THE BUDAPEST SOCIOLINGUISTIC INTERVIEW: VERSION 3

MIKLÓS KONTRA AND TAMÁS VÁRADI

THE BUDAPEST SOCIOLINGUISTIC INTERVIEW: VERSION 3

Compiled by
MIKLÓS KONTRA
(with the assistance of a dozen colleagues
at the Linguistics Institute,
Hungarian Academy of Sciences)

July 1988

Revised and prepared
for electronic publication by
TAMÁS VÁRADI

December 1997

Contents

Contents

Preface

Housed at the Linguistics Institute of the Hungarian Academy of Sciences, *The Survey of Spoken Hungarian* (alias *The Budapest Sociolinguistic Interview*) aims to investigate the social and stylistic stratification of the Hungarian spoken in the capital city of Budapest. In 1987, 50 pilot interviews were conducted with a quota sample of ten teachers over 50 years of age, ten university students, ten blue-collar workers, ten sales clerks, and ten vocational trainees aged 15-16.

In 1988-1989, 200 tape-recorded interviews were completed in Budapest. This latter phase of our study is called The Budapest Sociolinguistic Interview (BSI), Version 3. The 200 informants interviewed form the Budapest subsample of the 1000-strong national sample used for the pen-and-paper Hungarian National Sociolinguistic Survey (cf. Kontra 1995). The BSI project is heavily indebted to Labov (1984) and the Survey of Vancouver English (Gregg 1984). In the transcription and coding of the recordings we followed the Survey of English Usage to some extent.

Over the past decade the BSI project has undergone several changes both in the composition of the research team and in the methodology used in transcribing, coding and checking the data. We have increasingly felt the need to record and make available some of the fundamental principles and research tools of our project and for this reason we are publishing *The Budapest Sociolinguistic Interview, Version 3, From Cards to Computer Files: Processing the Data of the Budapest Sociolinguistic Interview* and *Manual of the Budapest Sociolinguistic Interview Data* as Nos 2, 3 and 4, respectively, of the Working Papers in Hungarian Sociolinguistics.

The original version of what is published here was discussed and critiqued by a small international gathering in Budapest in 1988. The current revised version contains the English translation of the original Hungarian protocol used in the fieldwork in 1988-89. It also lists references to relevant papers which have since been published.

December 12, 1997   *Miklós Kontra & Tamás Váradi*

References

Gregg, R. J. 1984. *Final Report to the Social Sciences and Humanities Research Council of Canada on An Urban Dialect Survey of the English Spoken in Vancouver.* Vancouver: Linguistics Department, University of British Columbia.

Kontra, Miklós. 1995. On current research into spoken Hungarian. *International Journal of the Sociology of Language* # 111:5-20.

Labov, William. 1984. Field methods of the Project on Linguistic Change and Variation. In: John Baugh and Joel Sherzer, eds., *Language in Use: Readings in Sociolinguistics,* 28-53. Englewood Cliffs, N.J.: Prentice-Hall.

1 Introduction

1.1 The aims of the Survey of Spoken Hungarian (SSH)

- to analyse the linguistic variations to be found in the speech of speakers in Budapest belonging to various socio-economic groups with a precise sociological profile

- to examine various language styles, that is, variations subject to the amount of speakers' audio-monitoring of their speech

- to accumulate data that would allow for the analysis of linguistic change if similar data are collected in 10-20-30 years' time.

1.2 The role of the sociolinguistic interview in the Survey

To achieve the objectives outlined above, various research instruments are needed, including:

- sociolinguistic interviews

- recording group sessions (e.g. card playing) by
  a) audio tape recorder
  b) video tape recorder

- use of speech material originally recorded for a different purpose such as the tapes provided by Főtaxi (the Municipal Taxi Company) containing the most relaxed style of speech possible, which were not recorded for our specific purposes, ie we are secondary users of this corpus

- the analysis of the language system of speech communities through participant observation

- experimental analysis of language use

It appears then that the sociolinguistic interview is only one of the research tools – albeit a very important one – employed in the Survey. If the interview does not cover a

particular linguistic phenomenon, it should not be concluded that the item in question is not considered important. Rather, it was thought that the interview is not an adequate means to investigate the phenomenon in question. For example, elicitation is the classical means to establish the degree of grammaticality of certain sentences. This means that a corpus of continuous spontaneous speech can only yield complementary evidence.

1.3   Is completeness our objective?

If our aim were to provide a comprehensive description of language use in Budapest, then our data collection would have to be representative in a) sociological and b) linguistic terms. We know from Hungarian sociologists that sociological representativeness would call for the use of 200-300 informants. What we don't know, however, is how much material should be recorded and in what communicative situations from a single informant in order to enable us to make inferences for the entirety of their language system with only a minimal margin of error.

It follows from the above that the Budapest survey can only aim at sociological representativeness at best.

We cannot strive for a comprehensive analysis of individual language features such as for example, the so-called "-*ik*" verbal conjugation. (Hungarian verbs fall into two categories: -*ik* and non-*ik* verbs, e.g. *alsz-ik* 'he sleeps' vs. *áll-0* 'he stands'. The two types of verbs used to be conjugated quite differently but today the differences are disappearing rapidly. The -*ik* or non-*ik* issue is a stylistic one today but speakers are sensitive to it, and this or that usage often invites a good deal of comments.) The Standard Prescriptive Grammar (=*Nyelvművelő Kézikönyv* 1980-1985; henceforth abbreviated as SPG) sorts these verbs into six groups "from a descriptive point of view", on the basis of their use. However, one finds variation not only in verbs belonging to the so-called "fluctuating -*ik*" and "fluctuating pseudo -*ik*" category but in those belonging to the "stable -*ik*" category as well. It is easy to realize that a comprehensive analysis of the issue would call for hundreds and thousands of relevant examples from a single informant in a great number of communicative situations ranging from careful formal style to the most casual relaxed speech situations. It is clearly beyond the means of our project to undertake such a detailed analysis, however justified and desirable it may be to compile such a close-up view of the use of these verbs. What the project can undertake is to provide solid empirical evidence to answer certain selected questions, including, for example, about the -*ik* conjugation. Our investigations will not yield a definitive answer to the question whether the classification in SPG is descriptively sound, whether the particular verbs are correctly assigned to their category in SPG etc. However, we will be able to answer the following questions:

1. How many speakers in the different socio-economic groups use the form *alszom* 'I sleep' and how many use *alszok* in an interview situation.

2. We will provide quantitative evidence to confirm or reject the statement that "The *-ik* conjugation prevails in imperative mood more than in the conditional" (SPG I.:1012)

3. Sociologically valid quantitative evidence will be gained on the use of certain hypercorrect forms (e.g. *"Én naponta kétszer* edzem*"* standard Hungarian has *edz-ek*, because *edz-0* is a non-*ik* verb; *esz-ek* vs. *esz-em* 'I eat sg.' is often commented on by speakers – *eszem* is supposed to be good usage – this may be the reason for the hypercorrect form *edzem*.)

4. How uncertain are informants with respect to the *-ik* conjugation?

5. To what extent are the particular groups of informants aware of style, i.e. can they use in a test situation the formal *-ik* forms called for by formal style, or is the *-ik* conjugation no longer a stylistic device that they actively use?

1.4   On the selection of research questions

Certain research questions of the interview will be set out below. The selection of these questions was made in the following way:

- In summer 1986 our fellow researchers at the Linguistics Institute of the Hungarian Academy were asked to state what they considered important questions for our project to focus on

- 22 answers came in from about 70 researchers and the answers were all processed.

- in a series of project meetings in 1986, our group discussed work in related projects abroad

- relevant conclusions in the literature were discussed and adopted in our work

- the following studies were commissioned:

  > Cseresnyési, László: *Hangtani kérdések. Ajánlások a budapesti köznyelvi vizsgálatok adatfelvételéhez.* (Phonological questions. Recommendations on data collection for the Survey of Spoken Hungarian.) Manuscript, 1986.

  > Komlósy, András: *Mondattani kérdések. Ajánlások a budapesti köznyelvi vizsgálatok adatfelvételéhez.* (Syntactic questions. Recommendations on data collection for the Survey of Spoken Hungarian.) Manuscript, 1987.

- On the basis of the above information Miklós Kontra defined the list of research questions and

- the list so derived was pruned to include only the phenomena amenable to a sociolinguistic interview.

With some simplification it can be claimed that out of the research questions proposed by colleagues the ones that made it to the interview were those that the project leader considered important and which can be most appropriately examined by means of an interview.

1.5 The rest of this manual

contains the interview itself in the following arrangement: the text of the three sections of the interview (phonology; morphology, syntax and lexicon; guided conversation) will be followed by a discussion of the research tools of the relevant sections together with the instructions to the field workers. Problems of transcription, coding and analysis are dealt with in the final chapter.

1.6 On the selection of informants

In Autumn 1987 Version Two of the Budapest Sociolinguistic Interview was used with a quota sample: 10 teachers of over 50 years of age, 10 university students, 10 sales clerks, 10 blue-collar workers and 10 vocational trainees (around age 16) were interviewed.

Version Three in 1988 was used with a stratified random sample, the (200 person) Budapest subsample of a national representative sample comprising 1000 persons. This sample has been used by Róbert Angelusz and Róbert Tardos to record more than a thousand sociological questions for their project "Social stratification – communicative stratification". In May 1988 the 1000 person sample was administered a linguistic questionnaire as a complement to the Budapest Sociolinguistic Interview (see Kontra 1995:10-11).

Therefore, it can be claimed that the sample used in 1988 meets any sociological standard. As an extreme example, should there be any correlation between a person's use of a linguistic variable and the date when their home was last remodelled, we would be able to precisely state that correlation.

Basic sociological data recorded about the informants include: age, sex, occupation, education, race (Gipsy or not), birthplace, in-migrant or not, teacher or not, mother tongue, time abroad, knowledge of foreign languages.

1.7 Field workers

The Survey hired field workers who are not trained linguists. The field workers are experienced interviewers in sociological research but had to be trained by us to do this

linguistic interview. This is why some of the things we say here may seem to the professional linguist too obvious to mention.

1.8   What to do about illiterate informants

The interview must be started with every informant and should be taken up to 'break'-point. If the informant turns out to be illiterate, then

1. all reading tasks (cards, passages, etc.) must be omitted

2. every other task (listening tasks, conversational modules, reporter's test) must be carried out,

3. the two-hour tape should be filled with guided conversations, and

4. a detailed explanation must be included in the log book as to why the interview is incomplete. In such cases, the word 'incomplete' should be written on the outside cover of the log book next to the informant's identifier.

If an informant should be unable even to read the letters A and K (for 'same' and 'different') or the figures '1' and '2', then the test must be done orally. In other words, the informant must be trained to give oral answers, the trial tests should be administered and when the test seems to be safely administrable, it should be started. The informant's microphone should record the stimulus on the test tape together with the informant's answers. The final result will be something like this:

> test tape: "one: *gyanu – gyanú* ('suspicion')"
> informant: "they are the same"

1.9   Primary and secondary data

In doing the oral sentence completion tests, the informant's attention mostly focusses on the word to be inserted and the suffix to be used. It can be presumed that when reading the full sentence out the informant pays less attention to the rest of the sentence. Therefore, we will call the item that the card is explicitly focussed on 'primary data', the rest secondary. For example,

> Ebben a ..... nem mehetsz színházba.   FARMER 'jeans'
> 'In-this the ..... not you-may-go to-theatre.'

The expected answer is *farmer-ban/ben*. Here the primary data are the vowel of the suffix and the presence or omission of the word-final nasal. On the other hand, presence or absence of the word-final nasal in *Ebben* as well as the short or long '*i*' of *színházba* are considered secondary data.

The field worker should have a clear view of the location of secondary data as well although they are not explicitely marked on his/her cards.

In processing the results, primary and secondary data should be handled strictly separately. After the transcription and coding of the data is completed, a separate study must be made to establish whether the distinction introduced in 1988 as a working hypothesis was indeed justified, in other words, whether informants' responses are significantly different as a result of the amount of attention devoted to the linguistic task. (For a preliminary answer to this question see Váradi 1995/1996.)

1.10   References

Kontra, Miklós. 1995. On current research into spoken Hungarian.*International Journal of the Sociology of Language* 111:9-20.

Váradi, Tamás. 1995/1996. Stylistic variation and the (bVn) variable in the Budapest Sociolinguistic Interview. *Acta Linguistica Hungarica* 43:295-309.

2 Phonological section

THE AIM OF OUR INQUIRY is to establish the distribution of certain optional phonological and/or phonetic rules in terms of

1) social position
2) speech tempo
3) style.

In other words: what differences can be observed between varieties dependent on speech tempo (cf. Elekfi 1973, Kerek 1977 and Dressler-Wodak 1982) and the varieties governed by the amount of audio-monitoring (cf. Labov 1966, 1972 and 1984)? Also, we are interested to find out the degree of homogeneity/heterogeneity of varieties along the slow-fast and casual-formal dimensions within a given socio-economic group. It is also unknown how speakers located at the same spot on the speed or style axes but of different social background differ in their speech performance i.e. what difference there is between the fast speech of uneducated speakers vs. college graduates or between the formal style of university graduates and those with only elementary school education.

2.1 Research tools

The varieties governed by speech tempo can be examined by asking informants to read out the same passage first slowly then at a fast rate.

The various speech styles ranging from formal to casual speech will be recorded in the way pioneered by Labov:

1. Minimal pairs will be read out (e.g. *sor – sör* 'row' – 'beer'),

2. word-lists will be read aloud,

3. short passages will be read out,

4. words will be elicited from informants in the course of conversation (in the same way that field workers on Language Atlas projects did),

5. all relevant data in the informant's speech from the more informal parts of the interview will be considered.

2.2 Research topics in phonology

The orthographic and phonetic symbols of Hungarian are reproduced in Table 2.1 on p. 15 and Table 2.2 on p. 16 from Vago (1980).

1. Affrication. If a dental or palatal stop is followed by a strident fricative, then geminate affricates result. The assimilating affrication between $t + sz \rightarrow cc$ (e.g. *látszik*) in suffixed forms often does not take place, with *ad+sz* yielding *atsz* instead of the expected *acc* (Deme 1962:102). Some authors, Szántó (1962:164) for example, think that this affrication does not take place across word boundaries. Elekfi (1973:20) holds that "in compound words *ts* and *tsz* become *cs* and *c* respectively only in fast, hasty speech but this assimilation does not work across word boundaries." Vago (1980:37) claims "there is no affrication in *két szék* 'two chairs'." Possible cases of affrication will be examined in three positions: word-internally (e.g. *hatszoros*), across word boundary (e.g. *hat szoros*), and with contrastive stress. Two types of contrastive stress will be analyzed:

   a) *ez nem hat völgy hanem hat szoros* 'these are not six valleys but six canyons', and b) *nem öt szoros csavar, hanem hat szoros csavar.* 'not five tight screws but six tight screws'

   On the role of contrastive stress in voice assimilation cf. Deme 1962:100.

2. The strident consonants *sz,z*, *s*, *zs*, *c*, *dz*, *cs*, and *dzs* show place assimilation with another strident consonant following them (cf. Vago 1980:38). Accordingly, *kis szoba* 'small room' → *KISZ szoba*, *rácsszerű* 'grid-like' → *rácszerű* etc. According to Elekfi (1973:15) place assimilation is a regional feature, in careful common style the constituent segments are pronounced one by one as written e.g. *malacság* [c-s]. The assimilated variety "is less careful, rather belongs to casual, everyday speech". We hypothesize that this rule operates as a function of speech tempo and style, in other words assimilation in a minimal pair like *kis szoba* 'small room' vs. *KISZ szoba* 'Communist Youth League office' will not take place in the most formal setting (reading of minimal pairs) but will do so in casual speech.

3. Palatal assimilation. We will examine to what extent the assimilation exemplified by *látja* 'he sees it' → *láttya* depends on speech tempo and speech style. (When the glide *j* follows *t*, *d*, *n*, or *ty*, *gy*, *ny*, in Hungarian, the result is a geminate palatal consonant – cf. Vago 1980:39.)

4. How do speakers' efforts to distinguish meaning manifest themselves? When is a pair like *bontsd fel* homophonous with *bonts fel* (as against the pair *rántsd le* – *ránts le*, cf. Kerek 1977:118)? When is the word *lombtalanít* 'defoliate' distinguished from *lomtalanít* 'clear sg. of junk'? Elekfi (1973:62) holds that "when in danger of

ambiguity ... we tend to pronounce the medial stop as well, because for example *bontsd fel [boncstfel]* is different from *bonts fel [boncsfel]*"

    1) *bontsd fel* 'open it up' → bončfel
    2) *bonts fel* 'open up sg.' → bončfel
BUT:
    3) *rántsd le* 'pull it down' → ra:nǰle
    4) *ránts le* 'pull down sg.' → ra:nčle

The dental stop is dropped in (1), resulting in homophony with (2). *l* does not induce voice assimilation in Hungarian. In (3) *d* causes the preceding affricate to become voiced before being dropped, whereas in (4) *d* is absent thus no voice assimilation takes place and hence (3) and (4) are different.

5. The elision of *l*. According to Deme (1962:105) this is a "strongly dialect" feature: e.g. *tanútam* instead of the common style rendering *tanultam* which is identical with the typographic image of the word. Elekfi (1973:5) regards this feature characteristic of "almost every dialect and the whole of less careful common style as well."

6. The elision of *t* as in *jelentkezik* → *jelenkezik*. According to Elekfi (1973:60) the *t*-less variant "cannot be accepted in common style". Also to be examined is the pair *bólintgat* vs.*bólingat*, which is listed in the orthographical dictionary in its *t*-less form but according to G. Varga (1968:151) the *t*-less variety is more frequent in the speech of less educated speakers whereas university graduates tend to prefer forms with the *t*. (*bólint* means 'to nod', *bólintgat* 'to keep nodding'. The form with *t* is more transparent because *-int* is an instantaneous verbal suffix.)

7. About two-thirds of Hungarian speakers have two *e* phonemes: a front low *e* (traditionally called open *e*) and a front mid *e* (traditionally called close *e*; henceforth written as *ë*). About one-third of Hungarians, including the standard Budapest Hungarian speakers, only have one (open) *e*. — Budapest is a melting pot with lots of in-migrants. In-migrating two-*e* speakers must be diagnosed in order for us to say something about the process of two-*e* speakers becoming one-*e* speakers in Budapest. According to G. Varga (1968:32), a traditional dialect study of 200 Budapest speakers, "the standard Budapest Hungarian open *e* sounds predominate in place of etymologically close *e*, with the close *e* being present in negligible number as a non-phonemic variant; not a single informant used it correctly and consistently." The question is what is the social distribution of close *ë* in Budapest. Györgyi G. Varga had access to "relatively limited data of continuous speech" (op.cit. 29). In the tests we are going to focus on a few words only in this regard, but this will be complemented by a massive amount of continuous speech.

8. How does typewritten text influence the speakers' reading performance? (N.B. Until about the early 1980s the keyboards of Hungarian typewriters lacked three keys: í, ú, ű. Instead of these high long vowels only their short high equivalents (i, u, ü) could be typed. It has been claimed several times that this deficiency of the keyboard has an influence on people's speech i.e. makes them use short vowels instead of standard long vowels, thus accelerating the change (?) or tendency to shorten these vowels. Szántó (1962:454) explains assimilation that takes place in spontaneous speech but not in reading aloud by the "spell" of the written text. We will examine whether informants will read out words differently if they are spelt on typewriters with the old and the new Hungarian standard keyboard (e.g. *hosszu* – *hosszú* 'long') or if they are spelt according to the 10th or the 11th edition of the Orthographical Rules of the Hungarian Academy (e.g. *zsüri* – *zsűri*). A further question is how the evidence obtained relates to data from spontaneous speech. For a VARBRUL analysis of this problem see Pintzuk et al 1995.

9. Morphology: *-ba* vs. *-ban*. *-Ba/-be* vs. *-ban/-ben* constitute an important grammatical difference in written Hungarian e.g. *ház-ban* 'in the house' vs. *ház-ba* 'into the house'. This distinction is often not observed in speech, *-ba* being used instead of *-ban*. To what extent is the realization of *-ba(n)* influenced by the "erosion and subsequent elimination of the sense of direction" (G. Varga 1987)? Can we corroborate the four types posited by G. Varga, namely, 1) concrete location, neutral context (e.g. *ülök a szobában* 'I'm sitting in the room'), 2) concrete location, non-neutral context (e.g. *benn ülök a szobában* 'I'm sitting inside the room'; the adverb *benn* 'inside' is related to the suffix *-ban/-ben* 'in'), 3) more abstract adverbial function (e.g. *gyermekkorában, nyomorban, kettesben* 'in his childhood, in poverty, in pairs'), and 4) governed complements (e.g. *gyönyörködik, bízik, csalódik valami*ben 'to take delight in, to have trust in, to be disappointed in'). What is the distribution of *-ba* forms in *-ban* function in terms of speech tempo, speech style and the socio-economic status of speakers?

10. Morphonology: Context effects in vowel harmony. Certain Hungarian (loan) words (e.g. *farmer* 'jeans' and *férfi* 'man' have vacillating suffixes, e.g. *farmer-ben* or *farmer-ban* 'in jeans'. Kontra and Ringen (1986) claim that such vacillation is not free but influenced by context in such a way that 1) if the word immediately before the test word with the vacillating suffix has a suffix morpheme identical with the vacillating suffix morpheme, then the vowel quality (front or back) of the preceding suffix may influence the choice of suffix vowel in the test word. For instance, subjects are more likely to use the front suffix *ben* in a sentence like

> *Eb-ben a farmer-* ... *nem mehetsz színházba.*
> 'this-in the jeans- not you-may-go to-theatre

than in

> *Ab-ban a farmer- ...   nem mehetsz színházba.*
> 'that-in the jeans- not you-may-go to-theatre

Written elicitation tests have shown this effect of context in vowel harmony (Kontra et al. 1990). The Survey will gather additional spoken data to further test the hypothesis.

11. What is the social distribution of certain stigmatized pronunciation variants e.g. *inekció, szofiáné*? What pronunciation variants do old loans like *nylon – nejlon* and some recent ones e.g. *spray – szpré, dzsúz – dzsúsz* have?

12. What is the pronunciation of words which show variation in vowel length in speech but are consistently spelt with a long vowel such as *színház, útiköltség, háború, fésű* and *hűvös* (cf. G. Varga 1968) as well as *bölcsőde* (cf. SPG I.:421)?

2.3   References

Deme, László. 1962. Hangtan (=Phonetics and phonology). In: Tompa, József (ed.) *A mai magyar nyelv rendszere,* (The system of present-day Hungarian, I:57-119.) Budapest: Akadémiai Kiadó.

Dressler, Wolfgang U. & Ruth Wodak. 1982. Sociophonological methods in the study of sociolinguistic variation in Viennese German. *Language in Society* 11:339-70.

Elekfi, László. 1973. *A magyar hangkapcsolódások fonetikai és fonológiai szabályai.* Manuscript. Linguistics Institute of the Hungarian Academy of Sciences.

Kerek, Andrew. 1977. Consonant elision in Hungarian casual speech. *Studies in Finno-Ugric Linguistics in Honor of Alo Raun,* ed. by Denis Sinor. IUUAS Volume 131:115-130. Bloomington.

Kontra, Miklós & Catherine Ringen. 1986. Hungarian Vowel Harmony: The Evidence from Loanwords. *Ural-Altaische Jahrbücher* 58:1-14.

Kontra, Miklós; Catherine O. Ringen; Joseph P. Stemberger, 1990. The Effect of Context on Suffix Vowel Choice in Hungarian Vowel Harmony. In: *Proceedings of the Fourteenth International Congress of Linguists,* Berlin, August 10 – August 15, 1987. pp. 450–453. Akademie-Verlag Berlin.

Labov, William. 1966. *The Social Stratification of English in New York City.* Washington, D.C.: Center for Applied Linguistics.

— 1972. *Sociolinguistic Patterns.* Philadelphia: University of Pennsylvania Press.

— 1984. Field Methods of the Project on Linguistic Change and Variation. In Baugh, John and Joel Sherzer (eds.) *Language in Use: Readings in Sociolinguistics,* 28-53. Englewood Cliffs, N.J.: Prentice-Hall.

Pintzuk, Susan; Miklós Kontra; Klára Sándor; Anna Borbély, 1995. *The effect of the typewriter on Hungarian reading style.* (Working Papers in Hungarian Sociolinguistics No 1, September 1995). Budapest: Linguistics Institute of the Hungarian Academy of Sciences. 30 pp.

Szántó, Éva. 1962. A magyar mássalhangzó-hasonulás vizsgálata fonológiai aspektusban. *Magyar Nyelv* 58:159-166 & 449-458.

Vago, Robert M. 1980. *The Sound Pattern of Hungarian.* Washington, D.C.: Georgetown University Press.

G. Varga, Györgyi. 1968. *Alakváltozatok a budapesti köznyelvben.* (Form variants in the everyday language of Budapest). Budapest: Akadémiai Kiadó.

G. Varga, Györgyi. 1987. Vélemény "A budapesti szociolingvisztikai interjú" című tervezetről. (On the Project Outline of "The Budapest Sociolinguistic Interview"). Manuscript.

Orthographic	Phonetic	Description
a	a	short low, slightly rounded back vowel ([ɐ])
á	a:	long low unrounded back vowel
b	b	voiced bilabial stop
c	t͡s	voiceless dental affricate
cs	č	voiceless alveo-palatal affricate
d	d	voiced dental stop
dz	d͡z	voiced dental affricate
dzs	ǰ	voiced alveo-palatal affricate
e	e	short low unrounded front vowel (=[ɛ])
é	e:	long mid unrounded front vowel
f	f	voiceless labio-dental fricative
g	g	voiced velar stop
gy	d͡y	voiced palatal stop
h	h	glottal glide
i	i	short high unrounded front vowel
í	i:	long high unrounded front vowel
j	j	palatal glide
k	k	voiceless velar stop
l	l	dental lateral
ly	j	palatal glide
m	m	bilabial nasal
n	n	dental nasal
ny	ñy	palatal nasal
o	o	short mid rounded back vowel
ó	o:	long mid rounded back vowel
ö	ö	short mid rounded front vowel
ő	ö:	long mid rounded front vowel
p	p	voiceless bilabial stop
r	r	dental trill
s	š	voiceless alveo-palatal fricative
sz	s	voiceless dental fricative
t	t	voiceless dental stop
ty	t͡y	voiceless palatal stop
u	u	short high rounded back vowel
ú	u:	long high rounded back vowel
ü	ü	short high rounded front vowel
ű	ü:	long high rounded front vowel
v	v	voiced labio-dental fricative
z	z	voiced dental fricative
zs	ž	voiced alveo-palatal fricative

Table 2.1: The orthographic and phonetic symbols of Hungarian (taken from Vago 1980 pp. 1-2)

15

Short

	Front		Back
	Unrounded	Rounded	Rounded
High	i	ü	u
Mid		ö	o
Low	e		a

Long

	Front		Back	
	Unrounded	Rounded	Unrounded	Rounded
High	í	ű		ú
Mid	é	ő		ó
Low			á	

Table 2.2: The systematic phonetic vowels of Hungarian (taken from Vago 1980 p. 2)

3 Morphological, syntactic and lexical section

THE AIM OF OUR INQUIRY is to explore some linguistic variables listed below as well as to elicit continuous spontaneous speech from every informant, which would also allow for the investigation of several questions that are not included in the present discussion. The first aim is to be achieved through various tests while the latter through guided conversation.

A LINGUISTIC VARIABLE is a linguistic element that has variants. The variants can be related to the speech style and the social position (socio-economic status) of the speakers. The variables are amenable to quantitative description and probably play a key role in language change. Language variables can be described in rules – such rules define the socio-regional conditions under which the variants appear. To take a simple example, it is not known today how the use of *-ba* forms in *-ban* functions relates to various speech styles, whether educated speakers use this variable in a different way than uneducated ones (irrespective of style and/or as a function of it), nor is it known whether the use of alternants is affected by linguistic context. Intuitively, one could presume, for example that "inconsistencies" like *ebbe a házban* 'into-this the house-in' do not occur, however, there is evidence for the occurrence of such forms (cf. Váradi 1990 and 1994). The two types of data collected in the Survey (roughly: the test data and the guided conversation) complement each other: without the test results we could not make COMPARATIVE analysis across informants, whereas without continuous speech data we could not analyse such characteristics of particular elements in speech as their frequency, contextual dependency etc.

The fact that some research questions are listed explicitely and others are not does not mean the latter are neglected. For example, it is important to find out in what contexts Hungarian sentences with so called flat and eradicating intonation are used (cf. Komlósy 1987). The answer to this question requires the prosodic analysis of a sizeable spoken corpus – but no specific data gathering is needed. Obviously, the spoken corpus makes it possible to investigate several problems not listed here or not even thought of today.

3.1   Research topics in morphology

1. *-ba* forms in *-ban* functions cf. (9) on p. 12

2. Loan words with alternating suffixes cf. (10) on p. 12

3. The social distribution of the so-called *-suk/-sük* conjugation. In Standard Hungarian there is a consistent distinction between the indicative and imperative verbal paradigms of verbs ending in root final *t*, e.g. *lát-ja* 'he sees it' vs. *lás-sa* 'he should see it'. This distinction does not obtain in most Hungarian dialects: speakers use the imperative form in place of the indicative forms, e.g. *lássa* for standard *látja*. This phenomenon, called *-suk/-sük* conjugation, is about as heavily stigmatized as multiple negation in English. The Survey will gather data about its social distribution via the cards, the reporter's test (cf. Ball 1986) and guided conversations. (For relevant findings in the Hungarian National Sociolinguistic Survey see Váradi and Kontra 1995.)

4. The first person singular conditional verbal suffix is *-nék* regardless of the vowels of the stem e.g. *en-nék* 'I would eat' and *alud-nék* 'I would sleep'. Instead of this invariant standard suffix, dialect speakers often use harmonic suffixes, e.g. *en-nék* but *alud-nák*. This dialect feature is heavily stigmatized. The Survey will gather data on its social distribution. (For relevant findings in the Hungarian National Sociolinguistic Survey see Pléh 1995.)

5. The variation

   • according to the social position of speakers and
   • the context awareness of speakers

   in the use of some verbs that can be conjugated according to *-ik* and non-*ik* paradigms. The *-ik* conjugation characterizes primarily "the educated standard" or careful style (cf. SPG I:1011), it is least stable in imperative and conditional mood, therefore this is where the biggest variation can be expected. It is presumed that educated speakers will be more sensitive to context than uneducated informants so our hypothesis is that educated speakers will use more *-ik* forms in formal contexts than uneducated speakers. In other words, educated speakers are more sensitive to context and this is shown in their choice of *-ik* vs. non-*ik* forms. Research tools: cards (*iszom* and *iszok*, *virágozzon* and *virágozzék*) together with a number of relevant parts of the interview as well as the total corpus of continuous speech.

6. *Jöttök* and *jösztök*. According to SPG the latter is the "familiar" variant of the verb-form meaning 'you-pl. come'. We are not certain it is just familiar, it might be regionally or socially tied.

7. *Szabadott.* The adjective *szabad* 'free' and the auxiliary *szabad* 'may' clash in Hungarian. In standard Hungarian 'he was allowed to' is *szabad volt.* In nonstandard it is *szabadott.* According to SPG (II:730) more careful style insists on the "more beautiful and judicious" construction *szabad volt.* The Survey investigates the social distribution of this nonstandard form. Research tool: cards.

8. *Miér* vs. *miért* 'why'. *t* is often deleted in speech. We want to find out the exact conditions under which *t* is deleted.

9. Possessive inflections. The suffix *-é* in Hungarian equals the English *'s* genitive, e.g. *Peter's = Péter-é.* The plurality of the things possessed is denoted by the Hungarian suffix *-i*, e.g. *The children are Peter's = A gyerekek Péter-é-i.* This final plural suffix is often dropped, "ungrammatically", in speech. Can we detect signs of some simplification in present-day Hungarian morphology here?

3.2   Research topics in syntax

1. The interrogative particle *-e.* "In standard speech it is a gross mistake . . . to append it to the preverbal particle, to the nominal part of the predicate or in a compound verb form to the main verb" (SPG I:458). What is the social distribution of this heavily stigmatized syntactic feature? (For relevant findings in the Hungarian National Sociolinguistic Survey see Kassai 1995.)

2. *Természetesen, hogy.* According to current prescriptive grammar (1) and (2) are correct:

    (1) Természet-es-en igazad van.
        'Nature-al-ly you are right'
    (2) Természet-es, hogy igazad van.
        'Nature-al that you are right'

but (3) is incorrect:

    (3) Természet-es-en, hogy igazad van.
        'Nature-al-ly that you are right'

SPG (II:803) holds that such structural blends are considered "not very serious mistakes". However, informal evidence suggests that a syntactic change is going on here. (For relevant findings in the Hungarian National Sociolinguistic Survey see Kontra 1992.)

Research tools: cards and the entire corpus, that is, a concordance of the word *hogy* of all guided conversations will give us all of the instances of such blends as well as all instances of the traditionally correct structures.

3. *ami* vs. *amely*. According to traditional grammar the word *ami* 'what (relative pronoun)' refers to antecedents expressed by non-nouns. The word *amely* 'which (relative)' should refer to antecedents expressed by nouns. Despite prescriptivists' guidance, however, *ami* tends to be used in both cases. *Amely* can also be used hypercorrectly, as in the following example:

> *Van valami ebben a dologban, amely nem világos*
> 'is something in-this thing-in which not clear'

SPG (I:206) also says that *ami* is "increasingly more frequently" used in sentences like

> *Megérkeztek a könyvek, amiket/amelyeket megrendeltünk*
> 'arrived the books what/which we ordered'

Research tools: cards and the entire corpus. A concordance analysis of *ami/amely* in the guided conversations, together with the test results is expected to yield reliable evidence that can make more precise, or indeed understandable at all, the qualification "increasingly more frequent" in SPG.

4. *objects with possessive personal suffix and a verb*. Next to such an object the verb can fluctuate between the definite and the indefinite conjugation. (Hungarian verbs can be conjugated definitely and indefinitely.) There is variation e.g. (a) next to a partitive object

> *Mari kimosott/kimosta egy ingemet,*
> 'Mary washed-indef./washed-def. a my-shirt'

(SPG II:960), and (b) next to the determiner *minden* + an object with possessive suffix, e.g.

> *Pista minden könyvemet elvitt/elvitte.*
> 'Steve all my-books took-indef./took-def.'

The use of definite conjugation verbs in such cases is "more frequent", SPG states (II:961), but it is not known what exactly this increased frequency actually means. Cf. Komlósy 1987:16-17.

Research tools: cards.

3.3   Three lexical issues

- *felolt* and *felgyújt*. The verb *felolt* 'extinguish up' used in the sense of 'turn on /the light/' is a semantic anomaly, but frequently used. *Felgyújt* is the correct standard form.

  Research tool: reporter's test.

- What does the word *demográfia* mean?

  Demográfiát akarunk? Meg kell adóztatni a gyerekteleneket 'Do we want demography? Childless couples should be taxed then.' – a speaker on a live TV program said on 10 December 1986. The speech of more or less uneducated speakers often shows the use of certain fashionable words with transferred meaning, e.g. *Nincs egyetértés a politika és az írók között,* 'There is no agreement between the policy and writers' i.e. between politicians and writers. It may be presumed that the spread of such "inaccuracies" correlates with the educational background and/or socio-economic status of speakers in that the more educated speakers interpret the word in its literal sense whereas less educated speakers allow transferred senses as well.

  Research tool: cards.

- *Tűzőkapocskiszedő*. The staple remover test

  The Hungarian equivalent of *stapler* is *fűzőgép* or *tűzőgép*. The Comprehensive English-Hungarian Dictionary by Országh lists staples as *fűző(gép)-kapocs*, *papírfűző/könyvfűző drótkapocs* or simply *kapocs*. Staple removers, the handy devices that serve to remove the staples easily and without damaging the hand or the paper, were unknown in Hungary in 1986. They made their first appearance in stationery stores in early 1987 selling at around 20-30 forints. In June 1987 shop assistants in the stationery store in Felszabadulás tér, Budapest put down the name of the article on the receipt as *tűzőgép* or *fűzőkapocskiszedő*. This term was used only on the receipt as the goods themselves were sold unpacked, without any brand name and description of the article. In short: we are currently witnessing the spread of a new device at a time when the object practically is without a name. Even those who have been using it for years (because they got it from abroad) are at a loss to name it, as shown by the following two conversations:

  1. - Do you know what this is called? (shows up the object)
     - This is what you've brought everyone from America.
     - What's its name?
     - I don't know. Some [pause] remover. [pause] Are you testing me?
  2. - What's this in my hand?
     - Clipremover.

- But it's not clips this thing handles ...
- What is it called, [pause] gee, what IS it called?

At this point the man in the second conversation took out his own stapler, looked at the package (which was Czechoslovakian, without any Hungarian script) and then said: It's got no Hungarian name, interesting ...

Apart from the official register of goods, we can claim that in 1987 the item in question has no received name, or it may have too many names, because a name is only in the making these days, a name that in a few years' time will uniquely serve to identify what people cannot name at the moment. In short: we have a unique opportunity here to catch the birth of a word *in statu nascendi.*

Recommended research procedure:

1. The field worker shows a staple remover to the informant and asks "Have you ever seen such a thing?"
2. Field worker to informant: "What is this?"
3. If answer is "I don't know what it is", then field worker says: "OK, I'll show you what it's for". Then s/he demonstrates how the remover is used.
4. Field worker: "Now, what is this?"
5. Field worker holds up stapler and asks: "What is this?"
6. Field worker holds up staple and asks: "What is this?"
7. Finally field worker holds up staple remover and asks for a name again.

3.4   Other items collected

After the planning phase of this research, when the resarch topics had been finalized, we realized that a number of variable items which were not listed originally (see 2.2, 3.1, 3.2, 3.3) could also be collected. The following list contains such items, as they have also been coded and arereadily retrievable from the BSI database.

pénze – péndze	egyed – edd
fel- – föl-	odébb – odább
l + j	pettyest – pettyeset
ezben (=ebben)	olvashatók – olvashatóak
se – sem	nála – nálánál
ablaka – ablakja	kinlódjanak – killódjanak
kell mennem – kell menni	javitással – javitásal
szőlője – szőleje	lom(b)talanít – lom(b)tanít

klozet – klozett
ajánlkozik – ajálkozik
kom(m)unisták – kom(m)onisták
mozga:mban
borból – borbul
hűtőben – hüttőben
csöndben – csendben

állította – álította
ezért – ezér – ezé (and other lenition)
elküld-ték
pénzért – pézért
utoljára – utóljára
posta – pósta

3.5 References

Ball, Martin J. 1986. The reporter's test as a sociolinguistic tool. *Language in Society* 15:375-386.

Kassai, Ilona. 1995. Prescription and reality: the case of the interrogative particle in Hungarian. *International Journal of the Sociology of Language* 111:21-30.

Komlósy, András. 1987. Mondattani kérdések. Ajánlások a budapesti köznyelvi vizsgálatok adatfelvételéhez. (Syntactic questions. Recommendations on data collection for the Survey of Spoken Hungarian.) Manuscript, 1987.

Kontra, Miklós. 1992. On an ongoing syntactic merger in Hugarian. In: Kenesei, István & Csaba Pléh (eds.) *Approaches to Hungarian, Volume Four: The Structure of Hungarian*, 227-245. Szeged: JATE.

Pléh Csaba. 1995. On the dynamics of stigmatization and hypercorrection in a normatively oriented language community. *International Journal of the Sociology of Language* 111:31-45.

Váradi, Tamás. 1990. *Ba* vagy *ban*: problémavázlat.(*Ba* 'into' or *ban* 'in': an outline of the problem). In: Szabó G., (ed.) *II. Dialektológiai Szimpozion*. Veszprém: VEAB, pp. 143-155.

— 1994. Hesitations between Inessive and Illative Forms in Hungarian (*-ba* and *-ban*). *Studies in Applied Linguistics* 1:123–140. [Debrecen]

— and Miklós Kontra. 1995. Degrees of Stigmatization: *-t*-final Verbs in Hungarian. In: ZDL-Beiheft 77: *Verhandlungen des Internationalen Dialektologenkongresses Bamberg 1990. Band 4*, 132-142. Wolfgang Viereck (Hrsg.), Franz Steiner Verlag Stuttgart.

4 Guided conversation

Out of the ten objectives of a sociolinguistic interview listed in Labov (1984), the following four are to be realized primarily through guided conversations:

1. To gain comparable answers to questions that enable us to contrast the different attitudes and experiences of particular sub-cultures (e.g. danger of death, fate, premonition, fights and the rules of fair fighting, attitudes towards ethnic minorities, ambitions relating to school and education).

2. To prompt the informant to relate personal experiences which would show up community norms and styles of personal interaction and where speech style tends to be close to the vernacular.

3. To stimulate group sessions and record conversations whereby informants engage in conversation among themselves and not with the field worker.

4. To locate the topics that are closest to the informants and also to give them a chance to raise topics of their own.

4.1   The Observer's paradox and the microphone

In order to gain optimum quality recordings, small sized lavalier microphones should be used, clipped to the informant's garment. This may serve to eliminate microphone fright but it has the drawback that the speech of the field worker may become too low or inaudible. Therefore, when making a test recording the field worker should take up a position that is not disturbingly close to the informant yet their voice should be audible on the tape.

Group sessions can only be recorded with a desktop microphone. Here two strategies should be followed: (1) If the field worker cannot leave the scene of the conversation, s/he should strive to keep a low profile. S/he should speak to the informants from an equal footing, but should withdraw from the conversation whenever possible. (2) Following Löfström (1982) the field worker should try to leave the scene of the conversation. Owing to the higher level of shared knowledge between informants, this ploy will yield conversation that may prove "too intimate" for the field workers, in other words they will

25

be unable to interpret every word, phrase or conversation topic during the transcription. At the same time, the absence of the field worker may reduce the observer effect.

As a general rule the field worker should go through the network of conversational modules but should try to leave the scene either in the middle or towards the end of the conversation. If another member of the family or a neighbour drops in on a *tête-á-tête* conversation, the new person should also be involved in it. It is not desirable, however, that this newcomer should take over the role of the informant. Whenever there is such a danger, the informant should be given back the turn with a question like "And what do *you* think about this?"

An unexpected telephone call in the course of an interview provides an opportunity to record speech outside the framework of the interview in a non-surreptitious way. In such cases the field worker should encourage the informant to answer the phone and whenever there is a chance for a longer conversation he/she should try to leave the room by asking to go to the lavatory (but without stopping the tape recorder).

4.2   On the role of the field worker

It is a point of fundamental principle that the field worker should not act from a position of authority but rather as a helpful inquirer who knows less about the local way of life, customs, problems and language. Information should go from informant to field worker and not vice versa (Labov 1984).

It may easily happen that the field worker may inadvertently raise a question that makes the informant stunned or outraged. In such a case the field worker should skilfully slip into another conversational topic. It must be made clear to the informants right at the beginning of the interview that whenever they are asked a question that they do not wish to answer, they should feel free to do so by simply indicating clearly to the field worker that they do not want to give an answer to the particular question. For example: if early on in the interview, perhaps in the demographic module, it turns out that the informant is (recently) divorced, it is quite understandable if s/he refuses to answer questions relating to his/her family life. However, if on the contrary, s/he suddenly opens up and starts to smear his/her divorced spouse, s/he should be encouraged to talk as long as possible.

4.3   Conversation modules

Conversation modules are a group of questions related to the same topic e.g. child rearing, one's purpose in life etc. (Labov 1984:33).

When engaged in modules one should pay particular attention to the use of colloquial style, the use of any feature that may be considered formal should be avoided.

The precise wording of the question is extremely important. The field worker should by no means resort to improvisation. Some of the questions are marked with two asterisks, meaning they should be put word by word without the slightest alteration. As for the rest of the questions, the field worker should try to say them in the shortest possible form. According to Labov (op. cit. 34) good module questions need not take more than 5 seconds to ask. This brevity can be acquired but at the expense of practice – it is a skill that will never come "of its own".

4.4   The network of modules

The modules can be arranged into a network, but there is no prescribed sequential order. The field worker should start the conversation with the least personal questions like e.g. *How long have you been living here?* and progress gradually towards more and more intimate topics (e.g. religion). The transition from one module to the other should be as smooth as possible (cf. tangential shifts, Labov 1984:37 ff). If the informant shows interest in a topic, it is desirable to return to it later on.

4.5   Practical hints on the guided conversation

1. Although the guided conversation will inevitably contain a lot of dialogue, the field worker should make sure it contains as many long stretches of speech from the informant as possible. To this end:

2. Yes-no questions (e.g. *"Do you like vegetable soup?"*) should be kept to a minimum, and information seeking questions (*why ...*, *when...*, *what happened...*, *please tell me...*) should be used instead.

3. The field worker should aim to speak as little as possible. For this reason, instead of following the informant's speech with constant PHATIC LANGUAGE (*yes*, *aha*, and *indeed* etc.) the field workers should use their eyes only to indicate that they are with the informant.

4. The informant should be given ample time to think and reflect. While the informant is holding a pause, s/he should not be interrupted. Instead, s/he should be encouraged to proceed with enquiring looks and gestures.

5. A good interview is characterized by much speech from the informant and little from the field worker. The field worker should convince the informant that s/he is genuinely interested in what the informant is saying. As far as possible, the field worker should sincerely relate to the personality and problems of the informant.

6. The field worker should avoid interrupting the informant, s/he should not be speaking simultaneously with the informant. The less there is such overlap the better the interview, and vice versa.

7. In the course of guided conversation the field workers should hold a piece of paper ready (this may be the interview log book) and they should jot down the questions that occur to them while the informant is speaking. Natural conversation would require the field worker to respond verbally straight away as the conversation progresses. However, this would unduly cut up the conversation into short exchanges. It is hoped that by "eliciting by eye" and noting down the odd question to be raised later on, the field worker will be able to gain more or less continuous monologues from the informant.

8. It is by no means required that all the modules should be worked through. It is, however, imperative to cover all the obligatory modules marked with K (for *kötelező* 'obligatory' in Hungarian).

9. The minimum time the obligatory and optional modules should take in an interview is 30 minutes. This means that if in a long-winding interview the tests, reading passages and the preliminary conversation last 90 minutes, the field worker should (a) either be prepared at the 90th minute to use the supplementary tape when the two hours run out or (b) to shift to half the recording speed used. If s/he chooses the latter option, s/he should immediately record in the log book the counter setting on the tape recorder and AFTER the interview s/he should record in the book the fact that s/he has changed speed. Not during the interview.

5 General instructions

Before starting the interview the field worker should have the following things ready:

- 2 pcs. of 13 cm diameter Polimer reel tape (one is replacement)
- 1 Uher tape recorder
- 1 (lavalier condenser) microphone
- 1 walkman cassette player
- 2 fresh batteries for the walkman cassette player
- 1 test cassette for the walkman cassette player
- 3 answer sheets entitled:

    Same or different?

    Which is correct?

    How do you say it?

- 7 sheets containing the following reading passages:

  1. Jóska barátom ...
  2. Meghirdettem az újságban ...
  3. A hatodik óra után ...
  4. Pista, bonts fel ...
  5. Felmerült a gyanú ...
  6. Ezerszer megmondtam ...
  7. Hol van a fésű ...

- 1 stapler
- 1 staple remover
- 1 sheet of paper to staple together

- cards:

morphology, syntax and lexicon:	60
minimal pairs	20
word lists	11
fill-in sentences	7
demográfia	1
total	99

- log book

- the profile sheet of the informant supplied by the Survey to the field worker in advance.

At the start of the interview the following headline information should be recorded onto the tape:

1. the field worker's name: e.g. "Gyula Molnár"

2. date of the interview, e.g. "12 October, 1988"

3. type of the interview, i.e."Budapest interview, 3rd version"

4. the name of the informant can be omitted: "TELL ME YOUR NAME IF YOU THINK YOU WANT TO GIVE IT BUT IT'S ALL RIGHT IF YOU DON'T FEEL LIKE IT."

5. tape identifier

For each interview a completely new set of batteries must be purchased (at the Survey's cost) and the batteries MUST be discarded when the interview is over.

The sound quality of the recording must be constantly monitored. Background noise must also be followed with attention. If, for example, the noise of a bus roaring by distorts the recording of a word list, the field worker should wait until the noise abates and the words should be read again.

5.1   The relation between informant and field worker and the confidential treatment of data

Let's try the following tactic:

(1) "We know that people speak in different ways, that is to say, not everybody speaks like the radio and television announcers. A lifeguard speaks a little differently than a lathe operator. Textbooks do not reflect this variety in speech because linguists and textbook writers do not know the way various groups of people use language. We would like to know better the different linguistic varieties, that's why we would like to record a conversation with you on tape."

(2) "Should I ask you anything that you do not feel like answering, please do not hesitate to let me know and then I will immediately pass on to another question."

(3) "The recording with you will be handled confidentially. This means that only a few researchers will be entitled to listen to the tape, no one else. If parts of this conversation are published in a scholarly article, then any item that might possibly serve as clue to your identity e.g. personal names, street names etc. will be changed in a way that no one could recognize you".

5.2   After completing the interview

– as soon as possible but no later than 12 hours after the interview was made, the field worker should record all the information that is indispensable for the analysis of the recording but which is not self-evident for an outsider e.g.

(a) possible gestures during speech, knowledge of which may help to interpret what is said and without which the text cannot be properly interpreted, e.g. "*This has fleas* (pointing at the dog), *but this one has none.* (pointing at the cat)".

(b) the interview setting changed at a given point because somebody (Mr/Mrs/ Miss So-and-so) entered the room whose presence visibly made the informant very tense.

– AFTER every interview the following page from the log book should be filled in:

- At the time of the interview the informant was

    - ill; in particular, he had:

    - was not ill

    - had a visible speech defect:

    - was hard of hearing

    - was under the effect of drugs or alcohol, its rate:

- In the opinion of the field worker, the informant was

    - extremely cooperative

    - cooperative

    - not cooperative

- general characterisation of the informant's speech by the field worker:

Also to be noted in the log book is any linguistically relevant event that is not recorded on the tape. For example, it is conceivable that an informant uses *-suk/-sük* forms in his vernacular, avoids such forms throughout the interview but as soon as the tape runs out starts producing them abundantly. Such an event must be recorded in as much detail as possible.

The field worker should also record every significant mistake that he/she was unable to correct on the spot and which may affect certain parts of the data. The precise location of such mistakes must be recorded in the log book and the mistake must be described in all possible detail.

6 Research tools

The order of research tools below is a recommendation, it is not obligatory. The morphological, syntactic and lexical cards are administered at the beginning of the interview so that we could collect relatively spontaneous data before the linguistic awareness of the informant begins to grow as a result of the "Which is correct?" questionnaire.

The ordering in the section is advisable to keep to.

The different modules of the guided conversation should be interlaced with the tests – to provide relaxation and distraction. When all the tests have been done, the rest of the tape should be filled with conversation.

The Uher tape recorder must not be stopped during the three perception tests either. This will result in minutes of idle tape but possible problems (e.g. misunderstanding of the instruction) and the resulting commentary will be well documented.

On the other hand, the Uher tape recorder could be stopped while the informant is reading the seven passages to himself.

6.1   Cards

There are sixty cards to elicit data on the major questions listed in sections 3.1 and 3.2. The general format is the following: The informant gets a card from the field worker. The card has a sentence with a blank in it. The citation form of the word to fill the blank is given on the card. For example:

> Mari .... egy ingemet tegnap.   KIMOS 'wash'
> 'Mary washed-def./indef. a my shirt yesterday'

So the informant receives the above card (without the gloss, of course) and has to say the sentence with the appropriate form of *kimos*. This means s/he has to choose between the form *kimosta* (definite conjugation) and the form *kimosott* (indefinite conjugation).

For the convenience of our field workers (who are not trained linguists) we provide another set of cards, the field worker's cards, which contain the variants the cards are supposed to elicit. Only the variants of primary data are listed.

PLEASE READ OUT THE SENTENCE ON THE CARD SO THAT YOU INSERT THE WORD ON THE RIGHT INTO THE SENTENCE ON THE LEFT. FOR EXAMPLE:

Én tegnap nem ... eleget.   ALSZIK 'sleep'
'I yesterday not ... enough'

6.2   Reading passages

Each passage will be handed over on a separate sheet with the following instruction:

HERE IS A PASSAGE. PLEASE READ IT CAREFULLY IN SILENCE THEN ALOUD NATURALLY, WITHOUT MUCH ADO AS IF YOU WERE READING IT TO A FRIEND WHOSE EYE HAS JUST BEEN OPERATED ON AND HE CAN'T READ.

After the informant has read the passage:

PLEASE READ THE SAME TEXT FAST. SUPPOSE YOU HAVE TO GO TO THE LOO URGENTLY BUT YOU CAN'T UNTIL YOU HAVE FINISHED WITH READING ALL THIS ALOUD.

(If the informant does not read fast enough, the field worker should urge him by waving his hand.)

The following is one of seven reading passages, slightly altered and provided with detailed explanations for the English reader.

Tibor barátom, akinek a kabátján /1/ két lyuk /2/ is van, fölbiztatta /3/ a vele hasonszőrü /4/ srácokat, hogy irjanak /5/ hosszu /6/ dolgozatokat /7/ – hadd kinlódjanak /8/ a tanárok a javitással! /9/ Mikor már a sokadik nagyon hosszu /10/ dolgozatot /11/ javitották, /12/ a tanárokban /13/ fölmerült a gyanu, /14/ hogy az ujabban /15/ irt /16/ dolgozatok /17/ nem véletlenül ilyen hosszuak. /18/

(Gloss: My friend Tibor on whose coat there are two holes, he encouraged guys like him to write lengthy school papers so that teachers will have a tough time correcting them. After the umpteenth lengthy paper which they corrected, teachers became suspicious that it is no accident that the papers are so lengthy.)

Notes:

1. *kabát-ján*. Palatal assimilation across morpheme boundary, cf. (3) on p. 10

2. *két lyuk*. Palatal assimilation across word boundary, cf. (3) on p. 10

3. *fölbiztatta*. Short or long front high vowels, cf. (8) on p. 12

4. *hasonszőrü*. Short or long *ü* cf. (8) on p. 12

5. *irjanak*. Short or long *i* cf. (8) on p. 12. Cf. the effect of the typewriter: the word *irjanak* 'they should write' is typed on this sheet with the old keyboard, but the same word occurs in a different reading passage typed with the new keyboard, that is, *írjanak*.

6. *hosszu*. Short or long high front vowel and typewriter effect.

7. *dolgozatokat*. Elision of *l*, cf. (5) on p. 11

8. *kinlódjanak*. Short or long *i* and the typewriter effect, plus see 1. above.

9. *javitással*. Same as in 8.

10. *hosszu*. See 6. above.

11. *dolgozatot*. See 7. above.

12. *javitották*. See 5. above.

13. *tanárokban*. *-ban* 'in' or *-ba* 'into'. Cf. (9) on p. 12

14. *gyanu*. See 6. above.

15. *ujabban*. See 6. above.

16. *irt*. See 5. above.

17. *dolgozatok*. See 7. above.

18. *hosszuak*. See 6. above.

6.3   Minimal pairs

The 20 minimal pairs will be handed over on 20 cards to the informant with the following instructions:

> NOW I WILL GIVE YOU A FEW CARDS SHOWING TWO WORDS EACH. PLEASE READ THEM ALOUD.

1.   ember – emberben
2.   bontsd fel – bonts fel
3.   KISZ-szoba – kis szoba
4.   adja – hagyja
5.   elmegyünk – felmegyünk
6.   hat szoros – hatszoros
7.   lehet – lehetett
8.   erdőben – erdőbe
9.   ránts le – rántsd le
10. rácsszerű – Rácz-szerű
11. atyja – adja
12. megtanultam – megtanultuk
13. beteges – betegesen
14. lombtalanít – lomtalanít
15. látják – látták
16. följössz – följösztök
17. kertbe – kertben
18. nem kap semmit – nem kapsz semmit
19. ezerszer – ezeregyszer
20. írt – irt

6.4   Same or different?

6.4.1   Instructions

The same or different test is administered by means of a Walkman cassette player and headphones. The instructions are the following:

> YOU WILL HEAR PAIRS OF WORDS. PLEASE DECIDE IF THE WORDS IN A PAIR MEAN THE SAME OR NOT. EXAMPLES:
> *alma* 'APPLE' AND *körte* 'PEAR'
> IF YOU ANSWERED K FOR 'DIFFERENT', YOUR ANSWER IS CORRECT.

There are four more examples to sensitize informants to different meanings, same meanings, and same meanings with variant pronunciations.

The recording on the Walkman player is loud enough for the field worker to hear through the earphone. This arrangement enables the field worker to monitor by eye and ear whether the informant has lost track, in other words, whether s/he was not filling in the answer to, say, item 5 when hearing item 6.

Answers to questions 5 and 9 should be paid particular attention to! The dialects where the *ë/e* distinction is operative can be diagnosed by the following answers:

5. *hëgyës* (pointed) vs. *hëgyes* (mountainous)

9. *értem* (for me) vs. *értëm* (I understand)

If either one or both answers are K (different), the field worker should ask the informant after the test "What is the difference between *hëgyës* and *hëgyes*?" as well as "What is the difference between *értem* and *értëm*?".

6.4.2   Contents of the test tape

1. gyanu – gyanú 'suspicion', variants

2. kinn – kint 'outside' , morphological variants

3. fel – föl 'up', attention check (see below)

4. sor – sör 'line – beer', attention check

5. hëgyës – hëgyes 'pointed – mountainous'

6. írjanak – irjanak 'write', variants

7. meglássa – meglátja

>   standard gloss: 'he should see – he sees'

>   nonstandard gloss: 'he sees – he sees'

8. jelenkezik – jelentkezik 'identifies himself', variants

9. értem – értëm 'for me – I understand' in two-*e* dialects

10. hat – hát 'six – back', attention check

11. házban – házba 'in vs. into house'

12. igér – ígér 'he promises', variants

13. csinálnák – csinálnék

>   standard gloss: 'they would do – I would do'

>   nonstandard gloss: 'I would do – I would do'

14. hosszú – hosszu 'lengthy', variants

15. ossza – osztja

>   standard gloss: 'he should distribute – he distributes'

>   nonstandard gloss: 'he distributes – he distributes'

16. kertbe – kertben 'into – in garden'

17. kínt – kint 'agony+acc. – outside'

18. fönn – fenn 'upstairs', attention check

19. elbirál – elbírál 'he evaluates', variants

20. engemet – engem 'me', morphological variants, attention check

21. kiírt – kiirt 'wrote out – exterminates'

22. fésü – fésű 'comb', vowel length variants

N.B. Attention checks (e.g. items 3, 4, and 18 above) are used to monitor the informant's attention level. These word pairs are linguistically uninteresting. If an informant scores wrong on these items, analysts know that the answers of this particular informant should not be considered.

Figure 6.1 shows the answer sheet which the informants were asked to fill in.

Azonos jelentésű szavak vagy sem?
'Do the words have identical meaning or not?'

Válaszlap
'answer sheet'

A = Azonos jelentésű szavak 'Words have identical meaning'
K = Különböző jelentésű szavak 'Words have different meaning'

Példák: alma – alma (A)K
  alma – körte A(K)

Próbafeladatok 'trial items':
1. A   K
2. A   K
Példa: kosut (=Kossuth) – kosút (=Kossuth) (A) K

Próbafeladatok 'trial items':
1. A   K
2. A   K
3. A   K

Kezdhetjük? 'Can we start?'
1. A   K   11. A   K
2. A   K   12. A   K
        ⋮
                21. A   K

Figure 6.1: Answer sheet for the 'Same or different?' test

6.5   Word lists

The words will be presented to the informant on 11 cards. The instruction is as follows:

> I WILL GIVE YOU A FEW CARDS. EACH HAS FIVE OR SIX WORDS. PLEASE
> READ THEM OUT!

(See Figure 6.2.)

1. injekció	2. konkurrál	3. röntgen
ember	zsűri	kis szobában
erdőbe	klozett	nem kap semmit
bontsd föl	nejlon	lomtalanít
egyszer	ventillátor	kertben
	hívő	
4. juice	5. rántsd le	6. kertbe
eljösztök	beteges	lombtalanok
adja	látják	tanultatok
nyolc szoros	spray	rácsszerűen
KISZ-szoba	te sem kapsz semmit	ránts le
7. erdőben	8. hűvös	9. grapefruit
lehet	ventilátor	hagyja
adja	színház	hatszoros
bonts fel	zsüritag	leírta
stress	útiköltség	Rácz-szerű
	klozet	
10. software	11. istenhivő	
elmegyünk	fésű	
atyja	konkurencia	
Sopianae	bölcsőde	
ajánlkozik	nylon	
	háború	

Figure 6.2: Word list cards

6.6   Which is correct?

6.6.1   Instruction to field workers:

> *Re* word pair 9 (*értëm* 'I understand' and *értem* 'for me'):

An informant who genuinely distinguishes between *e* and *ë* may well circle both numbers on the answer sheet despite the instruction to mark a single answer for each item. If this is the case, the informant must be asked why s/he marked both numbers. Any spontaneous response or commentary must be recorded with the Uher, but the field worker should avoid discussing right or wrong ways of doing the tests! If the informant is inquisitive, the field worker should try to pass on by saying "We will come back to this at the end". If they do return to the point, the resultant conversation should be recorded with the Uher!

6.6.2   Transcript of the tape recording

The following instruction was recorded on the tape:

> YOU WILL HEAR TWO DIFFERENT PRONUNCIATIONS OF CERTAIN WORDS. THE QUESTION IS WHICH PRONUNCIATION IS CORRECT? IF THE FIRST VERSION IS CORRECT, CIRCLE NUMBER ONE. IF THE SECOND VERSION IS CORRECT, MARK NUMBER TWO. HERE IS AN EXAMPLE: *szupermarket – supermarket*. IF YOU CIRCLED NUMBER ONE, YOU WERE RIGHT: THE CORRECT PRONUNCIATION IS *szupermarket*. CAN WE START?

Figure 6.3 contains the annotated list of test items in this module.

6.7   How do *you* say it?

In order to establish the linguistic insecurity index of the informants, two sub-tests are used: the first asks for the correct form (cf. "Which is correct?" test), while the other investigates the forms used by the informant. The discrepancies show the linguistic insecurity of the informant.

Before the test is delivered the field worker must get the informant to spontaneously produce the word *körút* 'boulevard'. A suggested tactic may be the following:

> IF SOMEBODY FROM A PROVINCIAL TOWN ASKED YOU IF THE WESTERN RAILWAY STATION IS ALONG NÉPKÖZTÁRSASÁG STREET, WHAT YOULD YOU TELL THEM?

Expected reply:
"*Nem, a körúton van.*" 'No, it is along the boulevard'.

1.	nejlon – nájlon	'nylon', variants
2.	hivők – hívők	'believers', variants
3.	ember – embër	'man', cf. open vs. close *e*
4.	tanultam – tanútam	'I studied', *l* deletion
5.	grepfruit – grépfrut	'grapefruit', variants
6.	eszem – eszek	'I eat', cf. *ik* verbs
7.	öccörös – ötszörös	'quintuple', cf. affrication
8.	aggya – adja	'he gives' cf. palatal assimilation
9.	értëm – értem	'I understand – for me', cf. *ë/e*
10.	injekció – inekció	'injection', variants
11.	jöttök – jösztök	'you-pl. come', variants
12.	dícsér – dicsér	'praise', variants
13.	bólintgat – bólingat	'keeps nodding', cf *t* elision
14.	meggyőződjön – meggyőződjék	'be convinced', cf. *ik* verbs
15.	gyanúsítja – gyanusítja	'he suspects', variants
16.	szopiáne – szofiáne	'cigarette brand', variants
17.	szivesen – szívesen	'you are welcome', variants
18.	etyszer – eccer	'once', cf. affrication
19.	elbírálás – elbirálás	'evaluation', variants
20.	jelenkezik – jelentkezik	'identifies himself', cf. *t* elision
21.	bölcsöde – bölcsőde	'crêche', variants
22.	akassza – akasztja	standard gloss: 'he should hang – he hangs' nonstandard gloss: 'he hangs – he hangs' cf. *-suk/-sük*
23.	pósta – posta	'post office', variants

Figure 6.3: Test items in the "Which is correct?" module

The field worker should note down on the answer sheet whether the informant uttered the word with a long or short *ö* by circling X if the vowel was long and Y if it was short. Then he should note whether the informant does the trial test correctly. If the informant does not seem to have understood the task correctly, then another example should be given, such as the following:

WHAT IS IT THAT CAUSES AN IRRITATING BITE IN SUMMER? (*szúnyog* 'MOSQUITO') OK. NOW I AM GOING TO SAY THIS WORD IN TWO WAYS. IF YOU USUALLY SAY IT THE WAY I SAID IT FIRST, THEN CIRCLE NUMBER ONE. IF THE WAY YOU SAY IT IS HOW I SAID IT THE SECOND TIME, CIRCLE NUMBER TWO.

1.	jösztök – jöttök	'you-pl. come', variants
2.	értëm – értem	'I understand – for me', cf. *ë/e*
3.	injekció – inekció	'injection', variants
4.	dícsér – dicsér	'praise', variants
5.	adja – aggya	'he gives' cf. palatal assimilation
6.	bólintgat – bólingat	'keeps nodding', cf *t* elision
7.	öccörös – ötszörös	'quintuple', cf. affrication
8.	meggyőződjön – meggyőződjék	'be convinced', cf. *ik* verbs
9.	grepfruit – grépfrut	'grapefruit', variants
10.	gyanúsítja – gyanusítja	'he suspects', variants
11.	ember – embër	'man', cf. open vs. close *e*
12.	szopiáne – szofiáne	'cigarette brand', variants
13.	tanultam – tanútam	'I studied', *l* deletion
14.	hivők – hívők	'believers', variants
15.	eszem – eszek	'I eat', cf. *ik* verbs
16.	jelentkezik – jelenkezik	'identifies himself', cf. *t* elision
17.	szívesen – szivesen	'you are welcome', variants
18.	eccer – etyszer	'once', cf. affrication
19.	nájlon – nejlon	'nylon', variants
20.	elbirálás – elbírálás	'evaluation', variants
21.	bölcsőde – bölcsöde	'crêche', variants
22.	gyanu – gyanú	'suspicion', variants
23.	megdícsérem – megdicsérem	'I praise', variants
24.	labdázom – labdázok	'I play ball', cf. *ik* verbs

Figure 6.4: Test items in the "How do YOU say it?" section

1.	lombtalanítják	17.	hegyes 'pointed'
2.	rántsd le	18.	látja
3.	ember	19.	konkurencia
4.	hatszoros	20.	bátyja
5.	lomtalanít	21.	nem kap soha semmit
6.	kis szoba	22.	hívő
7.	ránts le	23.	bontsd fel
8.	beteges	24.	rácsszerű
9.	ventilátor	25.	nem kapsz soha semmit
10.	adja	26.	hagyja
11.	zsűri	27.	bonts fel
12.	hegyes 'mountainous'	28.	tekinget
13.	nejlon	29.	elmegy
14.	följösztök	30.	bölcsőde
15.	klozet	31.	bólintgat
16.	egyszer	32.	postán

Figure 6.5: The list of elicited lexical items

6.8 Questions

The test words elicited in the minimal pairs section, word list section, reading passages and investigated in the listening tests are finally used in this part of the interview, which is like all traditional dialect questionnaires, i.e. questions are asked of the informant to elicit one particular word form. Figure 6.5 shows the list of lexical items elicited in this section.

6.9 Fill in the right word

Instruction: FILL THE BLANK WITH THE APPROPRIATE WORD IN EACH OF THE FOLLOWING SENTENCES.

Sok mindenre emlékszem, ... gyerekkoromban történt.   AMI/AMELY 'which'
'Many things I remember ... in my childhood happened.

This card elicits data for relative *ami* vs. *amely* cf. 3) on p. 20. In this section data are obtained for the following variables:
ragassza – ragasztja, cf. *-suk/-sük*
ami – amely
the question particle *-e* cf. 1) on p. 19
nyitja – nyissa cf. *-suk/-sük*
akassza – akasztja cf. *-suk/-sük*

6.10 The reporter's test

This test (cf. Ball 1986) is used to elicit words from informants without their concentration on linguistic form. The field worker asks the informant to act as a reporter doing a running commentary of what the field worker acts out. Here field workers have to elicit the following words:
felgyújtja or *felgyújcsa* (cf. *-suk/-sük*),
felgyújtja or *feloltja* (cf. turn on vs. extinguish up),
kinyitja or *kinyissa* (cf. *-suk/-sük*)
felakasztja or *felakassza* (cf. *-suk/-sük*)
szobában or *szobába* (cf. *-ban* 'in' vs. *-ba* 'into')
táskába or *táskában* (cf. *-ban* 'in' vs. *-ba* 'into', hypercorrection)

6.11 What does *demográfia* mean?

Field worker to informant: "Do you know the word *demográfia?*"Field worker records yes or no answer in log book then hands over the following card:

What does *demográfia* mean?

1. A policy which should result in increasing numbers of births.

2. The study of the changes in numbers of births, deaths, marriages and diseases in a community over a period of time.

3. Both meaning 1. and meaning 2. are correct, this word has two meanings.

6.12 Conversation modules

There are three kinds of modules:

1. those which are obligatory to do,

2. those which must be introduced verbatim (e.g. 'danger of death' question),

3. and the rest.

1. and 2. may overlap.

As regards guided conversations in the interview, field workers are encouraged to do a few personal data oriented modules at the beginning (e.g. your birthplace, your work, your mother tongue). Then, after all the test-like parts of the interview, field workers need to conduct conversations for at least thirty minutes.

Listed below are some conversational modules:

6.12.1 The mandatory ones:

1. Personal data

2. Parents' and spouse's personal data

3. Occupation

4. Danger of death à la Labov

5. A question on the Gipsy problem

6. Questions related to language e.g. "Where is the best Hungarian spoken?", "Which part of society speaks the best Hungarian?", "Is a dialect speaker hindered by his/her speech in upward social mobility?" etc.

6.12.2 Optional verbatim modules:

1. Question on abortion:

   "CZECHOSLOVAKIAN WOMEN DO NOT NEED A PERMIT TO HAVE ABOR-
   TION. HUNGARIAN WOMEN DO. THUS HUNGARIAN WOMEN CANNOT
   FREELY DECIDE WHETHER OR NOT THEY GIVE BIRTH TO A CHILD.
   WHICH POLICY DO YOU LIKE BETTER, THE HUNGARIAN OR THE
   CZECHOSLOVAKIAN ONE? WHY?"

2. A Gipsy question:

   "DO YOU KNOW WHAT THE ABBREVIATION *CMÖ* MEANS? (If the answer
   is "no": *Cigánymentes övezet* 'Gipsy-free area'.) THIS ABBREVIATION IS
   OFTEN SPRAYED ON BRIDGES OR WALLS OF HOUSES IN BUDAPEST,
   E.G. ON HIGHWAY 3 THERE IS A LARGE GRAFFITI. WHAT DO YOU THINK
   OF THIS?"

6.12.3 Neither verbatim nor mandatory modules:

1. WHAT ABOUT NUCLEAR POWER STATIONS?

2. WHAT ABOUT THE HUNGARIAN REFUGEES FROM ROMANIA?

3. DO GIRLS FIGHT AROUND HERE?

4. THERE ARE NO BROTHELS IN HUNGARY BUT THERE IS PROSTITUTION IN THE
   STREETS AS WELL AS IN HOTELS. IS THIS A GOOD THING? WHY?

7 The processing of data

The transcription is done by means of SONY BM-80 desktop transcribers and IBM PC/XT compatible personal computers. The entire material of the interview is transcribed and/or coded straight onto electronic medium.[1]

Computationally, the data collected through the interview fall into two broad categories: (1) test-like tasks and (2) continuous speech. These two kinds of data require different treatment:

CONTINUOUS SPEECH is transcribed with the help of a specially programmed word processing program in the form of a standard ASCII text file; TESTS are processed by means of a database management program so that only relevant parts of the informants' responses are recorded and coded. Both parts are integrated in the same system (dBASE III plus), which means that the transcriber can shift fairly easily between the transcription of the continuous speech and the coding of test items.

Although all key aspects of the interview are carefully controlled, the length and structure of each interview will inevitably be different. For our purposes the structure of the individual interviews can be seen as a sequence of conversation and test modules, where the number and ordering of such units are not rigidly controlled. It is essential, therefore, that the system should be flexible enough to accommodate such a varied material, yet every single item should be uniquely identifiable and amenable to further processing.

7.1 Transcription of guided conversations

7.1.1 The basic philosophy

The basic philosophy of transcription is the following:

1. Because automatic grammatical tagging of Hungarian will not be a reality in the near future, only partial grammatical analyses can be carried out. Thus, anything that can be investigated by means of concordances and other text processing soft-

---

[1] The present chapter is included for historic accuracy. Nearly ten years after it was written, it has been made largely obsolete by the major revision of the software implemented since. It is also made redundant by the appearance of Váradi 1998, which is dedicated to a thorough review of the questions briefly discussed here.

ware will be examined. Also, some selected grammatical phenomena will be manually coded in the transcripts (e.g. all occurrences of *-ba* 'into' instead of *-ban* 'in').

2. In principle the Budapest Survey should aim to transcribe all conversations in full prosodic transcription. We do have the transcription system to do this but we do not have the money and manpower.

7.1.2   Format conventions

The conversations are transcribed in the following format:

The whole of the spontaneous speech from a single informant is entered in a single file. Each conversation module (CM) is recorded in a single paragraph, that is, an empty line is used to set off one conversation module from another. Each CM is headed by an identifier line consisting of the following information:

columns	Content
1-5	identifier of the informant
6-8	identifier of the conversational unit
10-15	location of CM on tape
17-18	identifier of transcriber
20-30	date

Each line of transcribed speech is 80 characters long and consists of the following parts:

columns	Content
1-5	identifier of the informant
6-8	identifier of the conversational unit
10-13	line number within CM
15	identifier of current speaker[2]
16	continuity marker[3]
17-72	text
74-79	location on tape

The last line of a CM consists of information (in exactly the same format as the header information detailed above) of the unit (whether conversational or test unit) that follows the current unit.

7.1.3   Sample transcription

Figure 7.1 shows a sample page of transcript. It shows excerpts from two conversation modules. The first marked *bio* on personal data of the informant is followed by a test

---

[2] *t* (terepmunkás) 'field worker' or *a* (adatközlő) 'informant'
[3] i.e. :=new turn, >=old turn continued

B7003bio 1a0042 RA 1988.07.18.
B7003bio 0001 a:Ott volt mint □ ö ált= állattenyésztési vezető, □ *s*
B7003bio 0002 t:*Igen.*
B7003bio 0003 a> aaa édesanyám pedig háztartásbeli volt. □ ööö (a)
B7003bio 0004 a édesanyámnak a szakmája tanítónő volt □ valamikor, de
B7003bio 0005 a <0a> háború előtt tanított, □ <0a> háború után már nem
B7003bio 0006 a tanított.
B7003bio 0007 t:Igen, □ igen, □ értem. □ Namost, □ egészen tizennégy
B7003bio 0008 t éves koráig tehát akkor ott élt, ott lakott, ott járt
B7003bio 0009 t iskolába.
B7003bio 0010 a:Igen, igen.
B7003bio 0011 t:A Nyírségbe<n>, *ugye*?   1a0300
B7003bio 0012 a:*Igen.*

B7003vl1 1a0305 RA 1988.07.18.
⋮

1988.07.18.
B7003cmö 1a1529 RA 1988.07.19.
B7003cmö 0001 t:<0a> szomszédasszonyomnak már □ kinyitotta a táskáját.
B7003cmö 0002 a:Hát nekünk is □ az vo<l>t a szerencsénk, mer<t> a
B7003cmö 0003 a kolleganőnkkel mentünk az utcán,és □ aszondja nekem az
B7003cmö 0004 a Erika, hogy □ turkálnak a táskámba<n>. És hátranézek, és
B7003cmö 0005 a □ egy cigány férfi fogja a gyereke kezit, a másik
B7003cmö 0006 a kezivel az Erika táskájába nyú<l>ká<l>, a cigány nő az   1a1540
B7003cmö 0007 a ölébe<n> tartsa <s><=tartja> a gyerekét, és az én
B7003cmö 0008 a táskámba<n> *turká<l>*.
B7003cmö 0009 t:*Őrület*.

B7003vl2 1a1548 RA 1988.07.19.

The meaning of the codes used in the body of the transcription is as follows:

B7003bio 0001	□	pause
	ö	short hesitation
	=	unfinished word
	* *	simultaneous speech
B7003bio 0003	aaa	lengthening as hesitation
	ööö	long hesitation
	( )	the transcriber is uncertain of what s/he hears
B7003bio 0005	<0 >	ungrammatically omitted word
B7003bio 0011	<n>	*ba* is pronounced instead of *ban* (cf. 9) on p. 12)
B7003cmö 0002	<l>	*l*-deletion (cf. 5) on p. 11)
	<t>	*t*-deletion (cf. 6) on p. 11)
B7003cmö 0007	<s>	*-suk/-sük* conjugation (cf. 3. on p. 18)
	<= >	explanation; the standard spelling form

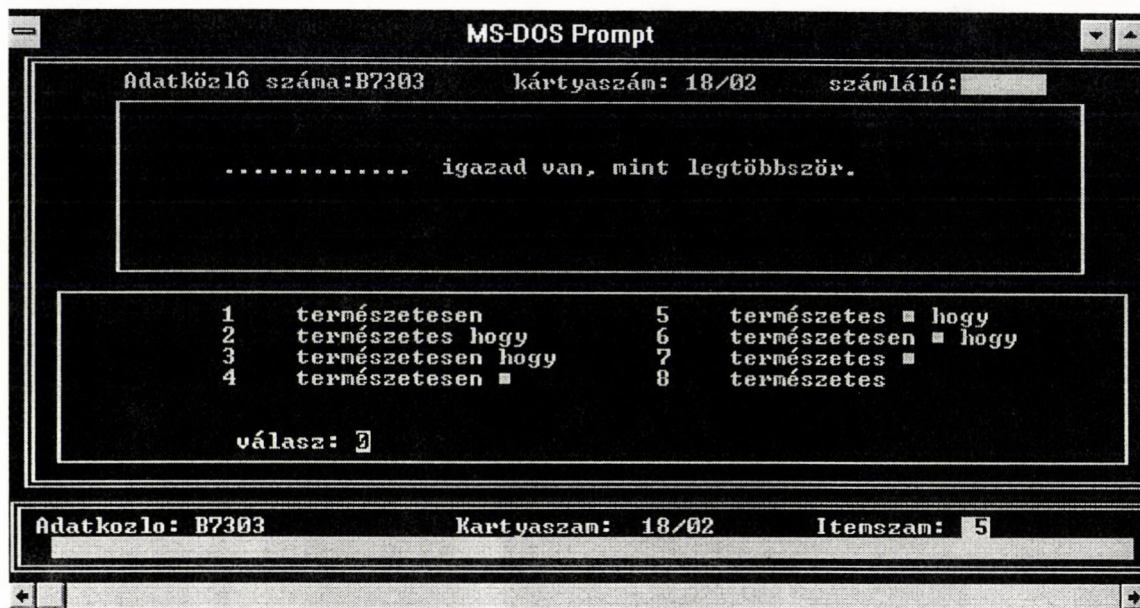Figure 7.1: Transcription of conversation modules

Figure 7.2: Screen print of old data entry program in 1988

coded *vl1* as well as other modules omitted here for lack of space. The second module (coded *cmö* on the Gipsy question) is followed by test unit *vl2*.

7.2 The coding of test-like material

The majority of the test materials involves the informant reading out or saying what s/he thinks is the correct response. Only the relevant parts of the informant's responses are coded by the transcriber. Coding is done through screen masks containing the original stimulus sentence and the anticipated reponses with the numerical codes supplied.

The assignment of the various items to primary or secondary data status is done automatically and so is the assignment of the individual cards to the various research questions they are aimed to survey. Owing to the intricate nature of the testing involved, not only is it the case that a single test sentence may examine a number of different research questions but also the same research question may be involved in a number of different test sentences (as well as in the guided conversations, of course).

Since the original compilation of this document the data processing software has been thoroughly revised. See Váradi 1998 for more details. Figure 7.3 shows a screen print of the revised data entry system.
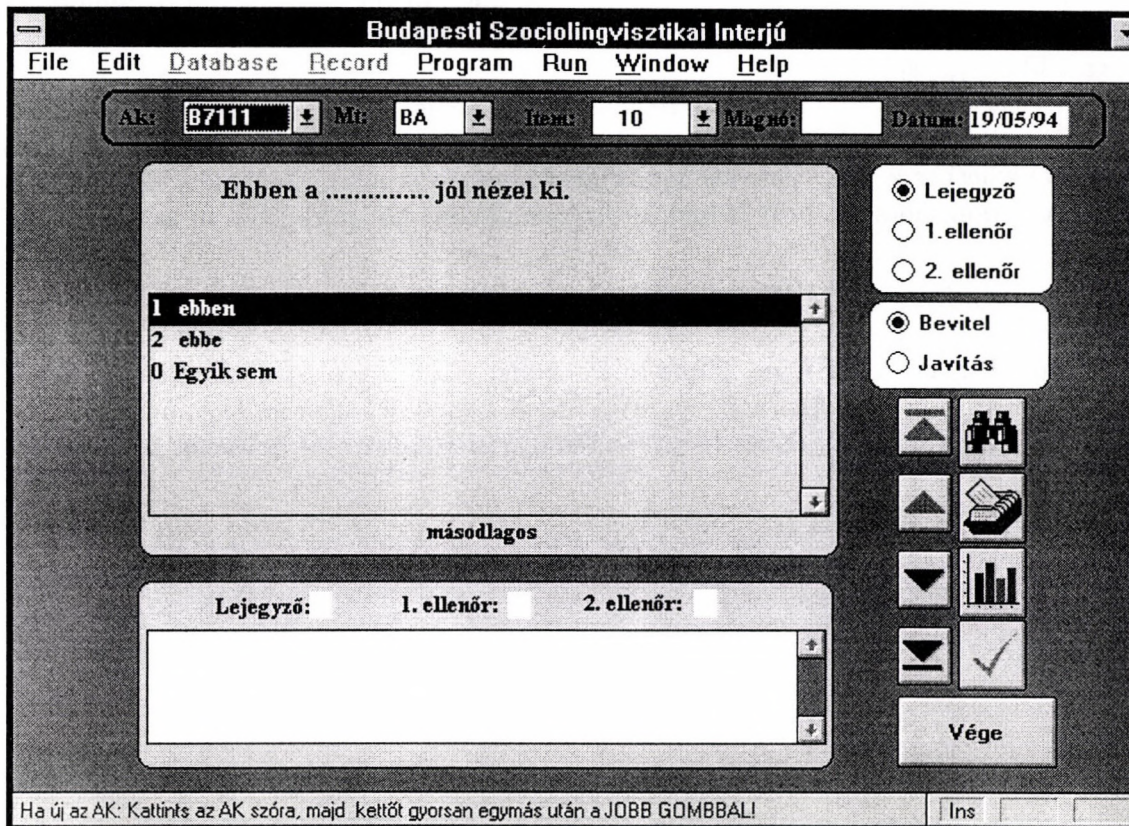
Figure 7.3: Screen print of revised data entry program in 1994

7.3   Some envisaged applications of the system

With the help of the present system it will be possible to tell exactly what cards examine the same research questions e.g. *-suk/-sük* conjugation as secondary and primary features; what was the distribution of the informants' responses over the total number of contexts in which the question was analysed or any subset of them.
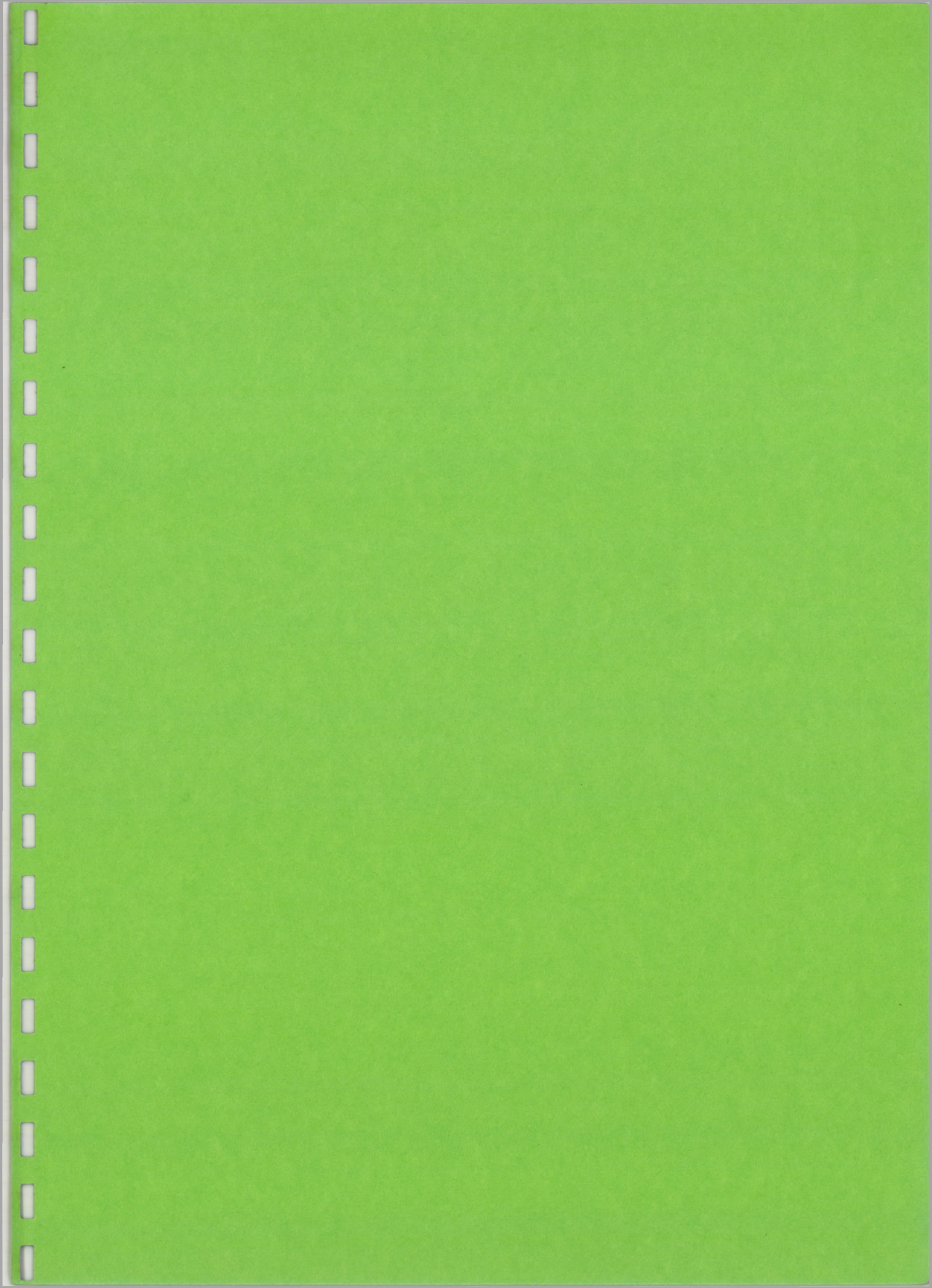
Furthermore, because this issue is manually coded in the transcription of guided conversation as well, one can also collect accurate information (through concordance searches) about the incidence of the same question in the entire set of one informant's utterances.That is, test data and conversational data can be collected for any selected variable (e.g. *-suk/-sük*).

As each line is equipped with reference to its locus, it will be possible to examine the distribution of certain features in a given conversation module only, e.g. it will be easy to say whether a particular lexeme or grammatical variable is spread evenly across all conversation modules, or it is frequent in one module but infrequent in another.

7.4   References

Váradi, Tamás. 1998. *From Cards to Computer Files: Processing the Data of The Budapest Sociolinguistic Interview.* Working Papers in Hungarian Sociolinguistics No. 3, January 1998. Linguistics Institute, Hungarian Academy of Sciences, Budapest.

Working Papers in Hungarian Sociolinguistics