. 1997). It is still common, however, both in research and in popular discourse, to talk and act as if such individuals would be more or less exceptional; the majority of speakers in a community or group is thought of as being either NS or NNS. Behind such a ‘dichotomy-with-exceptions view’ lies an underlying assumption of homogeneity, implying a strong association between the various nativeness criteria mentioned above. In this study we therefore wanted to explore the NS/NNS distinction in relation to a larger database, built up within the project mentioned above.
Background data about the 222 young people participating in the project was collected through a questionnaire filled out by the researchers during individual audio-recorded interviews. The questionnaire comprised 80 items, consisting of both multiple choice questions with follow-ups and open questions, generating a high degree of detail and heterogeneity in the replies. In addition to the pre-processing required in encoding data into SPSS, the profiling procedure illustrated in this paper involves making the database more manageable by further grouping values and merging variables at different levels of detail. It should be remembered that, while each such step increases comparability and generalizability, it is at the cost of loss of (possibly important) information. We will return to this below.

3.1 Operationalizing NS criteria

A large number of background variables in the database are potentially relevant to an assessment of our informants’ nativeness according to the criteria reviewed above, including also such variables as, for example, writing practices and media consumption preferences. For the present discussion, however, we have selected some variables that more immediately relate to the NS criteria. These are listed in Table 1 with brief descriptions.

Table 1: Native speaker criteria and some related background variables in the database, all reported data. (SS=Swedish/Sweden only, SO=Swedish/Sweden and other, OO=Other only. Values within parentheses are theoretically possible but non-occurring values.)

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Some related background variables in the database</th>
<th>Values</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of onset</td>
<td>age of onset of Swedish</td>
<td>early/late</td>
<td>AGEON</td>
</tr>
<tr>
<td></td>
<td>age of arrival to Sweden</td>
<td>early/late</td>
<td>AGEARR</td>
</tr>
<tr>
<td></td>
<td>where Swedish first spoken</td>
<td>home/other</td>
<td>SWFIRST</td>
</tr>
<tr>
<td></td>
<td>first learned language</td>
<td>SS/OO/OO</td>
<td>FIRSTLG</td>
</tr>
<tr>
<td>Proficiency</td>
<td>best language(s)</td>
<td>SS/OO/OO</td>
<td>LFPROF</td>
</tr>
<tr>
<td></td>
<td>proficiency in mother tongue, mother tongue 2 &amp; Swedish, respectively: speak/understand/write/read &amp; understand</td>
<td>1–5</td>
<td></td>
</tr>
<tr>
<td>Use</td>
<td>preferred language(s)</td>
<td>SS/OO/OO</td>
<td>LGPREF</td>
</tr>
<tr>
<td></td>
<td>language(s) used with parents</td>
<td>SS/OO/OO</td>
<td>LGPAR</td>
</tr>
<tr>
<td></td>
<td>language(s) used with siblings</td>
<td>SS/OO/OO</td>
<td>LGSIB</td>
</tr>
<tr>
<td></td>
<td>language(s) used with close friends</td>
<td>SS/OO/(OO)</td>
<td>LGFRIEND</td>
</tr>
<tr>
<td>Context</td>
<td>born in Sweden / abroad</td>
<td>SS/OO</td>
<td>BORN</td>
</tr>
<tr>
<td></td>
<td>parents born in Sweden / abroad</td>
<td>SS/OO/OO</td>
<td>PARBORN</td>
</tr>
<tr>
<td></td>
<td>mother tongues of close friends</td>
<td>SS/OO/(OO)</td>
<td>FRENGL</td>
</tr>
<tr>
<td>Mono-/bilingualism</td>
<td>several, including languages spoken, language practices in relation to various contexts, media etc.</td>
<td>[open]</td>
<td>[several]</td>
</tr>
<tr>
<td>Education &amp; social class</td>
<td>several, including education and occupation of parents/step parents in Sweden &amp; country of origin, living area and type of housing, etc.</td>
<td>[open]</td>
<td>[several]</td>
</tr>
</tbody>
</table>


For several of the relevant variables in our database it makes sense to divide responses into three main categories: Swedish/Sweden only (SS), Swedish/Sweden and other (language/country) (SO), and Other (language/country) only (OO), as shown under Values in Table 1. For reported age of onset and age of arrival, in the present illustration of the profiles we use a binary division of responses into early (before three years of age) and late. Just as in the case of selection of specific variables, this is but one of several possible options. Responses to the open question as to where Swedish was first spoken were divided into home, contrasted to all other settings, such as at pre-school or school, or with friends. The values Swedish/Sweden only (SS), early, and home will collectively be referred to as ‘NS-values’, as opposed to all other, ‘NNS values’. Almost all variables involve a number of theoretical and methodological problems, of which we can only briefly mention a few here.

Determining an individual’s age of onset (AO) is far from trivial, at least for individuals who were not born into families and communities where only one language is used. Somewhat surprisingly, the notion and its operationalizations are rarely problematized in the literature. Theoretically, it is not clear what conditions should be fulfilled for an individual’s contact with a language or language development to be counted as an onset of language acquisition. It is possible to imagine a broad continuum from active interaction to mere exposure to the language, which may be infrequent, only receptive, limited in mode (e.g. only television), etc. Methodologically, determining AO is problematic since in most cases one must rely on reported data, which may have low reliability due to respondents’ insufficient memory of details on language use early in life and their different interpretations of ‘onset’ (or ‘when you start to speak/learn the language’ or whatever wording is used in the question). Operationalizing onset by replacing it with (probably more reliable) reports of age of arrival at a specific place may reduce the validity of the data; arriving in a country does not necessarily imply starting to acquire the majority language in any other sense than that of minimal exposure. In our background database we are of course faced with the same reliability and validity problems, and we here try out the possibility of combining and comparing several onset-related variables. In addition to reported AGEON and AGEARR, we include the variables FIRSTLG, reported first learned language(s), and SWFIRST, whether Swedish was introduced within or outside the family, which in many cases indirectly indicates age of acquisition.

Language proficiency as assessed by analyses of oral and written production and various tests is clearly relevant in descriptions of linguistic profiles insofar as the tests are valid and relevant to the research questions. It should be noted, however, that such tests/analyses inevitably involve problems with the choice of norm against which to measure. This is something which is especially pertinent in our context, where the presence of alternative language models due to the emergence of new multiethnic youth varieties complicates the analysis of non-standard features (cf. Kotsinas 1988, 1992; Bodén 2004; Fraurud & Bijvoet 2004). Since our database does not include test data, we here only consider self-reported
data, in particular the responses to the question “Which language or languages do you think you know best?”.

With regard to language use, we include the responses to one general question, “Which language or languages do you prefer to use?”, and three questions of habitual use: language(s) used with parents (and step parents), siblings and friends. Even these questions tend to hide important quantitative as well as qualitative variation. The code ‘Swedish and other’ for language use with parents can e.g. include instances of code-switching, varying language choice in different situations, speaking one language to one parent and one to another, parents’ speaking one language and children speaking another (in a single conversation) and many other variants.

Context is in Table 1 represented by but a few of several possible variables. Country of origin (of parents and child) is often referred to in studies of bilingualism and SLA in Sweden, partly due to the fact that it – together with citizenship – is the only resource in official Swedish statistics for conjectures about ethnic or linguistic affiliation (cf. de Geer 2004). As elsewhere, people who were born abroad or have parents born abroad are often grouped together under the label ‘first and second generation immigrants’, subsequently used for generalizing about this ‘group’ as opposed to ‘(native) Swedes’. Obviously, the birthplace of an individual or of his/her parents does not necessarily imply anything about his/her mono-/multilingualism. Still BORN and PARBORN are included in Table 1 under ‘Context’, as two of several variables that indicate the potential presence of languages other than Swedish in an individual’s earlier and current environment. The variables AGEARR and SWFIRST, listed under Age of onset, can, or perhaps should, also be seen as context variables. Another possible source of linguistic influence, especially for this group of young people, is the mono-/multilingualism of friends, here reflected in the question: ‘What are the mother tongues of the friends you see most?’ (FRIENDLG).

Mono- and multilingualism (to different degrees) can in our database be traced in responses to several different questions, including one open question of currently spoken languages (see below). It is also reflected in the proficiency, use and preference variables that we described above, and which we will use here.

Education and social class, finally, are reflected in our database with answers to some open questions about parents’ education and earlier and current occupation etc. They are highly complex variables, especially in settings like ours, where migration sometimes involves dramatic changes in the socio-economic status both within and between generations, while information about education in countries of origin is difficult to assess. Furthermore, the linguistic literature using the native speaker concept gives different signals about which social class or level of education is ideal for a native speaker. The criteria of education and class is not included in our present analysis.

6 Recent statistics, however, use the term ‘foreign background’, applied either to persons born abroad and/or with both parents born abroad, or only to those born abroad and/or with a foreign citizenship.
3.2 Distribution with regard to nine ‘NS variables’

Let us consider the distribution of some putative nativeness variables (hereafter ‘NS variables’) over the 222 young people participating in our project. According to the simplistic ‘dichotomy-with-exceptions view’ described above, our informants would split into two main groups, clear native and clear non-native speakers – leaving a small number of unclear or mixed cases in between. In our database, there is one variable that produces this kind of distribution, viz. reported first learned language(s) (FIRSTLG, cf. Figure 1).

It is illuminating to evaluate the responses to questions about FIRSTLG and the other onset-related variables, AGEON, AGEARR, and SWFIRST, compared with each other. The separate frequencies for NS values on all four onset related variables are: AGEARR=early (73%), AGEON=early (58%); SWFIRST=home (45%), FIRSTLG=SS (45%). Pair-wise comparisons show that the overlap between these variables (including the two latter) in our data is only partial, and – more importantly – that there is no absolute implicational relationship among them. Thus, an ‘early’ age of arrival does not imply an ‘early’ age of onset. In our dataset there are 35 exceptions to this assumption, with a difference of up to seven years, while an ‘early’ age of onset does not imply learning Swedish at home. In this instance, there are 30 exceptions. Furthermore, the reliability and validity problems connected with AO, discussed above, may be even more serious for responses to the question of which language was learned first. For example, nine informants claimed Swedish (only) to be their first language and to have started to
speak Swedish outside the family, two of them also saying that this was after three years of age. This and other observations suggest that the questions of which language(s) one learnt first and when are open to different interpretations and are difficult to answer.\footnote{Some of these problems could, of course, be partially overcome by more extensive fieldwork, such as detailed interviews with informants, care-takers and other members of the families.} It may be easier to remember \textit{where} one started to speak/learn a language than \textit{when}, hence the variable \textit{SWFIRST} in an important way complements the other age related variables.

As may be expected in a group comprising many young people with immigrant parents or grandparents the distribution of \textit{FIRSTLG} and other onset-related variables contrasts sharply with those of current best language (\textit{LGPROF}) and preferred language (\textit{LGREF}). A large majority of young people consider Swedish to be the language they know best and a large (but not identical) majority prefers to speak Swedish. For the group as a whole, then, this may suggest a general trend of language shift, but, as we will show below, the picture is in fact more complex.

Translating central nativeness criteria to variables in our database, the NS group among our informants could, as a first approximation, be operationalized to include those who (i) have acquired Swedish in the family at an early age, as their only first language, and in Sweden, (ii) consider Swedish their best and preferred language, and (iii) are monolingual in the restricted sense of using only Swedish with parents, siblings or friends (besides possessing no additional first/best/preferred language). Nearly one third of the informants fulfill all of these nine NS criteria (the 30\% ‘All NS values’ in Figure 2). Very few, 7\%, did not have an NS value on any one of the nine criteria. An overwhelming majority of the 222 young people in our study, therefore, falls ‘in between’ – with a combination of NS and NNS values. The exclusion of this large group of informants from the object of study is of course no option if we want to study language and language use in current multilingual settings. Hence we need to develop tools for capturing possible general patterns among our informants without imposing new unwarranted categories or assumptions.

3.3 Multilingual profiles of 222 Swedish adolescents

As we have already pointed out, our initial analyses of the data indicated that there is no single nativeness criterion which produces a neat division of our informant group into two homogeneous groups. Instead, we need a method that allows us to explore and systematize the variation in our informant group prior to further both statistical and qualitative analysis. For our needs it is essential to be able to group informants according to linguistic background and practices (i) at different levels of detail and (ii) without losing track of either the individuals within groups or the different variables and possible combinations of values. What we call \textit{linguistic profiling} analysis provides such a characterization of groups and individuals through combinations of (a variable number of) values of (a variable number of) variables hypothesized to be relevant for the particular study at issue.
In the present application of this method, we wanted to explore common nativeness criteria against empirical data. We have here used eight of the nine variables in Figure 2 above: three with two, four with three, and one with four possible values. It should again be emphasized that this particular choice of variables and values is one of several possible alternatives for illustrating the complexity of the data with regard to nativeness criteria by means of the profiling procedure.

Altogether, 88 combinations of values were represented among the 222 informants (see Table 2). In order to save space, only 26 of the profiles are included in the table: all profiles shared by three or more informants plus a few others with only one or two members, illustrating different points in the discussion.

In Table 2, each row represents a unique combination of values on eight NS variables, a linguistic profile. For each profile, column one contains a reference number and column two the number of informants sharing the profile. The variables are ordered from left to right according to (non-absolute) implicational relationships established by means of pair-wise cross-tabulations comparing the probabilities that one NS-value predicts another NS-value. The profiles are then ordered from top to bottom by sorting the contents of the columns according to highest NS-value, starting from the left-most column.

The young people sharing profiles 1 and (without siblings) 2, are those forming the 30% with ‘All NS values’ in Figure 2 above. That is, they have the values early for AGEARR and AGEON, home for SWFIRST, and Swedish only for LGPAR, LGSIB if not missing, LGFRIEND, LGPREF and LFPROF. Depending on how strictly the monolingualism criterion is applied, however, some or all of these informants’ nativeness could also be questioned. The earlier mentioned context variables, PARBORN, BORN and FRIENDLG, reflect some potential sources for multilingual influence not revealed in the eight NS variables included in the profiles presented in Table 2. Twelve of the 67 ‘all NS’ young people have one or more parents born abroad and one is himself born abroad. More than half of them have close friends with mother tongues other than Swedish. If these context variables were included among the NS variables, the ‘NS group’ would be reduced to 12% of all the young people. Perhaps more importantly, all the young people (including this 12%) are multilingual in various degrees due to other factors than parents: school, of course, but also mass media and friends. When asked about which languages they speak, all 222 informants except one report that they in addition to Swedish speak at least one (often English), or more commonly, two or up to six other languages.

The variable FIRSTLG has here been excluded for reasons discussed above.
LGSIB: SS, SO, OO, and ‘no sibling’.
In Table 2, however, missing values due to no siblings are not replaced.
This response is probably due to a narrower interpretation of ‘speaking a language’.
Table 2. *Multilingual profiles on (only) eight variables. (SS=Swedish only, SO=Swedish and other, OO=Other only.)*

<table>
<thead>
<tr>
<th>Profile no.</th>
<th>N inf.'s</th>
<th>Age of onset</th>
<th>Language use</th>
<th>Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AGE ARR</td>
<td>AGE ONS</td>
<td>SW FIRST</td>
</tr>
<tr>
<td>1</td>
<td>65</td>
<td>early</td>
<td>early</td>
<td>home</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>early</td>
<td>early</td>
<td>home</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>early</td>
<td>early</td>
<td>home</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>early</td>
<td>early</td>
<td>home</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>early</td>
<td>early</td>
<td>home</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>early</td>
<td>early</td>
<td>home</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>early</td>
<td>early</td>
<td>home</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>early</td>
<td>early</td>
<td>other</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>early</td>
<td>early</td>
<td>other</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>early</td>
<td>early</td>
<td>other</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>early</td>
<td>early</td>
<td>other</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>early</td>
<td>early</td>
<td>other</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>early</td>
<td>late</td>
<td>home</td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td>early</td>
<td>late</td>
<td>other</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>early</td>
<td>late</td>
<td>other</td>
</tr>
<tr>
<td>16</td>
<td>5</td>
<td>early</td>
<td>late</td>
<td>other</td>
</tr>
<tr>
<td>17</td>
<td>3</td>
<td>early</td>
<td>late</td>
<td>other</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>late</td>
<td>late</td>
<td>other</td>
</tr>
<tr>
<td>19</td>
<td>5</td>
<td>late</td>
<td>late</td>
<td>other</td>
</tr>
<tr>
<td>20</td>
<td>8</td>
<td>late</td>
<td>late</td>
<td>other</td>
</tr>
<tr>
<td>21</td>
<td>5</td>
<td>late</td>
<td>late</td>
<td>other</td>
</tr>
<tr>
<td>22</td>
<td>3</td>
<td>late</td>
<td>late</td>
<td>other</td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>late</td>
<td>late</td>
<td>other</td>
</tr>
<tr>
<td>24</td>
<td>4</td>
<td>late</td>
<td>late</td>
<td>other</td>
</tr>
<tr>
<td>25</td>
<td>4</td>
<td>late</td>
<td>late</td>
<td>other</td>
</tr>
<tr>
<td>26</td>
<td>1</td>
<td>late</td>
<td>late</td>
<td>other</td>
</tr>
</tbody>
</table>

Subtotal: 152

27-88 70 (70 informants distributed over 62 different profiles not represented in the table)

Total: 88 222

The informants who do not have NS values for all eight variables are distributed over 86 different profiles. For a few young people, ‘non-heritage’ multilingualism turns up in the use of both Swedish and another language with friends (profiles 3-4) or as a preference for speaking another language (profile 5). Several young people report that they learn languages through friends, resulting in degrees of competence ranging from merely a familiarity (understanding words in many languages or when friends speak a particular language, picking up words and everyday phrases) to a more active command. Similar phenomena has also been observed in, e.g., Switzerland (Franceschini 1999) and Germany (Auer & Dirim 2003, Dirim & Auer 2004). An example in our data is Emmy\(^{12}\) (profile 3), who,\[^{12}\] All names are pseudonyms.
when asked which languages her friends have as mother tongues, reports that they can be ‘everything you can think of’. Then she specifies that most have Arabic or Farsi and describes her language use with them like this:

Emmy: “I mostly speak Swedish with them, but, you can say a little of this and that in Farsi and like that”.

She also claims later in the interview that she writes SMS in different languages (including Arabic) and can understand some of what her boyfriend and his mother say when they speak Farsi with each other. In another study which is part of this project (Bodén & Grosse forthcoming), Emmy’s speech in another speech situation is judged by young local listeners to be a good example of ‘Gårdstenska’, a label which is associated with language use in multilingual areas of Göteborg. Although she has ‘NS values’ on all other variables, language use with friends seems to have made an impact on how she speaks Swedish.

A majority of all informants are more or less multilingual in Swedish and one or more (other) heritage language(s). Many of them acquired Swedish in early childhood, report it to be their best and preferred language (e.g. profiles 6-10 and 12) and often also consider it their first language, either alone or in addition to another language. An important group of early learners (N=30) have started to acquire Swedish outside the family, usually at nursery school (in Table 2 represented by 14 informants in profiles 8-12). One example is Semra (profile 10), who is born in Sweden and started speaking Swedish at the age of one at nursery school. She regards Swedish as one of her first languages, along with Turkish. She uses Swedish and Turkish with parents, one older sibling and friends, but has a preference for and thinks she is best at Swedish. Conversely, there are late learners of Swedish who started learning Swedish at home; Maria (profile 13), who is also born in Sweden, started to speak Swedish by the age of three and a half, presumably with her parents and elder siblings.

As regards heritage languages other than Swedish, different patterns of language shift and maintenance are represented among the young people, from the radical shift of Julli (profile 18) to various degrees of maintenance. Heritage language(s) may be retained with parents only (e.g. profile 6), or in the entire family domain (e.g. profile 9), or also with friends – as in the case of Xavier (profile 11), from whom we know that it is Spanish, one of his first languages, that he sometimes uses with friends. He also says he is equally good at both Swedish and Spanish, and has no preference for one language over the other.

As mentioned above, most of the informants regard Swedish as their best language. The exceptions are naturally found among very late learners, such as Data (profile 26), who arrived and started to learn Swedish at school at 18 years of age. But there are also a few informants (not represented in the table) with an early age of onset of Swedish who prefer and/or think they are best at another language. And conversely, there are several with an age of onset up to ten who now favor Swedish.
4. Concluding discussion

Our presentation of the linguistic profiles which emerge from an analysis of data on our 222 informants shows, we believe, that binary distinctions such as native/non-native are not very useful in studies of language and language use in multilingual environments such as the ones in our study. When binary distinctions are used by themselves, different nativeness criteria lead to different categorizations of an informant group, while, within categories, important variation remains hidden. On the one hand, we have suggested that heterogeneity within the group of ‘good candidates’ for native speakership in Swedish can have significant consequences for studies of language variation in multilingual contexts such as ours. On the other, we have noted that there is enormous diversity among the young people in our informant group who do not qualify as ‘good candidates’ for native speakership in Swedish. In fact, significant numbers within our informant group don’t seem to qualify to be native speakers of any language. Considering them to be non-native speakers of all their languages seems like a very unsatisfactory solution, both theoretically and methodologically. Finally, we have shown that there is no absolute implicational relationship among ‘nativeness’ criteria, which would make the use of one or a small number of variables adequate to divide our informants into a small number of categories. We would now like to draw some methodological and theoretical conclusions of our analysis.

When working with a fairly large and diverse informant group in a multilingual setting, such as the one in our study, we suggest that the only defensible way of proceeding is to treat the informant group initially as a single group, defined by whatever geographical, age or other criteria that have been used to delineate the object of study. Rather than dividing the group into two or three a priori categories based on one or two background variables, we have found it rewarding to learn more about our informant group by carrying out a linguistic profile analysis using combinations of a larger number of variables.13

Another important aspect of working with a multitude of linguistic profiles instead of a few gross categories, is the way it inspires and facilitates going back into the individual data, taking into account further variables or the same variables at a higher level of detail, or a return to the original recordings, for example to try to find explanations when apparent anomalies arise in the profile analysis. The risk if one doesn’t do this is that assumptions about the relationships between background variables (e.g. age of onset and age of arrival) never actually get tested. It also allows the researcher to see the relevance of variables (e.g. in our study: mother tongues of good friends) that normally are not used as a basis for categorizing informants in terms of nativeness. In this way, less biased research can be carried out at both a macro and a micro level alternately. By moving between macro and micro, we both reduce the risk that interesting relationships and explanations are ‘swept under the rug’ in order to manage the data, and at the

---

13 As noted in footnote 4, our informant group included all students of eight selected high-school classes who were willing to participate.
same time allow ourselves to make generalizations above the level of individual cases.

In this paper, linguistic profiling has been used to explore the usefulness of the binary NS/NNS distinction and various native speaker criteria in the light of data from 222 young people in multilingual urban settings in Sweden. In the near future, we plan to apply profiling to analyses of linguistic variation in the same informant group.

In conclusion, we do not exclude the possibility that a binary NS/NNS distinction may be usefully operationalized, for example, in experimental studies focusing on one particular factor in language acquisition, where subjects are chosen according to this factor and other factors carefully controlled for. But we do not believe that this distinction is useful in studies of language use, variation and change in contemporary multilingual contexts for at least two reasons. One is that no single or small number of nativeness criteria can categorize informants into (in relevant respects) homogeneous groups. The other is that there is no implicational relationship between the criteria, which would allow one criterion (e.g. age of onset) to stand for many. We find it likely that the diversity of linguistic profiles illustrated in this paper is not unique to the 222 young people in our project, but is also to be found in other similar settings. The application of a binary NS/NNS distinction in such contexts will either – if a single criterion is used – result in a categorization of informants into two widely heterogeneous groups, or – if multiple criteria are combined and only clear cases considered – result in an exclusion of a considerable number, possibly a majority, of language users from the object of study.

References


KERSWILL, P. & E. TORGERSEN (this volume). “Ethnicity as a source of changes in the London vowel system”.
