The language attrition test battery

A research manual

by

Monika S. Schmid
Vrije Universiteit Amsterdam

TABLE OF CONTENTS

1 Before you get started

If you are planning to conduct research on language attrition, you are probably champing at the bit to go out there and collect your data. However, some good thinking at this stage will save you headaches later. Those considerations should include:

1. Subject characteristics (e.g. minimum age at emigration, minimum length of time in emigration. You may also have some requirements on language contact, e.g. you may not want anyone whose partner also speaks the L1 or who uses it on a daily basis)
2. Subject acquisition: how are you actually going to find people to test?
3. Test battery – what do you want your subjects to do
4. Equipment requirements – what do you need in order to conduct, record, store, classify and analyze your data

1.1 Subject characteristics

I think (at this point in time) it is wise to set as minimum the following: subjects who were **no younger than 17** at the time of emigration, and who have stayed in the country of emigration for **a minimum of 15 years**. Both these limits are quite cautious, but will ensure that there is no undesired within-group variation according to these variables (see Köpke & Schmid 2004:9-12). I think it is necessary to investigate and understand attrition in its 'stable state' before we can afford to turn to the development.

Other criteria, such as **amount of contact with the L1**, are for the researcher herself to determine – some research designs will be targeted towards people with as little contact as possible, while others may be particularly interested in the impact of intensity of use and therefore have to cast their net a bit wider. Bear in mind that, if you want to add extra criteria, it is wise to get that information from potential participants as early as possible.

1.2 Subject acquisition

The best way of finding informants is to advertise in both regular newspapers and in expat media (if such exist, see below). In order not to alert potential participants to the intent of the study, we have found it useful to make the claim that we are investigating language change and therefore are looking for people who, having lived outside the country, have 'missed out' on some of it. Here is an example of such an ad:

Research on German language change:
Participants wanted for linguistic experiment!
It has long been established by linguistic science that all human languages are undergoing a constant process of change. For many languages, this change has been sped up considerably over the past years by technological innovations such as the internet. Where the German language is concerned, an additional impetus was provided by the historical process of the reunification. We would like to investigate this, and are looking for Germans in the Vancouver area who have been living in Canada for at least the past 15 years (or longer).
There is no required specialized knowledge – we're looking for you, whether you speak German on a daily basis or virtually never.
The experiment will take around 2 ½ hours and will take place at a location of your choice between June 15th and July 31st. Participants will receive $ 30.
If you are interested, please contact:

In addition to that, a very useful way of finding contacts, expat newsmedia etc. is through the embassy (or consulate) of the country your intended subjects come from. You can find the embassy/consulate on the internet, and they are usually very helpful about providing you with the addresses of clubs, organisations, radio stations, newspapers, you name it… You can then write to these, asking them to circulate your request for participants among their members/audience.

A word of warning: you may get more than you bargained for. It is advisable

a) not to mention private phone numbers,
b) to keep very good track of who has contacted you, and
c) to send them all very nice and polite letters, whether you will use them as participants or not. Trust us, we know…

When initial contact is made, you should send those interested a preliminary questionnaire, e.g.:

Questionnaire

1. Personal information
   Last name:
   First name:
   female ☐ male ☐
   Year of birth:
   In what country were you born?
   In what country did you grow up?[1]
   Since what year have you been living in Canada?
   What is your highest school or university degree?
   What is your profession?
2. Family situation
   Are you married or living in a permanent relationship? ☐ yes ☐ no
   If so, what is the native language of your partner or spouse?
   ☐ German ☐ English ☐ other, namely:

[1] This question may save you going to meet and interview people and then finding out that they are not actually first-language speakers.

3.	Language use How often do you use German within your family? ☐ daily ☐ weekly ☐ monthly ☐ less frequently ☐ never How often do you use German with friends? ☐ daily ☐ weekly ☐ monthly ☐ less frequently ☐ never How often do you use German in your working life? ☐ daily ☐ weekly ☐ monthly ☐ less frequently ☐ never

Planning the actual experiments both chronologically and geographically is a nightmare. My office was papered with enlarged copies of maps that had little coloured pins stuck in them (making my colleagues speculate on the Germans finally trying to finish what they attempted 60 years ago...). Get a good map as early as possible, and make a lot of copies that you can write on. I have found it a helpful to sort my list of participants' addresses by postcode.

1.3 The test battery

The rationale for this test battery has been set out in Schmid (2004), so I will not go into it here. We suggest the following tests (detailed information on each can be found below):

1. Sociolinguistic questionnaire
2. Matched guise
3. Grammaticality judgment
4. C-test L1
5. Charlie Chaplin film re-telling task
6. Wug-test
7. Fluency in controlled association
8. C-test L2
9. Can-do scales

Obviously, this list can be augmented to include tests that will specifically focus on some aspects of particular interest to you.

1.4 Equipment

Hardware: You will need a laptop computer with a DVD drive, a minidisk recorder (or good quality tape recorder) and a foot pedal for transcribing.

Software: Microsoft Office, including PowerPoint, SPSS, some recording software such as TotalRecorder. Optional but very useful is sound editing software, such as SoundForge, which can help you improve the quality of your sound data and raise the volume (which is often necessary).

To many people, the actual recording and trasncribing is one of the biggest headaches, so here are some tips and info.

a) The minidisk or tape recorder: Both devices record audio data in a format that cannot directly be read into the computer. While the present generation of minidisk recorders comes equipped with a USB cable, this can only be used to digitally transfer data from the computer to minidisk (e.g. music you've illegally downlowaded from the Internet) and not vice versa. This will probably change soon. If your minidisk recorder cannot do this, or if you have a tape recorder, you will need a mono jack cable (one which has the same plug as headphones or a microphone). You plug this cable into the headphone jack of the recorder and the microphone jack of your soundcard. In order to record the sound and save it as .wav or .mp3, you will need recording software. I can recommend the program TotalRecorder (http://www.highcriteria.com/, costs about US$ 12,00 and is worth it).
Of course this means that recording the data onto your computer will happen in real time. I've found it saved me a lot of time to connect the minidisk recorder to the computer at the beginning of each experiment, and to record it 'live' via TotalRecorder. This also provides you with a backup, should the minidisk recorder choose to erase your data (which it did to me twice).

b) The transcription kit: do **NOT** attempt to transcribe your files using the buttons on your recorder to play and rewind. You will go mad. You can purchase a transcription kit, which comes with a program to install on your computer and a foot control pedal which you connect to the computer via a USB cable, for around 120 EUR. It will be the best investment you've ever made. One such transcription kit is the Sony Digital Voice Editor. Once you have saved your sounds as .wav files, either on CD or on your hard disc, you can open them in the Digital Voice Editor and use the foot control pedal for playback, rewinding etc. One very handy function is that you can also slow down playback. It makes people sound like they are totally stoned, but is very handy to keep up when you're typing.
Alternatively, you can use the program PRAAT which can be downloaded (for free) from www.praat.org.

2 Getting started – How to construct and conduct your tests

Obviously, this is the most important bit, and these guidelines can only be rather general, as many of the tests will vary quite a bit, depending on which language(s) you are investigating and what you are after.

2.1 Sociolinguistic questionnaire

a) Construction

You can find English, Dutch and German versions of the sociolinguistic questionnaire in the folder “Sociolinguistic questionnaire” on the CD. An English version is also included in Appendix 1.

b) Administration

We have found it useful not to hand the questionnaire to the informants, but to use it as a guideline for a semi-structured interview. Take the respondent through it as gently as possible and try to make the conversation natural. Just let them talk, to stimulate free speech. This typically takes between 30 min and 1 hour. Try to gently wind up after that – which will not always be easy.

Your own role will take some getting used to. What you want to do is encourage people to talk as much as possible (and to have long turns), while you yourself talk as little as possible.

c) Classification of results

Please do this in the format indicated in the Excel-file “Template for entering results.xls”, sheet ‘SQ Raw data’. on your CD.

In addition, the free spoken data you obtained through the interview should be transcribed. In order to cut down on the work load, I limited my transcriptions as much as possible to what the subjects said, leaving out my own contributions unless they were strictly necessary to understand the responses.

2.2 Matched guise

a) Construction

The first, and most difficult, thing you have to do here is find two speakers who are fully and convincingly bilingual in both the L1 and the L2 of your participants. We would suggest that you let them read (a translation of) the text we’ve used (see Appendix 2 for the Dutch, English and German version and the files on your CD for the actual readings), but of course you can use a different text if you want.

b) Administration

I have found that the best way to present the recordings is to incorporate the sound files into a PowerPoint presentation (see the files on the CD), but of course you can also play them back from an audio CD if you prefer. You explain the test (be sure to say very explicitly that each voice will be reading **the same text**, and that the text itself is not to be taken into account when rating the personality), and then play the first voice. After listening to each recording, give the informants the time they need to fill in the list of characteristics (see Appendix 2 for the English version and the files on your CD for the Dutch and German version). While they do this, please watch to make sure they place a cross in each line – often, they'll think that they have to select only one characteristic etc. Explain to informants that this is about how they interpret characteristics of speech and voice, not to see whether their judgment of someone is accurate. You'll sometimes encounter resistance to this test ("You can't really tell all that just from someone's voice!"), I found it helpful to give the example of someone you only know via telephone and then meet – often, you're very surprised because you've imagined someone entirely different.

c) Classification of results

In accordance with common practice in sociolinguistic research, we propose a subdivision of those 13 characteristics into two dimensions: solidarity and prestige.

Solidarity		Prestige	
+	-	+	-
sociable	aloof	intelligent	unintelligent
pleasant	unpleasant	a good leader	obedient
decent	rough	factual	approachable
attractive	unattractive	ambitious	easy to please
graceful	awkward	educated	uneducated
polite	impolite	high status	low status
tolerant	intolerant		

Enter each characteristic in the file 'Template for entering the results.xls', sheet 'MG Raw data' in the order in which they were presented. When entering this, **be sure to count the – pole in the table above as 1, and the + pole box as 5**, no matter which side of the boxes they are on! They will automatically be recalculated into average solidarity and prestige scales on the sheet 'MG recoded'.

2.3 Grammaticality judgment

a) Construction

Okay, major headache here, of course: what grammatical features to incorporate??? This question probably makes the grammaticality judgment task the most labour-intensive one to construct. The answer, needless to say, depends on what you want to find out. If your research question is very specific as to the investigation of a particular feature (say, for example, you are looking at the attrition of Finnish case marking in an English-speaking environment), it will be relatively easy. If your approach is more inductive, it is more complicated.

It is always a good idea to base the selection of linguistic variables on previously existing studies. The ideal situation is that there is previous research available on free speech in the attrition of the language you are investigating. In that case, it is a good idea to incorporate those features that were problematic in free speech.

If there is no such literature, the next best thing is to turn to research on L1 and L2 acquisition of that language. Which features are difficult for children to learn? Which features are difficult for foreigners to learn?

Once you have selected the features for testing, we suggest that you develop your grammaticality judgments in two parts. The first consists of isolated sentences, in the second, the sentences form a coherent text (see the examples on your CD). We would suggest that you present every one of the features you are after 4 times, twice correct and twice incorrect. Make sure to include filler items!

Pre-test the GJs on at least 10 (unattrited) native speakers to make sure they elicit the corrections you want. Usually, your pre-test subjects will stumble over a couple of other things which you would never have expected to be problematic.

b) Administration

The best way of presenting the GJ, we have found, is visually and orally at the same time. In a pinch, you can achieve this by giving them a printed version and reading them out yourself. However, if you are tired (as you will be!) there is the chance that you stumble – after having presented those bloody things about 100 times myself, I had no idea any more which sentences were correct and which weren't. So the neatest way is to have a native speaker read them out, record him/her, cut the recording up into the individual sentences (you can do that e.g. with TotalRecorder) and save each sentence.

Then you can assemble the whole into a PowerPoint presentation.[2] There are some examples of all this on your CD!

If you're thinking of using reaction times, you should use the relevant software and equipment.

c) Classification of results

0 = incorrect form, judged as correct

1 = correct form, judged as incorrect

2 = incorrect form and spotted, but wrong correction

3 = incorrect form and spotted, good correction

4 = correct form identified as correct

5 = don't know

While it is useful to keep these more detailed results for later investigations, we can recode them into a two-way taxonomy:

0 = unexpected answer (0-3 plus 5 above)

1 = expected answer (3 + 4 above)

2.4 C-test L1

a) Construction

Select texts of approx. 70-100 words. Delete the **2nd half** (half the letters if the word has an even number of letters, half + ½ if it is an odd number, so "this" would become "th_____" and "which" would become "wh_____") of every **2nd word**, starting with the **2nd sentence**. Go on deleting until you have exactly 20 gaps, and then leave the rest of the text intact.

The C-test texts should be **pre-tested** on approx. 15 people (non-attrited native speakers). Select texts with a success rate of 80 to 90%, counting as 'correct' the exact word or an acceptable alternative (disregarding number of letters). We have chosen 3 texts of relatively 'normal' written style (quality newspaper, encyclopedia, etc.), 1 more informal one (glossy magazine, opinion column etc.) and 1 rather formal one (user's manual, letter from bank etc.). Please check the file 'check list Dutch C-test formal and informal.doc' for the linguistic criteria of formality. The Dutch, British English, North American English and German texts we have used can be found in Appendix 3.

---

[2] This is very easy: you make a new slide (or copy one of the files on the CD) and enter the text of the sentence. Then you choose the menu 'Insert – Films and sounds – Sound from file' and specify the relevant sound. Make sure the sound files are in the same directory as the PowerPoint presentation, so you can transfer the lot to a CD or another computer without having to re-specify the file location.

b) Administration

Present each text to the subject on a separate page and ask them to fill in as many gaps as possible. (I have found it is a good idea to tell them that it is virtually impossible to get everything right – otherwise, this can be quite a stressful and frustrating experiment for them!). Please make a note of the time each informant needs to complete each text. After a maximum of **5 minutes** per text, please stop them and ask that they proceed to the next text.

c) Classification of results

We suggest the following classification:

0 = empty

1 = incorrect lexical stem and incorrect word class

2 = incorrect lexical stem but correct word class

3 = correct lexical stem but incorrect word class

4 = correct lexical stem, correct word class, agreement error

5 = all of above correct, but still slightly wrong

6 = acceptable variant with spelling error

7 = correct word spelling error

8 = acceptable variant

9 = correct word

If you want to use a binary right/wrong taxonomy this can then easily be reccoded by counting 6-9 as correct and 1-5 as incorrect.

Please do yourself and everyone else a favour by recording NOT ONLY the appropriate number but also what the informant has actually written (in all caes but 0 and 9, where this is obvious). In order to provide for this, I have included two sheets “CTest L1 Raw data” in the file “Template for entering the results.xls”. In the first sheet, enter the score (0-9) plus the actual answer from the test, seperated by a comma. In the second sheet, delete everything from the comma onwards, so that each cell only contains a number between 0 and 9. Once you’ve entered those results into the sheet “CTest L1 Raw data 2, the sheet “C-Test L1 Recoded” will automatically calculate how many instances of each score from 0-9 each informants has.

2.5 Charlie Chaplin film re-telling task

a) Construction

Not much to do here – all you need to do is get hold of a DVD of the Charlie Chaplin film "Modern Times", from which you show a 10-minute excerpt to the informants. This excerpt starts after the scene where CC is released from prison with a letter recommending him as an honest and trustworthy man. He takes this letter to a shipyard (this is where the sequence starts, at ca. 33 minutes into the film), is accepted for work there but messes up rather badly and leaves. He walks through the city where he meets a young girl who's just stolen a loaf of bread. She is apprehended by the police, but he tries to claim it was him. However, a bystander says it was her, so he is released again. He goes into a restaurant, eats a lot of things and then says he can't pay, so he is arrested again. After a bit more to-ing and fro-ing, he's loaded into a police van, into which the girl is then also put. During an accident they manage to escape, and then walk through the suburbs. They sit down in front of a house, and CC starts fantasizing how nice it would be for them to live in a house like that. They wake up back to reality and realize they're very hungry, and there is a policeman standing behind them. They get up and walk away, which is the end of the sequence.

b) Administration

Tell the subjects that they are about to see a bit of this film (give them the background that CC has just been released from prison, but was rather reluctant to leave), and that you would like them afterwards to re-tell it to you. I've found it useful to ask them to imagine that you've never seen the film. During the re-telling, please make sure subjects cover all the major points. Sometimes, you'll have to prod their memory – "But there was also something about a cow…".

c) Classification of results

This, together with the interview on the bsis of the Sociolinguistic Questionnaire is the relatively free spoken data which you can analyse and classify and do with whatever you like. How I feel about this you can read in Schmid (fortchc.)...

2.6 Wug-test

a) Construction

Like the GJ, this test depends very much on what you are after. If you are not interested in morphology (or if you are investigating a language that has little inflectional morphology, like English), you may decide to leave it out altogether. If you do decide to

use it, there are two possible ways of deriving the nonsense words that you want to use. The first is to base them on previously existing studies (if there are any) – always a nice way to cover yourself. If there aren't any such studies, choose the most frequent lexical items that follow the inflectional pattern you are after and change the initial consonant(s). Pre-test the items you have selected by either method on at least 30 non-attrited native speakers and select for your experiments those items which elicit the most agreement.

b) Administration

If you are looking for things like gender or plural on NPs, you may want to present the items in isolation. If you are looking for more context-dependent inflections, like e.g. tense, it is probably best to embed the nonsense item in a sentence that will force informants to choose the correct tense (see the examples on the CD).

c) Classification of results

Enter the full inflection that subjects have produced. It does not make sense to classify this test in terms of 'correct' and 'incorrect'. Since this test is so language-specific, I was unable to provide you with a template. When in doubt, ask me.

2.7 Fluency in controlled association (FiCA)

a) Construction

We have found it useful to use two FiCA sessions, one that asks for elements in the category 'animals' and one in the category 'fruit and vegetables'. Depending on what hypotheses/theories concerning lexical access your work is based on, you may want to replace one by a phonological criterium, e.g. words that start with the letter 'p'. In that case, the literature recommends that you use the most frequent word-initial consonant of the language under investigation. For some languages you can find out which consonant that is by using the Celex database (http://europa.eu.int/celex/htm/celex_en.htm) – but you're on your own here, I don't know how to use it…

b) Administration

Tell the informant that they have 60 seconds to name as many elements as they can. Tape this.

c) Classification of results

For each informant, enter all elements that were produced into the sheets "FiCA[1 or2] Raw data". Make sure not to enter items that were repeated. The number of items will automatically be counted on the 'Recoded' sheet.

2.8 C-test L2

Identical to C-test L1, see 2.4 above!

2.9 Can-do scales

a) Construction

We have based our Can-Do scales on the ALTE framework of the Common European Framework of Reference, see Appendix 3 for the English version and the files on your CD for the Dutch and German ones.

b) Administration

It is useful to provide the key to the Likert-Scale on a separate sheet, since informants can get confused as to whether '1' or '5' indicates the highest proficiency.

c) Classification of results

Enter the numbers that people have indicated into the sheets "CanDo L1 Raw data" and "CanDo L2 Raw data". The mean values for all four categories (Listening, Reading, Speaking, Writing) will automatically be recalculated on the recoded sheets.

3 What do we hope the tests will tell us?

3.1 Sociolinguistic questionnaire

As indicated above, there are two reasons for the sociolinguistic questionnaire:

i) to get information about people's background – age, education, contact with L1, social networks, attitudes etc. and
ii) this forms the basis of an interview with which we hope to elicit relatively free, unmonitored speech.

a) Background information

Again, the information we collect here falls into two categories: those that will allow us to explain some of our findings, and those that will force us to qualify them. A point in case is the question about educational background. This is a variable whose impact on the results from the other tests we have to assess both **within** the experimental group and **across** experimental and control condition. It is, for example, possible (if not likely) that experimental group subjects with less school education will perform worse on the C-test (or any of the other tests) than those with higher education. However, this is not necessarily linked to the attritional process: we may find the same effect within the control condition. On the other hand, if we compare the *relative* performance of the educational groups, we may find that one group compares worse to their unattrited

counterpart than another, which would allow us to assess the impact of educational level on the attritional process. These are the factors listed under a) Personal characteristics below.

On the other hand, the factors listed under b) – f) below, such as e.g. amount of contact with the L1, can only be relevant **within** the experimental group, as we would assume all participants in the control group to have constant contact. So, the factor education has to be assessed with great caution, since while it may influence performance a great deal, it is not necessarily an explanatory or contributing factor to the phenomenon we are investigating. If, on the other hand, we can establish a correlation between amount of contact and test performance, we can probably count this factor as linked to attrition.

b) Grouping variables

The data we collect with the sociolinguistic questionnaire can be grouped (and sometimes re-calculated) into a number of different subvariables

i) Personal characteristics:
   Age (Question 1)
   Sex (Question 2)
   Emigration length (Question 7)
   Education (Question 6 a & b)

ii) Language contact
   frequency of visits (Question 18)
   frequency of use (Question 26)
   native language of partner (and possibly of an earlier partner) (Question 34)
   native language of friends (Question 29, 56)
   amount of contact with friends/family back home (Question 51)

iii) Language choice
   language of church service (Question 21)
   use of L1 with partner (Question 38 & 39)
   use of L1 with children/grandchildren (Questions 42, 43, 45, 46)
   membership in L1 clubs (Questions 60, 61)
   use of L1 media (Questions 63, 64, 65, 66)
   network questions (Questions 58 & 59)

iv) L1 proficiency self-evaluation
   proficiency now (Question 25)

proficiency at emigration (Question 24)
change in proficiency? (Question 67)
fully bilingual? (Question 71)
judgment of others (Question 72)

v) L2 proficiency self-evaluation
lessons in L2 before emigration? (Question 11)
proficiency at emigration (Question 22)
proficiency now (Question 22)

vi) Attitudes
importance of maintaining L1 (Question 27)
importance that children acquire L1 (Question 28, 47, 48, 49, 50)
cultural preference (Question 30)
language preference (Question 31)
importance of L1 as medium of contact with home (Question 54)
homesickness (Question 62)[3]
embarassment (Question 69, 70)
bothered by heavy L1 accent in L2 (Question 73)
intention to return (74)

If you've entered the data from the sociolinguistic questionnaire into the sheet "SQ Raw data", you will find the re-grouped and recalculated values on the sheet "SQ recoded". The formulae I have used to calculate the "index" variables are sometimes rather complex, I have explained them below (4.3 a).

HYPOTHESES The hypotheses we can formulate in this context are:

i) Personal characteristics:

OPEN HYPOTHESIS While we have to collect and take into account the information about age, gender, age at emigration and emigration length for obvious sociolinguistic reasons, I find it unlikely that they will influence performance on the other tasks. Equally obviously, we cannot rely on my intuitions, so we should test them on the basis of the hypothesis that they *will* have an impact on the results from the GJ, C-Test, FiCA, Can-Do-Scales and the performance in free speech. (N.B.: Since we cannot formulate a directed hypothesis, these tests have to be two-tailed, see below!)

[3] N.B.: it might be wise to split this up into 'now' and 'in the early period of emigration'!

DIRECTED HYPOTHESIS The factors for which we can predict an impact are the two educational variables: The informants with a higher level of education will do better on the linguistic tasks than those with a lower level.

ii) Language contact

DIRECTED HYPOTHESIS The informants who have more frequent contact with their L1 will do better on the linguistic tasks than those informants who use their L1 only sporadically.

iii) Language choice

DIRECTED HYPOTHESIS The informants who choose to use their L1 wherever possible will do better on the linguistic tasks than those informants who prefer L2 in situations where a choice is possible.

iv) L1 proficiency self-evaluation

DIRECTED HYPOTHESIS The informants who evaluate their L1 proficiency positively will do better on the linguistic tasks than those informants who are doubtful about their level of L1 proficiency.

v) L2 proficiency self-evaluation

OPEN HYPOTHESIS Based on the multicompetence model, it should be possible here to predict that a higher level of self-evaluated L2 proficiency would entail worse performance on the L1 tests. However, based on my obeservations I am hestitant. I think there are two groups of informants who are positive about their L2 skills: those who have totally rejected their L1, and those who feel happily confident as highly proficient L2 (and L1) speakers. So, the hypothesis here (which it is very possible we will have to reject entirely) is that self-evaluated proficiency in the L2 may be linked to the scores on the linguistic tests, but in which way we do not know.

vi) Attitudes

DIRECTED HYPOTHESIS The informants who have a positive attitude towards their L1 will do better on the linguistic tasks than those informants who are more negatively inclined.

c) Linguistic data

This, too, is data our interest in which should be twofold. On the one hand, virtually all spoken data contains violations of grammatical rules and constraints, which for the sake of brevity I shall label '**mistakes**'. These will fall into different categories – misuse of words, violations of grammatical agreement, misuse of inflections, violations of word order rules etc. Counting and classifying mistakes in the data from attriters is an

interesting excercise. However, as I have pointed out before (e.g. Schmid, forthc.) it is not without its problems.

I therefore suggest that, on top of analysing mistakes, we have to look at **proficiency**. This is expressed in lexical richness (type-token frequencies) and morphosyntactic complexity.

For each speaker, therefore, the investigation of the free spoken data elicited with this interview should yield a count of mistakes on different categories, as well as a complexity index on different categories. These we can then correlate with the results on the more formal tasks.

HYPOTHESIS The hypotheses we can formulate in this context are:

i) the experimental group will have a higher number of mistakes on the linguistic features under investigation than the control group
ii) the experimental group will use a lower number of complex morphosyntactic constructions than the control group, i.e. they will use a higher number of relatively simple constructions (more main clauses, less ubordinate clauses and embedded constructions; more nominative cases, less oblique cases; more analytic tenses, less synthetic tenses; more singular NPs, less plural NPS etc.)
iii) the hypotheses formulated under a) and b) will interact with the features of language contact, choice, attitude and self-evaluation outlined above.

3.2 Matched guise

The assumption here is that a more negative evaluation of one of the subject's languages indicates a negative attitude towards that language, and that a negative attitude towards the L1 is conducive to L1 attrition.

HYPOTHESIS Speakers with a lower rating of the L1 guises than of the L2 ones will perform worse on the linguistic tasks than those who evaluate the two guises equally or have a more positive evaluation towards the L1 than the L2.

3.3 Grammaticality judgment

The GJ is interesting in several respects. You can simply analyse it globally, dividing the structures under observation into 'correct' and 'incorrect' ones, and compare the amount of expected answers between the experimental and the control groups.

If you use this criterion, it is most interesting to look at the incorrect structures, where the hypothesis would be

HYPOTHESIS

i) the experimental group will rate more incorrect sentences as correct than the control group
ii) the experimental group will apply more inacceptable corrections to sentences they have successfully identified as incorrect, i.e. they will achieve less target-like corrections of sentences they have successfully identified as incorrect
iii) the experimental group will misjudge more correct sentences as incorrect

In the recoded version of the GJ, this simply means that the total score of the experimental group will be lower than that of the control group. The last hypothesis has to be approached with caution, since experience shows that even if subjects do rate grammatically correct sentences as unacceptable, this rating will usually not concern the feature under observation. It is therefore advisable to divide the results into sentences that were presented in the correct and in the incorrect form, and focus on the latter.

In the second instance, it can be tested whether other extralinguistic variables, e.g. from the sociolinguistic questionnaire, have an impact on these results.

Beyond this purely quantitative approach, however, I think the GJ has very interesting potential, if investigated more closely in comparison with the results from the free spoken data. Let us illustrate this by a particular example, namely the V2 structure in German. The GJ contains 8 instances of this structure:

1. 2 instances of DO-V-S-X (correct)
2. 2 instances of DO-S-V-X (incorrect)
3. 2 instances of PP-V-S-X (correct)
4. 2 instances of PP-S-V-X (incorrect)

Based on previous research, we expect the experimental group to accept more instances of 1 than of 3, and to accept more instances of 2 than of 4. However, the really interesting questions arise in connection with the free data we collect with the sociolinguistic questionnaire and the Charlie Chaplin re-telling:

HYPOTHESIS

Hypothesis a) – mistakes:: speakers who do not successfully correct the sentences under 2. and 4. will commit more mistakes of the same sentence type in their free spoken data

Hypothesis b) - proficiency: speakers who do not successfully correct the sentences under 2. and 4. will use less structures of the same sentence type in their free spoken data (i.e. they will overuse straightforward S-V-X sentences and avoid preposing elements)

I think it is vital to investigate these hypotheses for all of the linguistic features we include in the GJs. The question of what exactly it is the GJ measures has often been

posed (e.g. Altenberg & Vago 2004): after all, it is a task that is totally unrelated to what speakers normally do. By demonstrating that correlations to speakers' performance in normal speech exist (or do not exist) we can add a very valuable dimension to these questions of competence/performance issues – and one which goes beyond attrition research as such.

3.4 C-test L1

This test, too, lends itself to investigation on a purely quantitative and on a more qualitative axis. Quantitatively, we can easily assign each test subject a score out of 100, based on how many of the blanks they were able to fill in correctly. This we can use as one of our straightforward proficiency scores, together with the result of the FiCA (see below 3.7).

HYPOTHESIS The assumption of this test is that a higher score will indicate a higher level of overall proficiency. From this, a number of hypotheses can be formulated:

i) the experimental group will score lower on the C-Test than the control group
ii) the groups with higher educational levels will achieve higher scores
iii) the groups with more language contact will achiever higher scores
iv) the groups who prefer their L1 to their L2 (choice and attitude) will achieve higher scores
v) the groups who evaluate their L1 proficiency better will achieve higher scores

Quantitatively, the C-Test can also provide a great deal of insight into morphological attrition, especially in highly inflecting languages.

3.5 Charlie Chaplin film re-telling task

The free speech produced by this experiment is to be investigated according to the same principles as mentioned under 3.1 b) above

3.6 Wug-test

The hypothesis and analyses of this test are too language-specific to be discussed in detail here.

3.7 Fluency in controlled association

The assumption behind this test is that the more items a subject can name in a limited amount of time, the better his/her lexical access. Since lexical access has always been considered one of the most vulnerable features in language attrition, we would expect the most striking differences between the experimental group and the control group on this test.

HYPOTHESIS The hypotheses to be formulated here are similar to the ones formulated for the C-Test:

i) the experimental group will score lower on the FiCA than the control group
ii) the groups with higher educational levels will achieve higher scores
iii) the groups with more language contact will achiever higher scores
iv) the groups who prefer their L1 to their L2 (choice and attitude) will achieve higher scores
v) the groups who evaluate their L1 proficiency better will achieve higher scores

3.8 C-test L2

The C-Test in the L2 was introduced as a controlling feature in order to allow for the analysis of the interaction between L1 and L2 comptence under the multicompetence approach.

HYPOTHESIS *Conny, can you please fill us in here?*

3.9 Can-do scales

Self-evaluations have often been used in language attrition research as indicationsof actual proficiency. However, these were usually of the more general type we elicit with the sociolinguistic questionnaire (indicated in the L1 self evaluation score under 3.1 a) above). The Can-Do scales we use here allow subjects a more differentiated self-assessment of a variety of skills and over a range of difficulty, in two productive and two receptive domains: speaking, writing, listening, reading. We propose to calculate an average score for each of those domains.

HYPOTHESIS
i) The scores on the writing scale will correlate with the results of the C-Test in both L1 and L2.
ii) The scores on the speaking scale will correlate with the results of the FiCA and with the results of the free spoken data.
iii) The scores on the receptive skills will correlate with the results from the linguistic tests.

4 How can we analyse our data statistically?

To begin with, there are two fundamental concepts you have to keep in mind when you are approaching your data. The first is the difference between *independent* and *dependent* variables, and the second concerns measurement scales.

4.1 Independent vs. dependent variables

If you consider the data described above, you will see that we are taking into account both *characteristics* and *performance* of our subjects. The former – essentially those factors represented in the data described under 3.1 a) above - are a given, which we are merely documenting, the latter we are trying to elicit by means of our tests. We will find variablity in both the *independent* variables, like sex, age, education, contact and so on and in the *dependent* variables, like the scores on the C-Test or the FiCA. The question we are asking is in what way the independent variables can account for and explain the variability we observe in the dependent variables.

An example: recall that the highest possible score on the C-Test is 100. If we make 50 potential attriters fill out this C-Test, we may find that the lowest score anyone has achieved is, say, 51; while the highest is 97. What we want to find out is to what degree the different independent variables we are investigating can account for that variability. It is possible that women do better on this test than men, that people with a higher educational level achieve better scores, or that older people find it more difficult. What the statistics can tell us, if we ask the questions in the right way, is *how these factors combine* to account for our findings.

4.2 Measurement scales

When looking at our data, we must always keep in mind that there are different kinds of measurements, which will allow us to do different things: *nominal*, *ordinal* and *interval* data.

Nominal data

We refer to data as nominal when they merely represent a label, such as 1=male, 2=female. We may encounter categories which we have to differentiate, but which are not logically ordered in ay way – the fact that the ‘label’ for the females has twice the numerical value of the label for the males here has no significance whatsoever, and we could equally well have switched them around or called men 13 and women 91. So, nominal data are values which serve to identify and differentiate categories, but do not imply any ranking order.

Ordinal

Like nominal data, ordinal data are also data which divide our sample into groups, but here we do assume a ranking. One example of this is the proposed classification scheme for the C-Test (s. 2.4 c) above), where we assume that the achieved result is better the higher the achieved score. A further example is the classification into educational levels in the sociolinguistic questionnaire.

However, while in ordinal data a ranking is implied, we cannot say that the distance between values is equal. Therefore, in the C-Test, it would not make sense to say that a '4' is twice as good as a '2'. This implies that with ordinal data, it is impossible to calculate sums, averages and the like.

Interval

Interval data are, so to speak, numbers proper, they represent measurements such as frequencies, time, length, weight etc. With interval data, it *is* possible to say that a figure that is mathematically twice as large as another also represents double that – 4 lbs are twice as much as 2 lbs, and someone who correctly answered 8 questions did twice as well as someone who only answered 4.

N.B. A slightly questionable case where measurement scales are concerned are the Likert-scales, where people are asked to indicate preferences, agreements and the like on a scale from (e.g.) 1 to 5. Very strictly speaking, these should be regarded as ordinal, since we cannot be absolutely certain that the distance between 1 and 2 will be the same as that between 2 and 3 and so on. However, it is general practice in statistics to ignore this difficulty and to treat such measurements as interval.

4.3 Types of data in our test battery:

1.a) Sociolinguistic questionnaire (independent variables) – Raw data

	Qu.	Variable	Values	# of levels
Nom.	2	Sex	male/female	2
	4	Nationality	L1, L2, both, other	4
	5	Standard/Dialect		2
	33	Family situation	marr., div., wid., partner, single	5
	34	Native lg. of partner	L1, L2, other	3
	48	Children L1 instruction	yes/no	2
	52	Medium of contact with home	Phone, Letters, Email, visits	4
	56	Language of most acquaintances	L1, L2, other	3
	57	Where did you meet most friends	L1 club, common friends, work/school, other	4
	60	Member of L1 club	yes/no	2
	61	Member of L1 club	yes/no	2
	62	Ever homesick?	yes/no	2
	63	L1 music	yes/no	2
	64	L1 TV	yes/no	2
	65	L1 radio	yes/no	2
	66	L1 newspapers/books	yes/no	2
	68	Has L1 proficiency changed?	yes/no	2
	70	Ever feel uncomfortable speaking L1 at home	yes/no	2
	71	Ever feel uncomfortable speaking L1 here	yes/no	2
	74	Does L1 accent bother you?	yes/no	2
	75	Ever want to go back	yes/no/don't know	3
	77	Was decision right	yes/no	2

Ord.	6a	School education	(Haupt., Real, Abitur)	3
	6b	Voacational tr.	none, apprenticeship, university	3
	18	Frequency of visits	never/seldom/frequently	5
	20	church	never/sometimes/regularly	3
	21	lg. of service	L2/L1/other language	3
Int.	1	year of birth		
	7	year of emigration		
	68	L2 before emigration	Likert-Scale 1-5	5
	69	L2 now	Likert-Scale 1-5	5
	72	L1 before emigration	Likert-Scale 1-5	5
	73	L1 now	Likert-Scale 1-5	5
	22	Frequency use L1	Likert-Scale 1-5	5
	23	Import. maint. L1	Likert-Scale 1-5	5
	24	Imp. children L1	Likert-Scale 1-5	5
	25	Acquaintances L2-L1	Likert-Scale 1-5	5
	26	Sense of belonging	Likert-Scale 1-5	5
	27	Preferred language	Likert-Scale 1-3	3
	28	L1 or L2 to partner	Likert-Scale 1-5	5
	29	L1or L2 from partner	Likert-Scale 1-5	5
	30	L1 or L2 to children	Likert-Scale 1-5	5
	31	L1 or L2 from children	Likert-Scale 1-5	5
	38	L1 or L2 to grandch.	Likert-Scale 1-5	5
	39	L1 or L2 from grandch.	Likert-Scale 1-5	5
	42	Admonish children	Likert-Scale 1-3	3
	43	Correct children's L1	Likert-Scale 1-5	5
	45	Regret children's lack of L1 proficiency	Likert-Scale 1-5	5
	46	Contact with family/friends back home	Likert-Scale 1-5	5
	47	L1 or L2 in contact with home	Likert-Scale 1-5	5
	49	L1 important medium for contact with home	Likert-Scale 1-5	5
	50	L1 with family	Likert-Scale 1-5	5
	51	L1 with friends	Likert-Scale 1-5	5
	53	L1 with pets	Likert-Scale 1-5	5
	54	L1 at work	Likert-Scale 1-5	5
	56a	L1 in church	Likert-Scale 1-5	5
	56b	L1 shopping	Likert-Scale 1-5	5
	56c	L1 in clubs	Likert-Scale 1-5	5
	56d	L2 with family	Likert-Scale 1-5	5
	56e	L2 with friends	Likert-Scale 1-5	5
	56f	L2 with pets	Likert-Scale 1-5	5
	56g	L2 at work	Likert-Scale 1-5	5
	56h	L2 in church	Likert-Scale 1-5	5
	56i	L2 shopping	Likert-Scale 1-5	5
	56j	L2 in clubs	Likert-Scale 1-5	5
	56k	Has L1 proficiency changed	worse/no change/better	3
	56l	Has L1 use changed	less/no change/more	3
	56m	consider yourself bilingual	no, L1 better/yes/no, L2 better/don't know	4
	56n	better idea of others who speak L1 or L2	L1 better/no difference/ L2 better/don't know	4
	Free spoken data	- errors per 1,000 words - TTF - morphosyntactic variables		

1. b) Sociolinguistic questionnaire (independent variables) – Recoded

a) Personal characteristics		Levels	Range
Nom	Sex (1=male, 2=female)	2	
Ord.	School education	3	

	Professional education	3	
Int.	Age Age at emigration Emigration length		32-70 > 17 > 15
b) Language contact			
Int.	1. Frequency of visits		1-5
	2. Use of L1		1-5
	3. Native language of friends		1-5
	4. Amount of contact with home		1-5
	5. L1 at work		1-5
	6. % L1 in networks		%
Contact index: $\frac{(1.+2.+3.+4.+5.)+(6./20)}{6}$			$1 \leq Contact_i \leq 5$
	Native language of current (last) partner Native language of earlier partner	1=L1, 0=L2 or other	
c) Language choice			
Ord.	1. Language of church service (1=L1, 0=L2), if applicable	2 (0,1)	
	2. Membership in clubs (0=none, 1 = now or earlier, 2=now and earlier)	3 (0,1,2)	
	3. Use of L1 media (0=none, 4=all four)	5 (0-4)	
Int.	4. L1 in church, if applicable		0-5
	5. L1 with partner now * native language of current partner4		0-5
	6. L1 with partner earlier * native language of earlier partner		0-5
	7. L1 with children now, if applicable		0-5
	8. L1 with children earlier, if applicable		0-5
	9. L1 with (grand)children now, if applicable		0-5
	10. L1 with (grand)children earlier, if applicable		0-5
	11. L1 with family		0-5
	12. L1 with friends		0-5
	13. L1 with pets, if applicable		0-5
	14. L1 in clubs, if applicable		0-5
Choice index: $\frac{(4.+5.+6.+7.+8.+9.+10.+11.+12.+13.+14.)}{\text{Number of arguments (4.-14.) >0}} + \frac{1.+2.+3.}{7}$			$0 \leq Choice_i \leq 5$
d) Language attitude			
	1. importance of maintaining L1	1-5	

[4]If the native language of the current or earlier partner is not the informant's L1, this value will be 0, since in that case, we assume that the informant does not have the choice to use the L1 with the partner.

	2. importance that children acquire L1, if applicable	0-5	
	3. cultural preference	1-5	
	4. language preference	3 (1, 2.5, 5)	
	5. importance of L1 for contact with home	1-5	
	6. homesickness	2 (0,1)	
	7. bothered by heavy L1 accent	2 (0,1)	
	8. intention to return	3 (0,2)	
	9. was decision right?	3 (0,2)	
Attitude index: $\left(\frac{(1.+2.+3.+4.+5)}{\text{Number of arguments (1.-5.)}>0} + \frac{(6.+7.+8.+9)}{6*5} \right) :2$			$1 \leq \text{Attitude}_i \leq 5$

2. a)Matched guise (Raw data)

Int.	Evaluation of two guises of the same speaker (L1 and L2 speaking) on 13 characteristics, 2 speakers and 1 filler voice = 5*13 = 65 variables per subject	65 * Likert-Scale 1-5.

2. b) Matched guise (Recoded)

Int.	Evaluation of two guises of the same speaker (L1 and L2 speaking) on a solidarity and prestige scale = 8 variables	$1 \leq \text{Sp1L1Solidarity}_i \leq 5$ $1 \leq \text{Sp1L1Prestige}_i \leq 5$ $1 \leq \text{Sp1L2Solidarity}_i \leq 5$ $1 \leq \text{Sp1L2Prestige}_i \leq 5$ $1 \leq \text{Sp2L1Solidarity}_i \leq 5$ $1 \leq \text{Sp2L1Prestige}_i \leq 5$ $1 \leq \text{Sp2L2Solidarity}_i \leq 5$ $1 \leq \text{Sp2L2Prestige}_i \leq 5$

3. Grammaticality judgment

Interval	Proportion of correct answers

4. C-test L1

Interval	Score out of 100

5. Charlie Chaplin film re-telling task

Interval	Linguistic data	- errors per 1,000 words - TTF - morphosyntactic variables

6. Wug-test (in the opposite order, i.e. VP first)

Nominal	inflections on particular items

7. Fluency in controlled association (FiCA 1 + FiCA2)

Interval	Number of items produced within a certain time

8. C-test L2

Interval	Score out of 100

9. a) Can-do scales – Raw data

Interval	Self-evaluation on 43 determinants in both L1 and L2	43 * Likert-Scale 1-5

9. b) Can-do scales – Recoded

Interval	Mean value on listening skills Mean value on reading skills Mean value on speaking skills Mean value on writing skills	$1 \leq Listening_i \leq 5$ $1 \leq Reading_i \leq 5$ $1 \leq Speaking_i \leq 5$ $1 \leq Writing_i \leq 5$

5 The statistics

5.1 Independent variables I - Sociolinguistic questionnaire

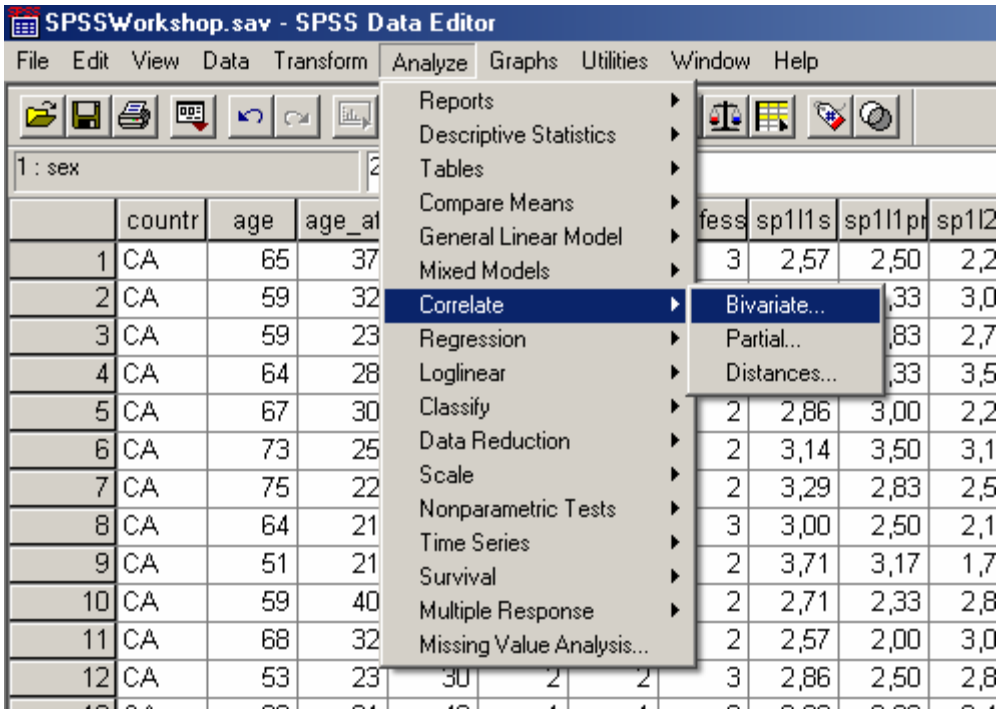
As said above (3.1) the hypotheses concerning the SQ can be divided into **within** and **between** subjects hypotheses. If we are dealing with within-subjects hypotheses, the independent variables can be all types of measurement – nominal (e.g. sex), ordinal (e.g. education) or interval (e.g. attitude index). Nominal and ordinal independent variables divide our sample up into discrete groups – e.g. men and women – and we can compare the results of the two or more groups. With interval data, such as age, no such discrete groups can be formed. For all hypotheses from the within-subjects part of this test, our aim is to determine if there is a statistical association between the independent variable (e.g. "emigration length") and the dependent variable (e.g. the score on the C-Test).

Where between-subjects hypotheses are concerned, i.e. where we compare the experimental and the control group, the independent variable is always nominal, and we are thus dealing with two groups.

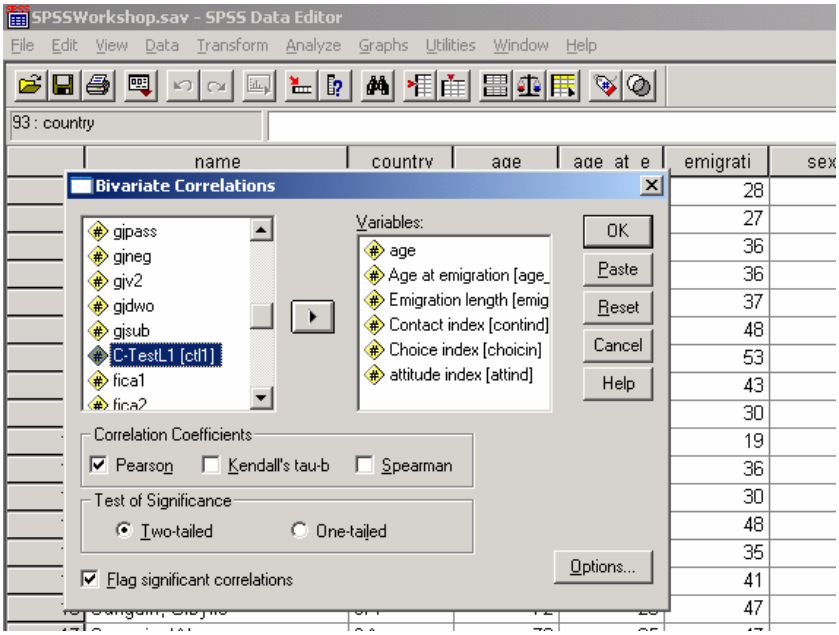
a) Interval data I: Pearson correlation

If the independent variables we are dealing with are interval data, one way of testing this statistical association is the "**Pearson correlation**", which allows us to test all correlations we are interested in at the same time.

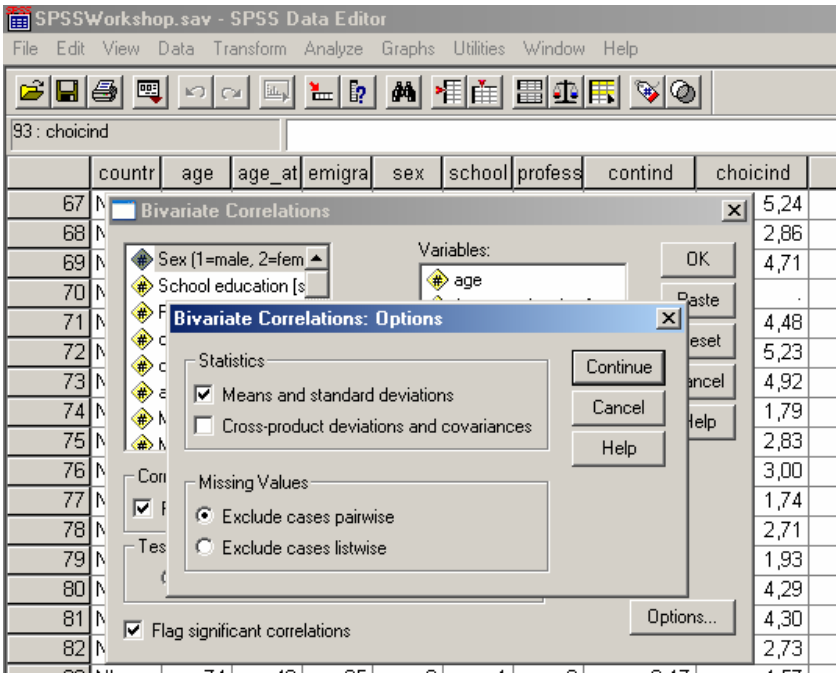
In SPSS, you choose the menu Analyze => Correlation => Bivariate



In the popup menu that follows, you can send anything you want from your list of variables on the left (this lists all the variables that are present in your data set) to the window on the right (this lists the variables you want to correlate). In the example in Fig. 2, we have already selected Age, Age at emigration, and Emigration length, Contact index, Choice index and Attitude index, and are about to select the independent variable we want to correlate this with, e.g. C-Test L1. (As was said above, you can do as many variables in one test as you like. I used only a small number so I would be able to fit the correlation table on the page.)



At this stage, it is usually handy to take a look at “Options”. In our case, we this offers us the following:



We can see here that it is also possible to get some descriptive statistics, which may be handy. So we tick that box, Click ‘Continue’ and in the next window click ‘Okay’.

This takes us to the output window, where SPSS calculates the results from the tests. The first thing we see is the promised descriptive overview:

Descriptive Statistics

	Mean	Std. Deviation	N
AGE	63,93	10,185	92
Age at emigration	27,27	7,165	92
Emigration length	36,66	11,670	92
Contact index	2,9597	1,10292	92
Choice index	3,2686	1,19144	92
attitude index	2,3788	,61407	92
C-TestL1	78,1739	12,15614	92

And then comes the good bit, the actual correlation:

Correlations

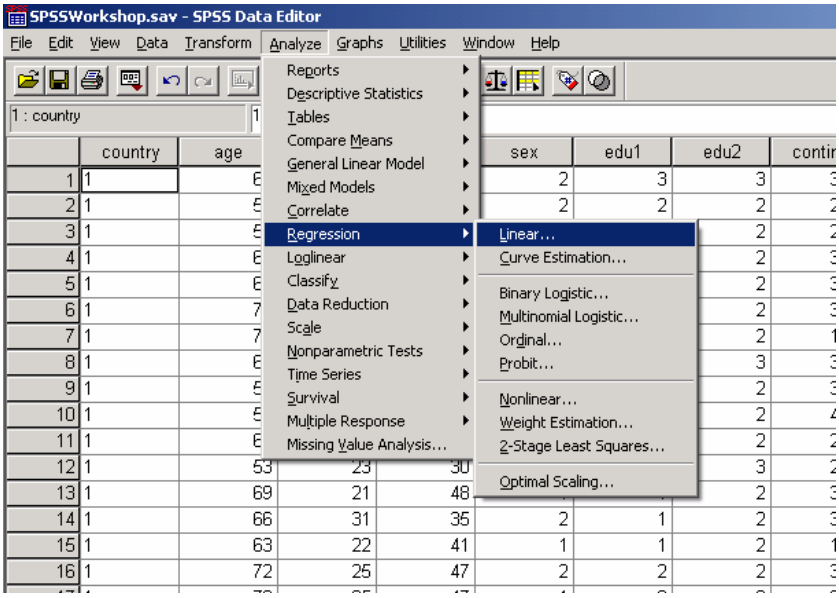
		AGE	Age at emigration	Emigration length	Contact index	Choice index	attitude index	C-TestL1
AGE	Pearson Correlation	1	,129	,793(**)	-,015	-,116	-,093	-,295(**)
	Sig. (2-tailed)	.	,219	,000	,887	,269	,379	,004
	N	92	92	92	92	92	92	92
Age at emigration	Pearson Correlation	,129	1	-,501(**)	,275(**)	,401(**)	,037	,105
	Sig. (2-tailed)	,219	.	,000	,008	,000	,725	,318
	N	92	92	92	92	92	92	92
Emigration length	Pearson Correlation	,793(**)	-,501(**)	1	-,182	-,348(**)	-,104	-,322(**)
	Sig. (2-tailed)	,000	,000	.	,082	,001	,324	,002
	N	92	92	92	92	92	92	92
Contact index	Pearson Correlation	-,015	,275(**)	-,182	1	,443(**)	,256(*)	,157
	Sig. (2-tailed)	,887	,008	,082	.	,000	,014	,135
	N	92	92	92	92	92	92	92
Choice index	Pearson Correlation	-,116	,401(**)	-,348(**)	,443(**)	1	,432(**)	,167
	Sig. (2-tailed)	,269	,000	,001	,000	.	,000	,111
	N	92	92	92	92	92	92	92
attitude index	Pearson Correlation	-,093	,037	-,104	,256(*)	,432(**)	1	,068
	Sig. (2-tailed)	,379	,725	,324	,014	,000	.	,521
	N	92	92	92	92	92	92	92
C-TestL1	Pearson Correlation	-,295(**)	,105	-,322(**)	,157	,167	,068	1
	Sig. (2-tailed)	,004	,318	,002	,135	,111	,521	.
	N	92	92	92	92	92	92	92

** Correlation is significant at the 0.01 level (2-tailed).

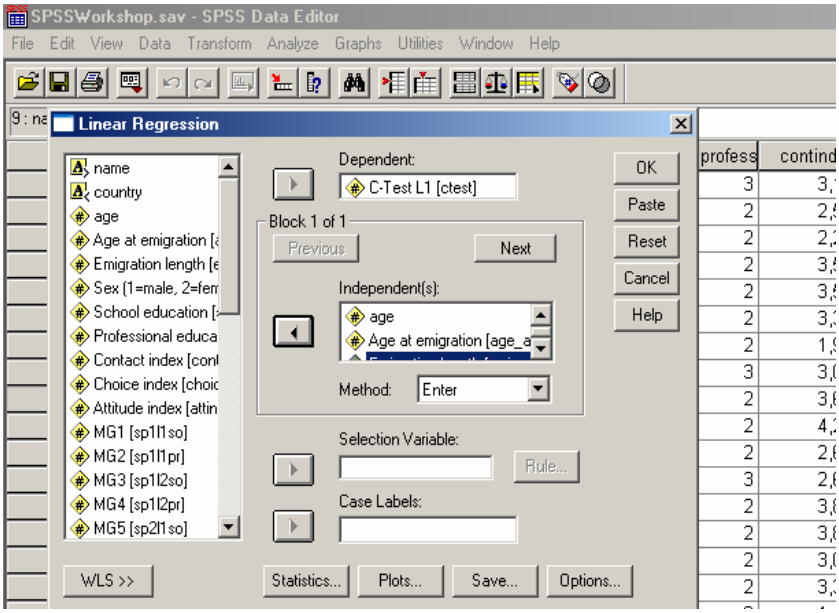
* Correlation is significant at the 0.05 level (2-tailed).

i Interval data II: Linear regression

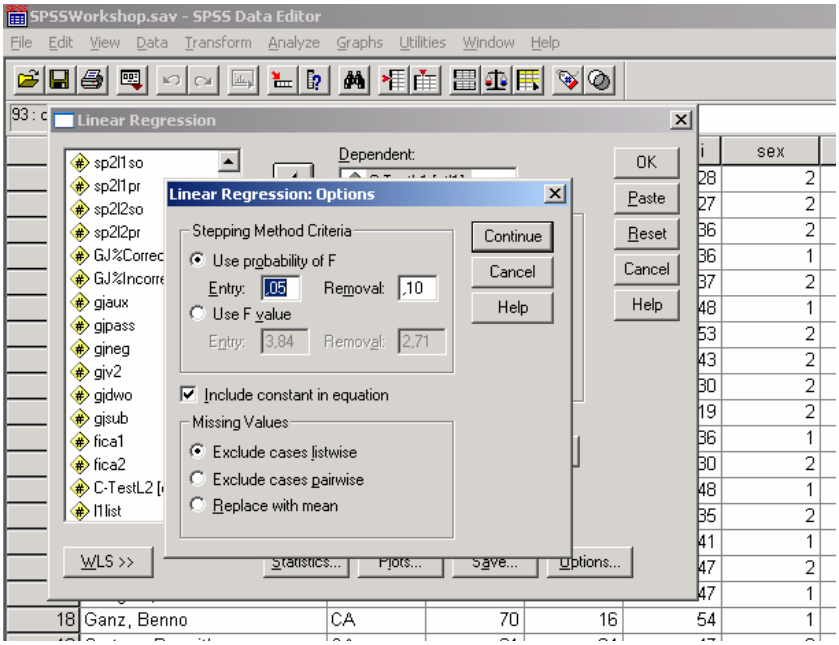
In the above table, we can see which correlations are significant. However, this test does not tell us anything about how the variables interact. This we can test by means of a multiple regression analysis for the independent variables which are measured on an interval scale. For this, we choose the menu Analyze => Regression => Linear



Like in the correlation analysis, we can select Age, Age at emigration, Emigration length, Choice, Contact and Attitude index etc. as our independent variables. This test can only be done for one dependent variable, e.g. the C-Test L1, at a time.



Again, it may be useful to take a look at “Options”:



After the descriptive statistics, the first result is given in the following table:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,348(a)	,121	,070	11,68754

a Predictors: (Constant), attitude index, Age at emigration, AGE, Contact index, Choice index

This is a very important figure: the R Square (in the case of one independent variable) or the Adjusted R Square (in the case of more than one independent variables) tells us how much of the variation within the dependent variable can be explained by all independent variables together. In this case, the variables age, age at emigration, emigration length, choic, contact and attitude index together account for only 7% of the variation in our data. In the result from the ANOVA, we can see that this very small proportion nevertheless is significant:

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1613,470	5	322,694	2,362	,047(a)
	Residual	11747,487	86	136,599		
	Total	13360,957	91			

a Predictors: (Constant), attitude index, Age at emigration, AGE, Contact index, Choice index
b Dependent Variable: C-TestL1

In the following table, each of the factors is assessed individually for significance, and we can see that Age is the only independent variable with a significant impact (the column Beta tells us how strong the variable is in this context).

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	92,203	10,213		9,028	,000
	AGE	-,359	,124	-,301	-2,905	,005
	Age at emigration	,159	,195	,094	,819	,415
	Contact index	1,177	1,257	,107	,936	,352
	Choice index	,543	1,343	,053	,404	,687
	attitude index	-,277	2,259	-,014	-,122	,903

a Dependent Variable: C-TestL1

Our selection has also produced an interesting effect: without really telling us why, the program chose to eliminate “Emigration length” as an independent variable:

Excluded Variables(b)

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	Emigration length	.(a)	.	.	.	,000

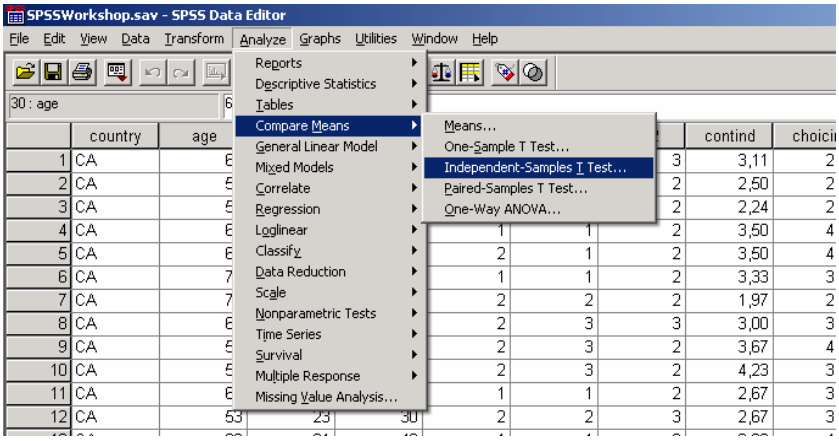
a Predictors in the Model: (Constant), attitude index, Age at emigration, AGE, Contact index, Choice index
b Dependent Variable: C-TestL1

The reason for this is that the regression analysis does not like independent variables that influence each other too much. If you think about it, of course, Age at emigration will always equal Age – Emigration length; and this connection between the independent variables has led to the elimination of one of the variables.

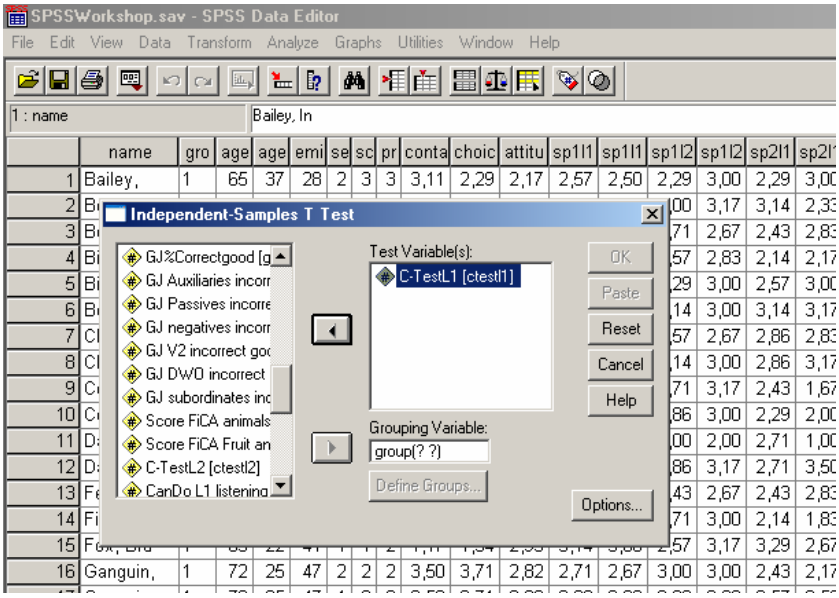
b) Ordinal data: T-Test and ANOVA

If the independent variable whose impact you want to assess is ordinal or nominal, you can use an ANOVA (if there are more than 2 groups, e.g. “education”) or a T-Test (if there are only two groups, e.g. men and women or experimental and control group).

To find out whether interval data results on the linguistic tests differ between your experimental and your control group, you perform an “Independent samples T-Test”[5], in the menu “Analyze => Compare Means = Independent-Samples T Test”

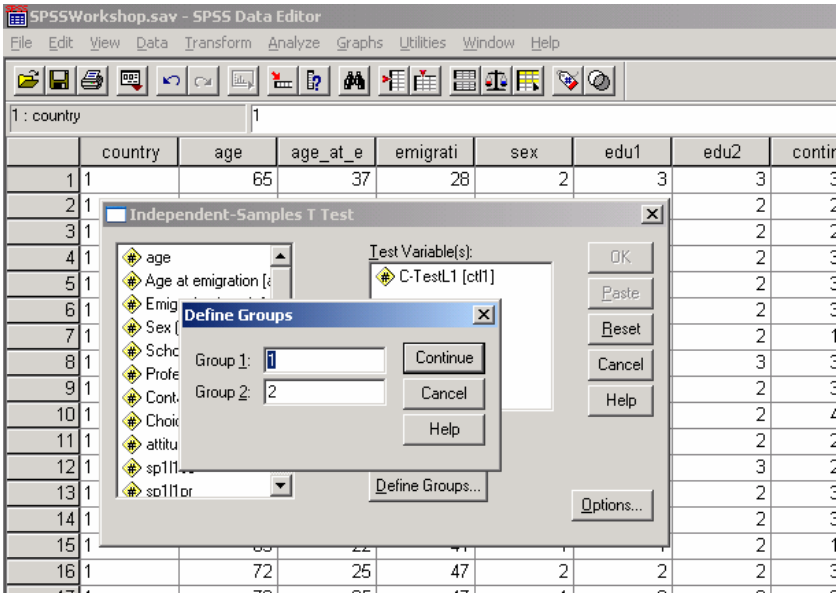


We could perform this test for any dependent variable on an interval scale, for example “number of mistakes per 1,000 words”, “number of subordinate clauses” and so on. Since I do not yet have those data available, I’ll again use the results from the C-Test.



Note that in this case, the independent variable is called “grouping variable”, since we do not correlate two interval variables on a sliding scale here but want to assess if the results between the two groups – experimental and control – are different from each other. Click on the grouping variable to make the option “Define groups” accessible, and then click on “Define groups”.

[5] “independent samples” in this context means you are not using two measurements from the same group, e.g. before and after a training session, when you would perform a “Paired samples T-test”.



In my data, the experimental group is labelled “1” and the control group is labelled “2”, so this is what I enter here (you can, of course, use any number you like, but it does have to be a numerical value!). Click “Continue” and then “OK”.

The descriptive statistics show that the Mean for the control group (NL) is slightly higher than that for the experimental group (CA).

Group Statistics

	1=CA, 2=NL	N	Mean	Std. Deviation	Std. Error Mean
C-TestL1	CA	54	77,4630	11,14520	1,51667
	NL	38	79,1842	13,55441	2,19881

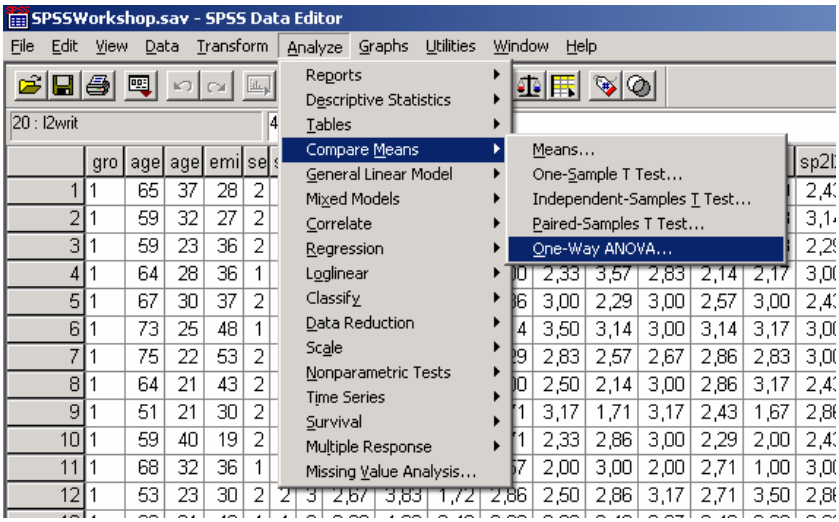
However, the results from the actual T-Test show that this difference is not significant.

Independent Samples Test

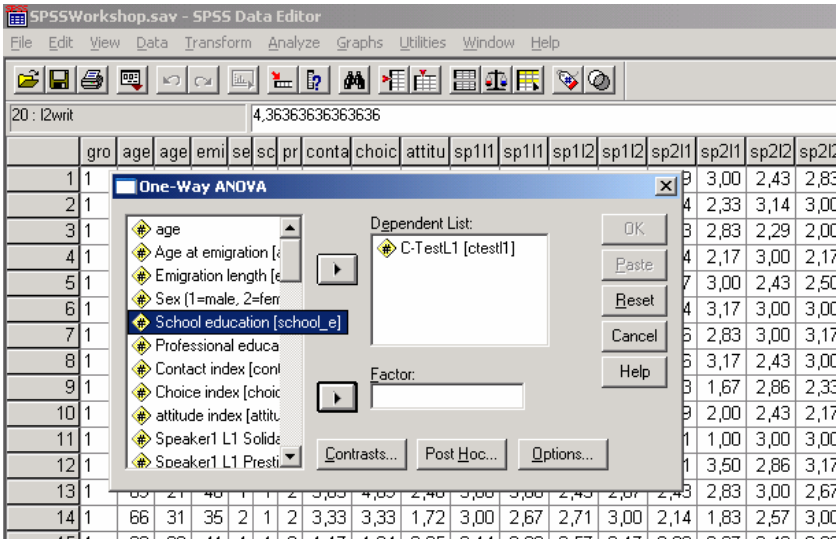
		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
C-TestL1	Equal variances assumed	1,067	,304	-,667	90	,507	-1,7212	2,58185	-6,85054	3,40804
	Equal variances not assumed			-,644	69,586	,521	-1,7212	2,67116	-7,04926	3,60676

In other words, the differences that we observe are accidental – which means that there is no difference between the performance of the two groups.

In the case of a variable that has more than two values, e.g. education, we use the One-Way ANOVA.

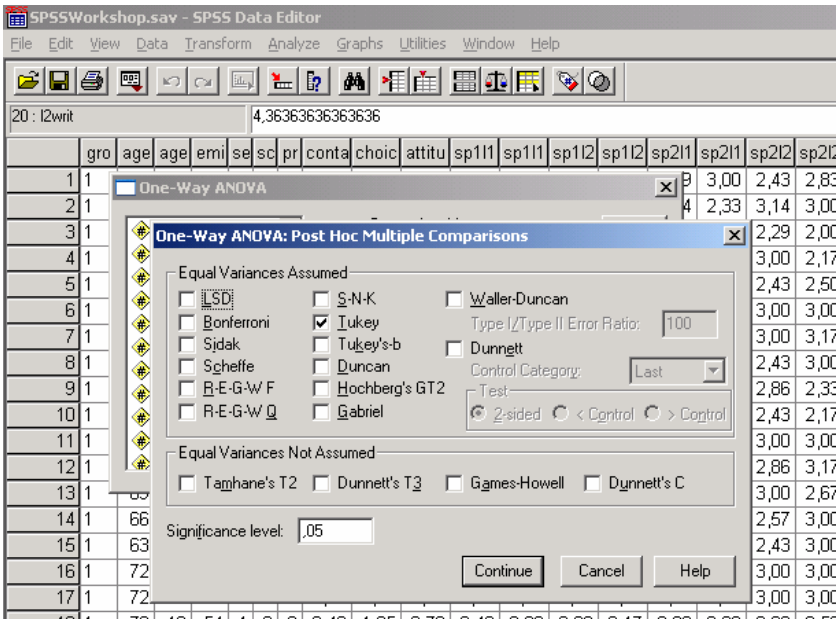


We can assess any number of dependent variables at the same time. Our independent, or grouping, variable we choose as the 'Factor'



If you want to perform an ANOVA, an important option is the **Post Hoc** test. Since the ANOVA assesses more than 2 groups, you will not only want to know whether the overall difference is significant, but also which groups are actually different from each other – it may be that groups 1 and 2 behave in exactly the same way and are only different from group 3. You can find this out by clicking on Post Hoc and selecting one (or more) of the options. They are all very similar – I prefer Tukey simply because Cor Koster, a colleague and

statistics teacher I am very fond of, recommended it, while Jules likes Bonferroni. Make your choice and click on ‘Continue’, then click on ‘Okay’.



The result from the ANOVA tells us that the overall difference between the three groups is significant:

ANOVA

C-TestL1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3561,188	2	1780,594	16,030	,000
Within Groups	9886,030	89	111,079		
Total	13447,217	91			

However, the Tukey Post Hoc shows that only the difference between group 3 on the one hand, and groups 1 and 2 on the other, is significant.

Multiple Comparisons

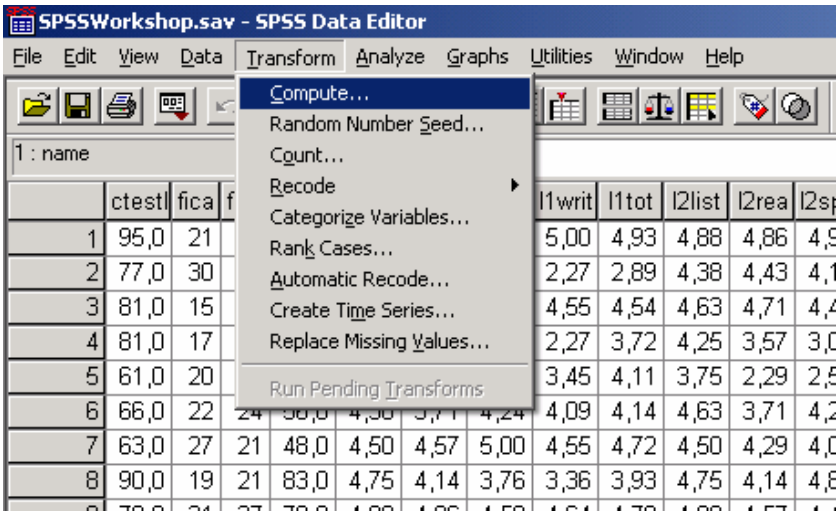
Dependent Variable: C-TestL1
Tukey HSD

(I) School education	(J) School education	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-2,4786	2,95426	,680	-9,5202	4,5630
	3	-14,1257(*)	2,92508	,000	-21,0977	-7,1536
2	1	2,4786	2,95426	,680	-4,5630	9,5202
	3	-11,6471(*)	2,48512	,000	-17,5705	-5,7237
3	1	14,1257(*)	2,92508	,000	7,1536	21,0977
	2	11,6471(*)	2,48512	,000	5,7237	17,5705

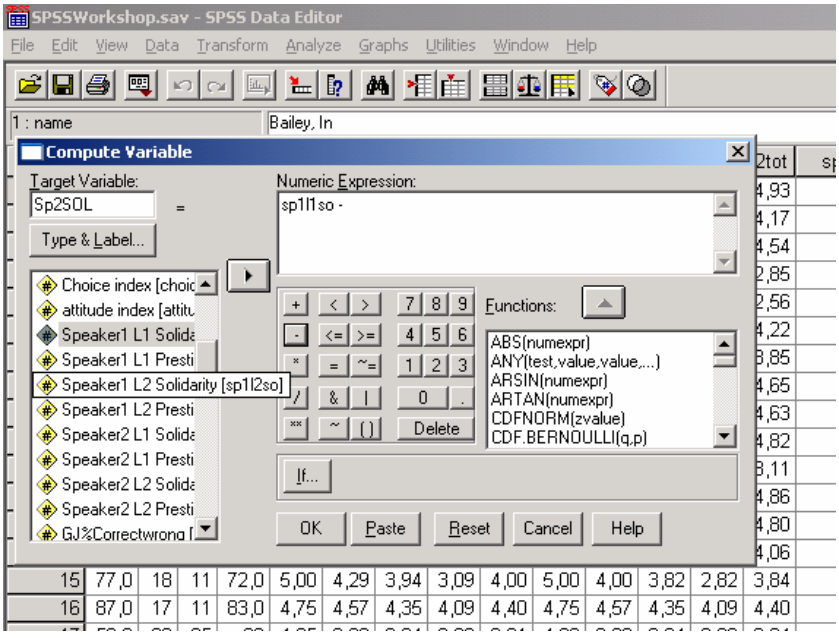
* The mean difference is significant at the .05 level.

5.2 Matched guise

In order to test whether there is a link between the differences in rating the two speakers in their two languages and the results from the linguistic tests, we first have to have SPSS calculate the differences between the two measurements. To do this, we choose “Compute” in the menu “Transform”.



In the next window, we have to give a name to the new variable we want to calculate, e.g. SP1SOL, and to specify that we want this variable to be “Speaker 1 L1 Solidarity” – “Speaker 1 L2 Solidarity”.



This step we repeat, until we have 4 new variables. These we can then correate with e.g. the results from the C-Test, or include in a Linear Regression (see above)

5.3 Dependent variables

The results from the other tests are all interval data which we can include in the above tests by means of the T-Test, the ANOVA or the Linear Regression.

References:

Altenberg, Evelyn/Robert Vago. 2004. “Grammaticality judgments in first language attrition”, , in: Monika S. Schmid/Barbara Köpke/Merel Keijzer/Lina Weilemar (eds). *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*. Amsterdam/Philadelphia: John Benjamins, pp. 105-129.

Köpke, Barbara & Monika S. Schmid. 2004. “First language attrition: the next phase”, in: Monika S. Schmid/Barbara Köpke/Merel Keijzer/Lina Weilemar (eds). *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*. Amsterdam/Philadelphia: John Benjamins, pp. 1-45.

Schmid, Monika S. 2004. “A new blueprint for language attrition research”, in: Monika S. Schmid/Barbara Köpke/Merel Keijzer/Lina Weilemar (eds). *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*. Amsterdam/Philadelphia: John Benjamins, pp. 349-363.

Schmid, Monika S. forthc. “Measuring language attrition in free speech: what’s in a mistake?”, to appear in *International Journal of Bilingualism* 5 (3).

6

Appendix 1: Sociolinguistic questionnaire – Version Dutch speakers in Canada

1) When were you born?……19……..
2) Gender ❑ male ❑ female
3) Where were you born? Place…… Region……
   Country:……
   ……
4) What is your nationality? ❑ CA…❑ CA & NL ❑ NL.
5) Would you say that you spoke a standard variety of Dutch while you lived in the Netherlands or a dialect? ❑ Standard Dutch ❑ Dialect, ……
6) What is the highest level of education you have completed?
   ❑ primary school ❑ secondary school, level ❑ higher education, namely:
   ❑ no vocational training. ❑ apprenticeship ❑university
7) When did you come to Canada (year)? 19……
8) Why did you emigrate and why to Canada in particular?
   ❑ Job ❑ Job of partner ❑ partner ❑ other,:.....
9) Apart from Canada, have you ever lived in a country other than the Netherlands for a longer period of time (that is, more than 6 months)? ❑ none, ❑ less than 1 year, ❑1 year or more, (Ort).....in (Land) .....
10) What language(s) did you acquire before starting school?
    ❑Dutch, ❑Dutch and other, ❑other),
11) Did you attend any English classes before coming to Canada? ❑ yes, ❑ no
    ❑ less than 1 month, ❑ less than 3 months, ❑ less than 6 months, ❑ less than 1 year,
    ❑ more than 1 year
12) Have you pursued further education while living in Canada (this does not have to be language-related) ? ❑ yes, ❑ no ……
13) What language or languages did you learn professionally or at school?
    …….
14) What language or languages did you learn outside of an educational environment (so outside of school or work)?
    …….

15) What is your current profession? If you are retired, could you please indicate your last profession before retirement?

16) If you have had several professions, could you indicate each one of them in chronological order?

1……..from……...……..to……..

2……..from……...……..to……..

3……...…...from……...……..to……..

4……..from……...……..to……..

17) Have you ever attended Dutch heritage classes while living in Canada? ❑ yes, ❑ no, how long:…… hours per week……

18) Have you ever been back to the Netherlands since leaving for Canada? ❑ never, ❑ seldom, ❑ regularly, 1-2 times, ❑ regularly 3-5 times, ❑ regularly, > 5 times

19) If you have indicated that you have been back to the Netherlands, could you please indicate what the reason or reasons for such a visit were (you may tick more than one box here)?
❑ important family event; ❑ visit without particular reason, ❑ other

20) Do you ever go to church in Canada?
1 = never, ❑ sometimes, ❑ regularly

21) If you have indicated you go to church, could you please indicate in which language the services are held?
❑ EN, ❑ NL, ❑ EN & NL, ❑ other

22) In general, how would you rate your English language proficiency before you moved to Canada?
❑ none, ❑ very bad, ❑ bad, ❑ sufficient, ❑ good, ❑ vg

23) In general, how would you rate your English language proficiency at present?
❑ none, ❑ very bad, ❑ bad, ❑ sufficient, ❑ good, ❑ vg

24) In general, how would you rate your Dutch language proficiency before you moved to Canada?
❑ none, ❑ very bad, ❑ bad, ❑ sufficient, ❑ good, ❑ vg

25) In general, how would you rate your Dutch language proficiency at present?
❑ none, ❑ very bad, ❑ bad, ❑ sufficient, ❑ good, ❑ vg

26) How often do you speak Dutch?
❑ rarely, ❑ few times a year, ❑ monthly, ❑ weeky, ❑ daily

27) Do you consider it important to maintain your Dutch?
❑ unimportant, ❑ relatively unimportant, ❑ not very important, ❑ important, ❑ vi

28) Do you consider it important that your children can speak and understand Dutch?
❑ unimportant, ❑ relatively unimportant, ❑ not very important, ❑ important, ❑ vi

29) In general, do you have more Dutch- or English-speaking friends in Canada?
❑ only Canad., ❑ more Canad., ❑ equal, ❑ more Dutch, ❑ only Dutch

30) Do you feel more at home with Dutch or with Canadian culture?
❑ only Canad., ❑ more Canad., ❑ equal, ❑ more Dutch, ❑ only Dutch

31) Do you feel more comfortable speaking Dutch or English?
❑ English, ❑ Dutch, ❑ no difference

32) Could you elaborate on your answer: why do you feel more comfortable speaking either Dutch or English or why don't you have any preference?

33) What is your current marital status?
❑ married, ❑ divorced, ❑ widowed, ❑ with partner, ❑ single

34) With what language(s) was your (ex)partner brought up?
❑ NL, ❑ EN, ❑ other

35) If your (ex)partner was not born in Canada, what were the reasons that he or she came to Canada? ❑ Job ❑ Job of partner ❑ partner ❑ other,:
.....

36) When did your (ex)partner come to Canada (year)?

37) How did you meet?❑ NL, ❑ CA, ❑ other

38) What language or languages do you mostly use when talking to your (ex)partner?
❑ only English, ❑ more English, ❑ equal, ❑ more Dutch, ❑ only Dutch, ❑ other or n.a.

39) What language or languages does your (ex)partner mostly use when talking to you?
❑ only English, ❑ more English, ❑ equal, ❑ more Dutch, ❑ only Dutch, ❑ other or n.a.

40) What is the current profession of your (ex)partner? Falls er oder sie pensioniert ist, welchen Beruf hat er oder sie vor der Pensionierung ausgeübt?

41) Do you have children? ❑no, ❑ yes, number: ……
their names are .....
..... and they are……..……..years old.

42) What language or languages do you mostly use when talking to your children?
❑ only English, ❑ more English, ❑ equal, ❑ more Dutch, ❑ only Dutch, ❑ other or n.a.

43) What language or languages do your children mostly use when talking to you?
❑ only English, ❑ more English, ❑ equal, ❑ more Dutch, ❑ only Dutch, ❑ other or n.a.

44) Do you have grandchildren? ❑no, ❑ yes, number: ……
their names are .....
..... and they are……..……..years old.

45) What language or languages do you mostly use when talking to your grandchildren?
❑ only English, ❑ more English, ❑ equal, ❑ more Dutch, ❑ only Dutch, ❑ other or n.a.

46) What language or languages do your grandchildren mostly use when talking to you?
❑ only English, ❑ more English, ❑ equal, ❑ more Dutch, ❑ only Dutch, ❑ other or n.a.

47) Do you encourage your children to speak Dutch? ❑ never, ❑ sometimes, ❑ often

48) Did your children ever follow Dutch heritage classes (Saturday classes for example)? ❑ yes, ❑ no

49) Did /do you ever correct your children's Dutch?
❑ never, ❑ seldom, ❑ sometimes, ❑ often, ❑ very often

50) If your children do not speak or understand Dutch, do you regret that?
❑ not at all, ❑ no, ❑ don't care, ❑ a bit, ❑ very, ❑ n.a.

51) Are you in frequent contact with relatives and friends in the Netherlands?
❑ never, ❑ seldom, ❑ sometimes, ❑ often, ❑ very often

52) How do you keep in touch with those relatives and friends in the Netherlands?
❑ phone, ❑ letters, ❑ email, ❑ other

53) What language or languages do you mostly use to keep in touch with relatives and friends in the Netherlands?
❑ only English, ❑ more English, ❑ equal, ❑ more Dutch, ❑ only Dutch, ❑ other or n.a.

54) Do you think Dutch plays an important role in the relationship between your direct family members?
❑ not at all, ❑ no, ❑ probably, ❑ a bit, ❑ very, ❑ n.a.

55) Have you made many new friends in Canada? ❑ yes, ❑ no

56) What is the mother tongue of the majority of these people?
❑ EN, ❑ NL, ❑ equal, ❑ other

57) How did you meet most of these people?
❑ Dutch club, ❑ common friends, ❑ work or school, ❑ other……..

58) Could you please name those people that you are most frequently in touch with in the following table? These people can live in the Netherlands or in Canada. I want to see through this table which language you most frequently use in your daily life: Dutch or English. You don't have to give the name of the person if you prefer not to. I would like to ask you, however, to provide the rest of the information asked for.

Name (optional)	**Does this person live in Canada or the Netherlands?**	**What language(s) do you use when communicating with each other?**	**How did you meet this person?**	**How long have you known this person?**	**What is your relationship with this person?**

59) Could you, in the following tables, please indicate to what extent you use Dutch (table 1) and English (table 2) in the domains provided? You may simply tick the box. If a certain domain is not applicable to you (for example, if you don't have any pets), you may leave the box empty.

I speak Dutch					
	all the time	frequently	sometimes	rarely	very rarely
With relatives					
With friends					
To pets					
At work					
In church					
In shops					
At clubs or organisations					

I speak English					
	all the time	frequently	sometimes	rarely	very rarely
With relatives					
With friends					
To pets					
At work					
In church					
In shops					

At clubs or organisations					

60) Have you ever been a member of a Dutch club or organisation in Canada? ❑ yes, ❑ no……..

61) Are you now a member of a Dutch club or organisation in Canada? ❑ yes, ❑ no……..

62) Do you ever get homesick in the sense of missing the Netherlands? ❑ yes, yes, what I then miss most is/are …….❑ no

63) Do you ever listen to Dutch songs? ❑ yes, ❑ no

64) Do you ever listen to Dutch radio programmes? ❑ yes, ❑ no, ❑ would like to but can't

65) Do you ever read Dutch newspapers, books or magazines? ❑ yes, ❑ no, ❑ would like to but can't

66) Do you ever watch Dutch television programmes? ❑ yes, ❑ no

67) If you have indicated that you never listen to Dutch songs or radio programmes, nor read Dutch newspapers, books or magazines and that you don't watch Dutch television programmes, could you indicate why you think that is?

68) Do you think your Dutch language proficiency has changed since you moved to Canada? ❑ yes, worse, ❑ no, ❑ yes, better

69) Do you think you use more or less Dutch since you moved to Canada? ❑ yes, less, ❑ no, ❑ yes, more

70) Do you ever feel uncomfortable when speaking Dutch with a Dutch person who has never spent a considerable amount of time in an English-speaking country? ❑ yes, ❑ no

71) If you ever do feel uncomfortable in such a situation, could you indicate whether this is also the case when you speak Dutch with someone who, like you, has lived in Canada for a long time? ❑ yes, ❑ no

72) Do you see yourself as bilingual? ❑ no, English better, ❑ yes, ❑ no, Dutch better, ❑ don't know
…….

73) Are you better at guessing a person's social position/status when they speak Dutch or English? ❑ EN, ❑ equal, ❑ NL, ❑ don't know
…….

74) How do you feel about Dutch people (tourists for example) who speak English with a heavy Dutch accent? ❑ no, ❑ yes

75) Do you ever intend to move back to the Netherlands? ❑ yes, ❑ don't know, ❑ no

76) Reason……
……

77) Looking back, do you think you have made the right decision in moving to Canada?
❑ yes, ❑ no, ❑ don't know

77) You have come to the end of this questionnaire. Is there anything you would like to add? This can be anything from language-related comments to remarks about the questionnaire or research itself.

7 Appendix 2: Matched Guise instructions

Visualizing a person on the basis of his/her voice

This is a language game. You are about to listen to five different speakers, some of whom speak English while others speak Dutch. Every speaker tells the same story, which can be found on the last page of this booklet. After each speaker the playback will be stopped. You can then fill out a questionnaire.

The first part of the questionnaire, part A, consists of a list containing opposite character traits. Please indicate which of the two opposite character traits you find more applicable to the speaker by ticking one of the boxes. For each speaker, please work through the whole list.

The second part, part B, offers you the possibility to add another character trait to this list in case you think it applies to the speaker but is not mentioned in part A.

Part C asks you to place the voice in terms of regional accent.

Finally, part D asks you to indicate whether you think this speaker has a high social status or a low social status.

The intention of this game is to visualize the person on the basis of the voice that is heard. If you would like to hear the voice once more, it will be played back again. If you have any questions, please do not hesitate to ask.

Please turn the page to start the test

Speaker 1

Part A						
The speaker is:						
sociable	☐	☐	☐	☐	☐	aloof
unintelligent	☐	☐	☐	☐	☐	intelligent
a good leader	☐	☐	☐	☐	☐	obedient
pleasant	☐	☐	☐	☐	☐	unpleasant
rough	☐	☐	☐	☐	☐	decent
factual	☐	☐	☐	☐	☐	approachable
unattractive	☐	☐	☐	☐	☐	attractive
awkward	☐	☐	☐	☐	☐	graceful
polite	☐	☐	☐	☐	☐	impolite
ambitious	☐	☐	☐	☐	☐	easy to please
intolerant	☐	☐	☐	☐	☐	tolerant
educated	☐	☐	☐	☐	☐	uneducated

Part B
Is / are there any other character trait(s) that you find applicable to this speaker?

Character trait 1:
Character trait 2:

Part C
Could you place this speaker in terms of regional accent?

Part D		
What social status do you think this speaker holds?		
High social status	☐ ☐ ☐ ☐ ☐	Low social status

Text to read:

DUTCH

Ergens moet nog een foto liggen waarop mijn oma voor de nieuwe auto van mijn vader staat. Waarschijnlijk rond 1974 gemaakt. Mijn vader was trots op zijn net verworven bezit, een Triumph geloof ik, en wilde dat luister bijzetten door zijn schoonmoeder te portretteren op de motorkap van dat ding. Hij was niet de enige die mijn grootmoeder zag als model. Onlangs onthulde mijn oom Jos, starend naar een geschilderd portret van mijn oma, dat hij dolgraag een keer met haar uit was gegaan. De enige die dat voorrecht echt heeft gehad, was haar grote liefde en mijn grootvader. Ze ontmoette hem op dansles. Een lange, mooie man die door zijn vrienden met recht werd omschreven als een heer. Eentje met veel humor en tot op hoge leeftijd interessant voor vrouwen van alle leeftijden.

ENGLISH

There should be a photo somewhere with my grandmother standing in front of my father's new car, probably taken around 1974. My father was proud of his newly acquired possession, I think it was a Triumph, and he wanted to emphasize that by portraying his mother-in-law on the hood of the thing. He wasn't the only one who saw my grandmother as a model. Recently, my uncle Tommy revealed, staring at a painting of grandma, that he would have loved to have gone out with her. The only one who actually had that privilege was her lifetime love, my grandfather. She met him at a dancing class, a tall, handsome man who was correctly described as a gentleman by his friends. He was a man of great humour and up to an advanced age remained attractive to women of all ages.

GERMAN

Irgendwo muss ich noch ein Foto von meiner Oma haben, auf dem sie vor dem neuen Auto von meinem Vater steht, so um 1974 herum aufgenommen. Mein Vater war stolz auf seine Neuerrungenschaft – es war glaub ich ein BMW, und das wollte er dann zeigen, indem er seine Schwiegermutter auf der Motorhaube verewigte. Er war übrigens nicht der einzige, der in meiner Grossmutter ein Modell sah. Mein Onkel Tommy hat sich neulich noch ein Bild von Oma angesehen und dann erzählt, wie gerne er mit ihr ausgegangen wäre. Der einzige, dem das dann vergönnt war, war ihre grosse Liebe, mein Grossvater. Sie hat ihn in der Tanzstunde kennengelernt – ein grosser, gutaussehender Mann, den seine Freunde zu Recht als einen echten Herrn bezeichneten. Er war ein Mensch mit viel Humor, der bis ins hohe Alter bei Frauen aller Altersgruppen Interesse weckte.

8 Appendix 3: C-Test texts

I. North American English

1.

We all live with other people's expectations of us. These are a refl_____ of th_____ trying to under_____ us; the_____ are predic_____ of wh_____ they th_____ we will think, d_____ and feel . Gene_____, we acc_____ the sta_____ quo, but these expec_____ can be ha_____ to han_____ when they co_____ from our fami_____ and can be diff_____ to ign_____, especially wh_____ they come from our par_____.

2.

Founded in 1878 by Bisshop Isaac Hellmuth and the Anglican Diocese of Huron as "The Western University of London Ontario", Western is one of Canada's oldest and best universities. The fi_____ students grad_____ in ar_____ and medi_____ in 1883. To_____, The University of Western Ontario is a vib_____ centre of lear_____ with 1,164 fac_____ members and alm_____ 29,000 underg_____ and graduate stud_____. Through i_____ 12 Facu_____ and Sch_____, and three affi_____ Colleges, the University off_____ more th_____ 60 diffe_____ degree and dip_____ programs to London's comm_____.

3.

The BBC's core purpose is broadcasting. Since the lau_____of Radio Times in 1923 it h_____ also eng_____ in comme_____ activities. If pur_____ properly, su_____ commercial activities he_____ to rea_____ the va_____ of lic_____ payers' ass_____ and gene_____ income to be plou_____ back in_____ the public ser_____ programming. The_____ commercial Policy Guidelines s_____ out the fram_____ which ens_____ that the BBC's commercial activities supp_____ its public purpose.

4.

The decision to remove soft drinks from elementary and junior high school vending machines is a step in the right direction to help children make better choices when it comes to what they eat and drink. Childhood obe_____ has bec_____ a ser_____ problem in th_____ country a_____ children cons_____ more sugar-based fo_____ and sp_____ less ti_____ getting the nece_____ exercise. Many par_____ have quest_____ schools' deci_____ to al_____ vending machines which disp_____ candy and so_____ drinks. Many schools, th_____, have co_____ to re_____ on the mo_____ these machines generate through agreements with the companies which makes soft drinks and junk food.

5.

In the last federal election, 61% of eligible voters cast a ballot. That's a fright_____ lack of inte_____ by the elect_____, but is not_____ compared to the turn_____ in provi_____ and munic_____ elections, which s_____ even lo_____ turnouts. It's diff_____ to bel_____ there's so lit_____ interest in elections. In Canada, we're fort_____ to have pol_____ stations wi_____ a short wa_____ or dr_____. There are volun_____ more th_____ willing to pro_____ rides to someone unable to walk or who doesn't have a car.

Missing words:

1. Reflection
2. Them
3. Understand
4. They
5. Predictions
6. What
7. Think
8. Do
9. Generally
10. Accept
11. Status
12. Expectations
13. Hard
14. Handle
15. Come
16. Family
17. Difficult
18. Ignore
19. When
20. parents
21. First
22. Graduated
23. Arts
24. Medicine
25. Today
26. Vibrant
27. Learning
28. Faculty
29. Almost
30. Undergraduate
31. Students
32. Its
33. Faculties
34. Schools
35. Affiliated
36. Offers
37. Than
38. Different
39. Diploma
40. community
41. Launch
42. Has
43. Engaged
44. Commercial
45. Pursued
46. Such
47. Help
48. Realize
49. Value
50. Licence
51. Assets
52. Generate
53. Ploughed
54. Into
55. Service
56. These
57. Set
58. Framework
59. Ensures
60. Support
61. Obesity
62. Become
63. Serious
64. This
65. As
66. Consume
67. Food(s)
68. Spend
69. Time
70. Necessary
71. Parents
72. Questioned
73. Decisions
74. Allow
75. Dispense
76. Soft
77. Though
78. Come
79. Rely
80. money
81. Frightening
82. Interest
83. Electorate
84. Nothing
85. Turnouts
86. Provincial
87. Municipal
88. See
89. Lower

90. Difficult	93. Fortunate	96. Walk	99. Than
91. Believe	94. Polling	97. Drive	100. provide
92. Little	95. Within	98. Volunteers	

2. British English

1:

We all live with other people's expectations of us. These are a refle_____ of th_____ trying to under_____ us; th_____ are predic_____ of wh_____ they th_____ we will think, d_____ and feel. Gene_____ we acc_____ the sta_____ quo, but these expec_____ can be ha_____ to han_____ when they co_____ from our fami_____ and can be diff_____ to ign_____ , especially wh_____ they come from our par_____ .

2:

The decision to remove soft drinks from elementary and junior high school vending machines is a step in the right direction to helping children make better choices when it comes to what they eat and drink. Childhood obe_____ has bec_____ a ser_____ problem in th_____ country a_____ children cons_____ more sugar-based fo_____ and sp_____ less ti_____ getting the nece_____ exercise. Many par_____ have quest_____ schools' deci_____ to al_____ vending machines which disp_____ candy and so_____ drinks. Many schoold, tho_____ , have co_____ to re_____ on the mo_____ these machines generate through agreements with the companies which makes soft drinks and junk food.

3:

Two former US navy ships contaminated with chemicals were expected to arrive in the English Channel last night. The Maritime and Coastguard Agency sa_____ the ves_____ , at the cen_____ of an enviro_____ row, we_____ being to_____ through the cha_____ , before hea_____ up the east co_____ to Hartlepool.

Pl_____ to dism_____ them in north-east England have been she_____ after being dee_____ to fl_____ international ru_____. Last we_____ , the gover_____ said the ships could be sto_____ in Hartlepool before go_____ back acr_____ the Atlantic.

4:

Don't get me wrong. I love magazines. I've been addicted to them since my teenage years. There's some_____ about wom_____ magazine superfi_____ that I of_____ enjoy. But oh b_____ , they are ju_____ so, so frustr_____ predictable. I rec_____ you co_____ cobble o_____ together very eas_____ in five min_____. Take the co_____ for example: the cover im_____ : get a he_____ and shou_____ shot of a smi_____ , heavily make-uped and airbr_____ model (or optio_____ a fam_____ person).

5:

In the last Canadian federal election, 61 per cent of eligible voters cast a ballot. That's a fright_____ lack of inte_____ by the elect_____ , but is not_____ compared to the turn_____ in provi_____ and munic_____ elections, which s_____ even lo_____ turnouts. It's diff_____ to bel_____ there's so lit_____ interest in elections. In Canada, we're fort_____ to have pol_____ stations wit_____ a short wa_____ or dr_____ . There are volun_____ more th_____ willing to pro_____ rides to someone unable to walk or who doesn't have a car.

III. Dutch

1:

Ik houd van Nederland en niet zo'n beetje ook. Waarom ik van het land houd is niet alleen omdat velen van wie ik houd hier leven, nee, het is me_____ dan d_____. De groo_____ reden v_____ mijn lie_____ voor het land ko_____ voort u_____ het feit dat al_____ zo geor_____ en syste_____ is. Er i_____ een systeem e_____ het wer_____. Je kan, ni_____ zonder twi_____, maar to_____ met dic_____ ogen er_____ uitgaan d_____ het recht zege_____.

2:

Als je reist, heb je de kans om te zijn wie je wilt zijn óf degene die je echt bent. Dat komt om_____ niemand een ste_____ op je dr_____. Toen ik n_____ het rei_____ in Nederland teru_____, werd ik hele_____ gek. A_____ na vier dagen. A_____ ik z_____ dat men_____ zich opwo_____ over een honde_____ op de st_____, werd ik pan_____. Dan da_____ ik, mens, waar ma_____ je je dr_____ over? Ik ben gel_____ naar de psycholoog ges_____, want ik trok dat echt niet.

3:

Openlijke narcisten zijn mensen met een opgeblazen gevoel over zichzelf. Ze ei_____ vaak ande_____ aandacht o_____ en ko_____ charmant ov_____, ond_____ het feit d_____ ze wei_____ besef he_____ van de beho_____ van anderen. Verb_____ narcisten zijn weli_____ net z_____ hevig met zichzelf be_____ en ev_____ arrogant a_____ openlijke narcisten,

ma_____ ze do_____ dit o_____ een subti_____ manier.

4:

Het internationaal perscentrum Nieuwspoort discussieert weer eens over de code. De Haagse soci_____ waar h_____ journaille en de poli_____ in een onged_____ samenzijn verk_____, hanteert si_____ jaar en d_____ de ongesc_____ regel d_____ wat er t_____ plekke gez_____ wordt ni_____ naar bui_____ mag ko_____. Alt_____: niet her_____ mag worden tot de betre_____ persoon en pla_____. Voorzitter van het bestuur van Nieuwspoort Max de Bok maa_____ onlangs pla_____ voor Casper Becx, maar het beleid bleef ongewijzigd.

5:

Prins Claus was een intellectuele gentleman. Voor zijn echtgenote koningin werd had hij een serieuze baan in de ontwikkelingshulp. Na 1980 voe_____ hij zich ste_____ meer een orna_____ van de tr_____. Hij raa_____ depressief, ge_____, maar we_____ nooit meer de ou_____. De la_____ van een onve_____ bestaan a_____ prins-gemaal le_____ op Claus een gr_____druk, g_____ hem het gev_____ een ha_____marionet te zijn, een man die acht_____ veel lie_____ een zelfst_____ positie had gehad dan een afge_____.

Missing words:

1.	meer	4.	van	7.	uit	10.	systematisch	13.	werkt
2.	dat	5.	liefde	8.	alles	11.	is	14.	niet
3.	grootste	6.	komt	9.	georganiseerd	12.	en	15.	twijfel

16. toch
17. dichte
18. ervan
19. dat
20. zegeviert
21. omdat
22. stempel
23. drukt
24. na
25. reisje
26. terugkwam
27. helemaal
28. al
29. als
30. zag
31. mensen
32. opwonden
33. hondendrol
34. stoep
35. panisch
36. dacht
37. maak
38. druk
39. gelijk
40. gestapt
41. eisen
42. andermans
43. op
44. komen
45. over
46. ondanks
47. dat
48. weinig
49. hebben
50. behoeften
51. verborgen
52. weliswaar
53. zo
54. bezig
55. even
56. als
57. maar
58. doen
59. op
60. subtielere
61. sociëteit
62. het
63. politiek
64. ongedwongen
65. verkeren
66. sinds
67. dag
68. ongeschreven
69. dat
70. ter
71. gezegd
72. niet
73. buiten
74. komen
75. althans
76. herleid
77. betreffende
78. plaats
79. maakte
80. plaats
81. voelde
82. steeds
83. ornament
84. troon
85. raakte
86. genas
87. werd
88. oude
89. last
90. onvervuld
91. als
92. legde
93. grote
94. gaf
95. gevoel
96. halve
97. achteraf
98. liever
99. zelfstandige
100. afgeleide

IV: German

1:

Die Geschichte der Kernspaltung reicht zurück in das frühe 19. Jahrhundert. In d_____ Folgejahren leg_____ Chemiker d_____ Grundstein f_____ den mode_____ Atombegriff. S_____ erkannten, da_____ die chemi_____ Elemente a_____ Teilchen aufg_____ sind, d_____ untereinander völ_____ gleichartig reag_____ , sich jed_____ von and_____ Elementen unters_____ . 1871 erschien d_____ erste tabell_____ Aufstellung d_____ Eigenschaften al_____ bekannten Elem_____ , das Periode_____ .

2:

Neben den regulären Teilen der Bibel gibt es, wie die Kenner wissen, allerlei apokryphe Schriften, darunter auch die etwas andere Schöpfungsgeschichte, die „Pseudo-Genesis“. Sie

wei_____ in ein_____ Details
erhe_____ von d_____ gängigen
Ver_____ ab, insbes_____ im ers_____
Kapitel, d_____ damit beg_____ , dass
d_____ „Chöre“, al_____ die En_____ ,
beim Erze_____ Michael zusamme_____ und
üb_____ ein rätsel_____ Phänomen
ber_____ .

3:
Eine Wünschelrute ist ein gegabelter Zweig, ursprünglich meist vom Haselnussstrauch, später verwe_____ man au_____ ähnliche
Instr_____ aus untersch_____ Materialien.
S_____ dient d_____ so gena_____
Rutengänger, ei_____ Person, d_____ für
si_____ eine beso_____ Begabung
bean_____ t, als Hilfs_____ zum
Auff_____ von unterir_____ »Reizzonen«,
z_____ Beispiel Wasse_____ , Erdölvorkommen
od_____ Erzlagerstätten.

4:
Bedienungsanleitung bitte vollständig vor Inbetriebnahme des Bügeleisens durchlesen und aufbewahren.
Reparaturen an Elektro_____ dürfen n_____ von Fachk_____ durchgeführt wer_____ Durch unsach_____ Reparaturen kön_____ erhebliche Gefa_____ für d_____ Benutzer entst_____ Wird d_____ Gerät zwecken_____ oder fal_____ bedient, ka_____ keine Haf_____ für dad_____ verursachte Sch_____ übernommen wer_____ .

Das Ge_____ wurde v_____ uns sicherheitstechnisch geprüft.

5:

Schon in ältester Zeit haben die Menschen den Himmel beobachtet. Je stä_____ frühe Kult_____ von d_____ Natur abhä_____ waren, de_____ näher l_____ es f_____ sie, a_____ den o_____ periodischen Ersche_____ der Na_____ und d_____ Sternenhimmels besti_____ Faktoren abzul_____ , die i_____ tägliches Le_____ beeinflussten. I_____ Verlauf d_____ Entwicklung d_____ mensch_____ Zivilisation verl_____ diese natür_____ Zyklen im_____ mehr a_____ Bedeutung.

Missing words:

1. den
2. legten
3. den
4. für
5. modernen
6. Sie
7. dass
8. chemischen
9. aus
10. aufgebaut
11. die
12. völlig
13. reagieren,
14. jedoch
15. anderen
16. unterscheiden.
17. die
18. tabellarische
19. der
20. aller
21. Elemente,
22. Periodensystem.
23. weicht
24. einigen
25. erheblich
26. der
27. Version
28. insbesondere
29. ersten
30. das
31. beginnt,
32. die
33. also
34. Engel,
35. Erzengel
36. zusammenkommen
37. über
38. rätselhaftes
39. beraten.
40. verwendete
41. auch
42. Instrumente
43. unterschiedlichen
44. Sie
45. dem
46. genannten
47. einer
48. die
49. sich
50. besondere
51. beansprucht,
52. Hilfsmittel
53. Auffinden
54. unterirdischen
55. zum
56. Wasseradern,
57. oder
58. Elektrogeraeten
59. nur
60. Fachkraeften
61. werden.
62. unsachgemaesse
63. können
64. Gefahren
65. den
66. entstehen
67. das
68. zweckentfremdet
69. falsch
70. kann
71. Haftung
72. dadurch
73. Schaeden
74. werden
75. Geraet
76. von
77. stärker
78. Kulturen
79. der
80. abhängig
81. desto
82. lag
83. für
84. aus
85. oft
86. Erscheinungen
87. Natur
88. des
89. bestimmende
90. abzuleiten,
91. ihr
92. Leben
93. Im
94. der
95. der
96. menschlichen
97. verloren
98. natürlichen
99. immer
100. an

9 Appendix 4: Can-do scales (English)

Listed below are a number of “can-do” scales. They consist of statement about your language proficiency in both Dutch and English. What I am interested in is how well or bad you perceive your current language proficiency in both languages to be. Please read each description carefully and circle the appropriate number to indicate whether, at the present time, you would be able to carry out each task in each language. Thus, you can only circle one number per language and per statement. Please use the following scale:

1 = I cannot do this at all
2 = I can do this, but with much difficulty
3 = I can do this, although with some difficulty
4 = I can do this fairly easily
5 = I can do this without any difficulty at all

An example:

When I am in a noisy place, such as a bar or pub, I can still understand and participate in a conversation.

Dutch 1 2 3 4 5 English 1 2 3 4 5

	Listening comprehension	**Dutch**	**English**
1.	I can understand most TV news and current affairs programmes.	1 2 3 4 5	1 2 3 4 5
2.	I can understand the main points of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.	1 2 3 4 5	1 2 3 4 5
3.	I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided that I have some time to get familiar with the accent.	1 2 3 4 5	1 2 3 4 5
4.	I can understand extended speech even when it is not clearly structured and when relationships are	1 2 3 4 5	1 2 3 4 5

	only implied and not signalled explicitly.		
5.	I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc.	1 2 3 4 5	1 2 3 4 5
6.	I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar.	1 2 3 4 5	1 2 3 4 5
7.	I can understand the majority of films in standard dialects.	1 2 3 4 5	1 2 3 4 5
8.	I can understand television programmes and films without too much effort.	1 2 3 4 5	1 2 3 4 5
	Reading proficiency	**Dutch**	**English**
9.	I can understand long and complex factual and literary texts, appreciating distinctions of style.	1 2 3 4 5	1 2 3 4 5
10.	I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints.	1 2 3 4 5	1 2 3 4 5
11.	I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.	1 2 3 4 5	1 2 3 4 5
12.	I can understand the description of events, feelings and wishes in personal letters.	1 2 3 4 5	1 2 3 4 5
13.	I can understand texts that consist mainly of high frequency everyday or job-related language.	1 2 3 4 5	1 2 3 4 5

		Dutch	**English**
14.	I can understand specialised articles and longer technical instructions, even when they do not relate	1 2 3 4 5	1 2 3 4 5

	to my field.		
15.	I can understand contemporary literary prose.	1 2 3 4 5	1 2 3 4 5
	Speaking ability	**Dutch**	**English**
16.	I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible.	1 2 3 4 5	1 2 3 4 5
17.	I can present a clear, smoothly flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points.	1 2 3 4 5	1 2 3 4 5
18.	I can use language flexibly and effectively for social and professional purposes.	1 2 3 4 5	1 2 3 4 5
19.	I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel, current events).	1 2 3 4 5	1 2 3 4 5
20.	I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms.	1 2 3 4 5	1 2 3 4 5
21.	I can narrate a story or relate the plot of a book or film and describe my reactions.	1 2 3 4 5	1 2 3 4 5
22.	I can deal with most situations likely to arise whilst travelling in an area where the language is spoken.	1 2 3 4 5	1 2 3 4 5
23.	I can take an active part in discussion in familiar contexts, accounting for and sustaining my views.	1 2 3 4 5	1 2 3 4 5
24.	I can present clear, detailed descriptions of complex subjects integrating sub-themes, developing particular points and rounding off with an	1 2 3 4 5	1 2 3 4 5

	appropriate conclusion.		
25.	If I do have a problem I can backtrack and restructure around the difficulty so smoothly that other people are hardly aware of it.	1 2 3 4 5	1 2 3 4 5
26.	I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions.	1 2 3 4 5	1 2 3 4 5
27.	I can present clear, detailed descriptions on a wide range of subjects related to my field of interest.	1 2 3 4 5	1 2 3 4 5
		Dutch	**English**
28.	I can express myself fluently and convey finer shades of meaning precisely.	1 2 3 4 5	1 2 3 4 5
29.	I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	1 2 3 4 5	1 2 3 4 5
30.	I can briefly give reasons and explanations for opinions and plans.	1 2 3 4 5	1 2 3 4 5
31.	I can express myself fluently and spontaneously without much obvious searching for expressions.	1 2 3 4 5	1 2 3 4 5
32.	I can formulate ideas and opinions with precision and relate my contribution skillfully to those of other speakers.	1 2 3 4 5	1 2 3 4 5
	Writing proficiency	**Dutch**	**English**
33.	I can select style appropriate to the reader in mind.	1 2 3 4 5	1 2 3 4 5
34.	I can write simple connected text on topics which are familiar or of personal interest.	1 2 3 4 5	1 2 3 4 5
35.	I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember	1 2 3 4 5	1 2 3 4 5

	significant points.		
36.	I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view.	1 2 3 4 5	1 2 3 4 5
37.	I can write personal letters describing experiences and impressions.	1 2 3 4 5	1 2 3 4 5
38.	I can express myself in clear, well-structured text, expressing points of view at some length.	1 2 3 4 5	1 2 3 4 5
39.	I can write summaries and reviews of professional or literary works.	1 2 3 4 5	1 2 3 4 5
40.	I can write clear, detailed text on a wide range of subjects related to my interests.	1 2 3 4 5	1 2 3 4 5
41.	I can write clear, smoothly flowing text in an appropriate style.	1 2 3 4 5	1 2 3 4 5
42.	I can write letters highlighting the personal significance of events and experiences.	1 2 3 4 5	1 2 3 4 5
43.	I can write about complex subjects in a letter, an essay or a report, underlying what I consider to be the salient issues.	1 2 3 4 5	1 2 3 4 5