Phonetic Transcription

Steven Gillis[*], Griet Depoorter[*], Simo Goddijn[**]

[*] *CNTS, University of Antwerp*

[**] *SPEX, University of Nijmegen*

This chapter was meant to appear in a handbook accompanying the 'final' release of the *Spoken Dutch Corpus*. Unfortunately the handbook will continue its virtual existence until etenity, … except for this chapter.

Phonetic Transcription

Steven Gillis[*], Griet Depoorter[*], Simo Goddijn[**]

[*] *CNTS, University of Antwerp*

[**] *SPEX, University of Nijmegen*

Abstract

This chapter describes the broad phonemic transcription in the CGN. First a broad overview of phonetic annotations in Dutch corpora is provided and a number of crucial dimensions are discussed: the source of annotation (human or automatic), the type of material involved, the level of transcription and the symbol set and transcription conventions. These dimensions serve as a guide through a number of aspects of the broad phonetic transcription in the CGN. In section 2 the level of transcription is discussed: methodological as well as fairly practical considerations are elaborated on with respect to the detail of phonetic annotations (or transcriptions) as well as the required level of expertise of the transcribers. In section 3 a pilot study is summarized that was meant to address issues such as the source of annotation and the transcription task (transcription 'from scratch' or verification of an automatically generated transcription) relative to practical matters such as the estimated transcription time, expected errors and variability. The next sections deal with the protocol: in section 4 the CGN set of phonetic symbols is defined and in section 5 the manual provided for the transcribers is described. The actual transcription procedure is dealt with in section 6: the entire corpus received an automatically generated broad phonetic transcription, ten percent of which, i.e. the 'core corpus', is manually verified. In section five the details of the grapheme-to-phoneme conversion of the orthographic annotation is described, as well as the manual verification procedure. The final section of this chapter is a first attempt to assess the quality of the manually verified transcriptions.

1. Introduction

From recent overviews of annotated Dutch corpora (Piepenbrock 1999, Bouma & Schuurman 1998, 2000, Daelemans & Strik 2002) it appears that phonetically annotated corpora of Dutch are relatively scarce. In the literature references can be found to the following corpora: the ANNO corpus[1], the COGEN corpus[2], the Flemish SpeechDat(II)[3], the Flemish/Dutch SpeechDat-Car corpus[4], the IFA corpus[5], the PBS corpus[6], the Speech Styles corpus[7], and the EUROM corpus[8]. In addition to these corpora of adult speech, there are several phonetically annotated corpora of adult-child spontaneous interactions in the CHILDES database[9]. The phonetic annotations in those corpora concern the young children's speech and not so much the adults'. In this context, the two lexical databases with phonetic annotations, viz. CELEX[10] and FONILEX[11] can also be mentioned.

These corpora differ substantially in a number of respects. First of all, there are quite substantial differences as to the amount of data that are phonetically annotated: as far as could be determined from the documentation, ANNO contains 646,500 word tokens, IFA ca. 50,000, PBS 11,518, SpeechStyles 118,000. But they also differ in other respects. The main dimensions are summed up in what follows, and these dimensions will provide the skeleton of our discussion of the phonetic transcription process of CGN in the remainder of this chapter.

- *Source of annotation*: some corpora are phonetically annotated by human transcribers who actually listen to the sound recordings (as was the case for PBS). In a fair amount of other corpora, the transcription is arrived at automatically by applying a grapheme-to-phoneme conversion program. The latter is done in the case of the ANNO corpus, the Speech Styles corpus, the

---

1 Schuurman 1997, see also http://www.ccl.kuleuven.ac.be/about/ANNO.html

2 Webstek? → Jean-Pierre!

3 http://www.speechdat.com/SpeechDat.html

4 http://www.elda.fr/catalogue/speech/S0139.html

5 Pols & Van Son (2002), Van Son, Binnenpoorte, Van den Heuvel & Pols (2001), Van Son & Pols (2001a, b), and see also http://www.fon.hum.uva.nl/Service/IFAcorpus

6 Cremelie (2000)

7 http://www.mpi.nl/world/tg/corpora/speechstyles/speechstyles.html

8 http://www.phon.ucl.ac.uk/resource/eurom.html

9 MacWhinney (2000) and see also http://atila-www.uia.ac.be/childes/data/germanic/dutch/ or http://childes.psy.cmu.edu/data/germanic/dutch/

10 http://www.kun.nl/celex/

11 http://bach.arts.kuleuven.ac.be/fonilex/

Flemish SpeechDat(II), and the Flemish/Dutch SpeechDat-Car corpusFor some corpora a mixed approach is taken: first the orthographic transcription is automatically converted into a phonetic one, which is subsequently checked by hand. The IFA corpus and the COGEN corpus exemplify this approach.

- *Type of material*: the language material in some corpora is balanced in the sense that they contain a set of words (pronounced in isolation or embedded in a sentence context) in which all phonemes or segments of the language are represented (possibly in all their phonotactically legal contexts). The PBS corpus is an example of a balanced corpus. Another aspect of balancing of the materials is whether a number of speakers produce the same words and/or sentences or whether the speakers engage in free conversation. SpeechStyles and COGEN have a mix of materials, while PBS is restricted to a very specific set of language materials. The type of material has far-reaching repercussions for the use of automatic grapheme-to-phoneme conversion on the basis of an orthographic transcription, as will be further discussed below.
- *Level of transcription*: transcriptions can range from fairly abstract to tightly connected to the speech signal. On one end of this continuum, transcription may carry morphophonological information (such as the final /d/ in /hud/ <hoed>, Eng. 'hat' which is normally devoiced in fluent speech). This type of transcription is not found in any of the corpora mentioned thus far, except for the FONILEX lexical database which contains a similar abstract layer. At the other extreme of the continuum transcription may be closely tied to the speech signal, requiring a multitude of diacritics.
- *Set of segments*: the level of transcription has obvious consequences for the set of segments used in the transcription. Though there is no uniformity as to the symbols used in various corpora, a consensus seems to appear to use some form of SAMPA[12] as a common set of symbols.
- *Endproduct*: in some corpora, a lexicon is provided with a phonetic transcription of the words in the corpus while other corpora provide a full transcription of the recordings. As a matter of course, a lexicon cannot represent phenomena that transcend the word level, like assimilation processes

[12] http://www.phon.ucl.ac.uk/home/sampa/dutch.htm

across word boundaries, while in the latter approach such phenomena are represented.

2. Level of transcription

In the foregoing we used the term 'phonetic transcription' in a fairly loose way. In fact various labels have been used for designating the different types of transcriptions (see Gibbon et al. 1997, Gillis 2000, Vieregge 1986). The following levels are distinguished in the EAGLES handbook (Gibbon et al. 1997):

- *Citation-phonemic representation*: a conversion of an orthographic transcription into phonetic symbols. Either a lexicon that contains a phonetic transcription for every orthographic entry is used, or a grapheme-to-phoneme converter, or both.
- *Broad phonetic or phonotypic transcription*: a transcription in which running speech phenomena like place assimilation, consonant deletion, vowel reduction etc. are manifest, as long as they can be described by symbols that have the status of phonemes. The symbols are used to mark the output of connected speech processes that delete, insert or substitute one phoneme for another.

  Broad phonetic and phonotypic transcriptions are categorized under one chapter in the handbook, but they can nevertheless be distinguished: phonotypic transcriptions are intended to be derived purely by rule, whereas broad phonetic transcriptions are intended to contain a "phonemic-level representation of the speaker's tokens" (p. 159). In other words, a phonotypic transcription is arrived at without actually listening to the source, while a broad phonetic transcription is made by humans. The phonotypic level may not be very different from a citation-phonemic representation made by concatenating lexical items since frequently occurring word-internal processes can be accounted for in the lexicon.
- *Narrow phonetic transcription*: a transcription in which detail that goes beyond the phonemic level is represented, such as allophonic variation, labialization, diphthongization of monophthongs, etc. This level of detail can only be obtained by listening to the acoustic signal and if necessary inspecting waveform and spectrogram. Gibbon et al. (1997: 160) formulate the following recommendation:

  "It is better not to embark without good reason on this level of representation,

which requires the researcher to inspect the speech itself, as this greatly increases the resources needed (in terms of time and effort). If the broad phonetic (i.e. phonotypic) level is considered sufficient, then labeling at the narrow phonetic level should not be undertaken."

- *Acoustic-phonetic transcription*: a transcription in which "every portion of speech that is recognizably a separate segment of the acoustic waveform or spectrogram" (p. 160) is manifest. For example, plosives would be segmented in closures and release bursts and aspiration phases would be segmented separately. This is clearly a very labor intensive procedure, and therefore not suitable for large amounts of speech. Apart from that, this kind of transcription requires hi-fi speech quality without background noises, which would rule out large parts of the CGN.

The choice between these various possible annotations was made considering, on the one hand, the needs of the prospected users of CGN, and on the other hand the practical feasibility of the enterprise given time and budget limitations. As to the users[13], some phoneticians phrased a preference for narrow transcriptions, while others were more interested in broad or even canonical (citation-phonemic) transcriptions because they were very skeptical with regard to the level of accuracy of narrow transcriptions. Speech technologists on the other hand, were in favor of (broad) manual transcriptions because this type of transcription was deemed absolutely necessary for i.a. automatic alignment.

Eventually broad phonetic transcription was opted for. Several reasons led to this decision. First of all, finer granularity of the transcription carries the risk of less reliability. In other words, more phonetic detail implies less agreement between transcribers, as Shriberg & Lof (1991) pointed out. They reported mutual agreement of only 33% for the use of diacritic symbols. Moreover it is not inconceivable that more detail also elevates the risk of less consistency in the products of a single transcriber. It was felt therefore that for a large scale project like CGN -- the output of which would constitute a de facto standard and reference for many years -- a corpus should result that carries as much consensus as possible. Hence, a transcription was opted for that guarantees as much potential mutual agreement as possible.

---

[13] In CGN workshop prospected users were asked to phrase their desiderata with respect to various aspects of the corpus, including the level of the phonetic annotation.

Second, the higher the level of required detail, the more time (and money) required for transcribing the data. From a pilot study on broad transcriptions (see below), it appeared that for one minute of speech the transcription time varied from 35 to 60 minutes, depending on the transcribed speech variety. Probably (much) more time would have to be spent to obtain narrow transcriptions.

Third, if a form of narrow transcription were chosen, two problems would turn up. Narrow transcription requires the experience of an expert phonetician, whereas broad transcription can be done satisfactorily by students with an appropriate training in phonetics (see below). In addition to the increase of transcription cost that this would imply, a survey of the available human resources made it clear that it would be very hard to find phoneticians who are willing to devote a considerable amount of their time to this routine job. Moreover, it turned out to be rather difficult (if not impossible) to arrive at a consensus regarding the specific phonetic details of the annotation scheme. Details that are important to one linguist may hamper the research of another.

The combination of these factors did not justify a choice for narrow phonetic transcription. The question remains which type of broad transcription to opt for: a transcription 'from scratch' or a transcription on the basis of an automatically generated transcription? Moreover at this point a clarification is required as to the level of detail of the transcriptions and the set of phonemic symbols. Last but not least it should be decided which transcribers should be envisaged to carry out the task. In order to address these issues, a pilot study was carried out.

3. Broad phonetic transcription: requirements and practical feasibility

The aim of the broad phonetic transcriptions was to arrive at high quality transcriptions of 10% of the entire corpus. In a sense this clear-cut aim is fairly ambiguous: a qualitative and a quantitative reading of it do not necessarily coincide.

In order to arrive at a qualitatively high-level transcription, one should address the issues discussed in i.a. Wester, Kessens, Cucchiarini & Strik (2001) with respect to the reliability of phonetic transcriptions: care should be taken so as to minimize the inter-subjective and intra-subjective variability. Ideally speaking, transcriptions

should be produced by more than one transcriber in order for a 'best' transcription to emerge.

In the same vein, in order to arrive at the aim of producing the required amount of transcriptions, a procedure should be selected that enables reaching the goal within the financial and time limits of the project. Wester et al. (2001: 378) characterize the task as "extremely time-consuming", "costly" and "often tedious", a characterization that phrases a consensus in the literature.

Ideally a procedure should emerge that enables reaching the qualitative as well as the quantitative requirements. A pilot experiment was therefore run in order to gain some insight into the time requirements for producing a broad phonetic transcription (detailed results and analyses can be found in Gillis 2000b). The goals of the study were

- to estimate the amount of time required to make transcriptions;
- to establish if transcriptions could be made by students as opposed to professional linguists;
- to establish the most efficient way to make transcriptions (i.e. start 'from scratch' or start from automatically generated transcriptions).

For the study, a 12-minute sample of speech data was selected, varying from read aloud speech to spontaneous multilogues. Transcribers were two professional linguists and two students, and their task consisted of either transcribing a fragment 'from scratch' or verifying a transcription automatically produced from the orthographic transcription using a grapheme-to-phoneme conversion program.

As to transcription time, the study showed a difference in the time required to verify the output of an automatic grapheme-to-phoneme converter (AT) against the speech signal and the time required to transcribe 'from scratch' (HT). On the whole, the transcribers spent 341 minutes (5.6 hours) on the AT task, thus verifying 2.9 words per minute (177 words per hour). In the HT task, the transcription time was higher: 406 minutes (6.8 hours) which amounts to transcribing 2.5 words per minute (148 words per hour). Extrapolating these figures to the entire corpus to be transcribed phonetically, this would mean that if the AT procedure were used the one million words of the CGN 'core corpus' to be transcribed would require ca. 5750

man-hours sheer verification time, and using the HT procedure ca. 6666 hours would be needed.

In Table 1 the average transcription time (measured over two transcribers) is displayed in terms of words transcribed/verified per hour as well as the time factor (time required for transcription/verification relative to length of the sound fragment). These data are displayed for different types of fragments[14] and the results for HT and AT are given separately.

Table 1: Average HT and AT transcription times and time factors

	Verification of AT		Transcription 'from scratch' HT	
Type of fragment	# words / hour	Time factor	# words / hour	Time factor
Read aloud speech	446	20	241	37
'Easy' dialogue	447	26	344	34
'Difficult dialogue	285	45	312	42
Multilogue	331	36	296	41
Radio comments	303	41	262	48
Radio interview	249	46	295	39
Mean:	344	36	292	40

[14] The 'easy dialogue' consisted of a dialogue with minimal overlap between the speakers and relatively favorable background noise conditions. The 'difficult' dialogue did not have those characteristics. The 'multilogue' consisted of a conversation between various people having lunch in a relatively noisy room. The radio comments consisted of the speech of a commentator following a football game. The radio interview consisted of a football commentator interviewing a player on the pitch after a game.

The results show that AT is indeed faster than HT. In general, as could be expected, the 'easier' speech samples (read aloud speech, 'easy' dialogues, and the like) benefited more from the AT procedure than the more difficult ones (like multilogues, and interviews); for the latter both procedures seem to take up an equal amount of time. Thus, the AT procedure involves a fair amount of time gain, which is a good reason for adopting it.[15]

As to the quality of the transcriptions, there was no 'reference transcription' prepared for the pilot study that could be used as a touching stone. Nevertheless, the quality of the transcriptions could be assessed in several ways. Since we have an automatically generated transcription (AGT), we expected – irrespective of the quality of that transcription – both the AT and the HT to diverge from the AGT in a comparable way. Even if the latter were of poor quality, the transcriptions produced by our subjects should diverge from it quantitatively and qualitatively in a similar way. In Gillis (2000b) this comparison is documented from various angles. We restrict ourselves here to some striking results.

The results in Table 2 show a comparison between the automatically generated transcription (AGT) on the one hand and the transcriptions of the subjects on the other hand. Two comparisons are made: a comparison of the 'reference' and the transcriptions resulting from the AT procedure and those resulting from the HT procedure. The comparisons are made in terms of the differences with the AGT.

Table 2: Comparison of AGT with AT and HT transcriptions (percentage of total number of symbols in transcriptions)

	Deletions	Insertions	Substitutions	Total
AT	12.6%	2.1%	6.3%	21.0%
HT	13.7%	3.3%	10.9%	27.9%

[15] The time factors computed on a total of approximately 12 minutes of recording, will be checked against the time factors that were calculated on the basis of approximately 70 hours of recordings in section 6.3.

The figures in Table 2 show a clear-cut picture: AT and HT diverge from AGT in a similar way. Deletions predominate over substitutions and insertions, and the amounts are highly comparable.

A comparison of the transcriptions produced by the subjects of the pilot study reveals a segmental overlap between 82.3% and 86.6%. Moreover, the types and the amount of the divergences are highly similar.

Table 3: Comparison of AT and HT transcriptions (percentages of total number of segments)

	Deletions + Insertions	Substitutions	Total
AT	4.6%	6.8%	11.4%
HT	4.5%	11.3%	15.8%

Table 3 shows that the number of segmentation differences (deletions and insertions) are highly comparable for the HT transcriptions and the AT transcriptions. There is more agreement between the AT transcriptions as to substitutions: 6.8% of the segments differ between AT transcriptions while the difference is 11.3% for the HT transcriptions. However, the type of substitutions found in both comparisons is very similar. In Table 5 we show the substitution patterns per group of segments and in addition the percentage of each pattern relative to the total number of substitutions in each category is indicated.

Table 4: Substitution patterns

Substitution pattern	AT	HT
Short, lax vowels: Replaced by long, tense counterpart: /I/ > /i/, /E/ > /e/, /A/ > /a/, /O/ > /o/, /Y/ > /y/	79%	81%
Long, tense vowels:	89%	84%

Replaced by short, lax counterpart: /i/ > /I/, /e/ > /E/, /a/ > /A/, /o/ > /O/, /y/ > /Y/		
Obstruents: Voiced obstruent replaced by voiceless counterpart, or vice versa: /t/ > /d/, /d/ > /t/, /x/ > /G/, /G/ > /x/, etc.	92%	87%
Nasals: Nasal replaced by another nasal: /N/ > /n/, /m/ > /n/, /n/ > /m/, /n/ > /N/	91%	96%

The substitution patterns in Table 4 are very outspoken: in AT as well as HT transcriptions, the percentages of each category are very high, thus showing that a small number of categories accounts for the majority of cases. Moreover the patterns that we find in AT and in HT are also highly similar, thus showing that at least in this respect the discrepancies between AT and HT transcriptions are not very different.

In sum, on the basis of the pilot study (Gillis 2000b) it was decided that there were enough compelling reasons for preferring a procedure in which an automatically generated broad phonetic transcription would form the basis of the transcribers' verification work.

In the previous section, it was already pointed out that professional phoneticians would be hard to find to complete the verification task. For this reason the pilot study also involved a comparison between the transcriptions of professional linguists (i.c. two dialectologists with extensive transcription practice) and linguistically trained students (i.c. two students recently trained in phonetics).

When comparing the transcription time of the professionals and the students it appeared that the professionals took much more time to complete the task than the students. Students benefited more from the provisional transcriptions than professional linguists. Evidently, professional linguists took more time to judge the provisional transcriptions critically, while the students used this transcription to

transcribe at a higher pace. Agreement between transcribers varied from 83.6% (HT) to 88.6% (AT) and showed no differences between professional linguists and students. The variation in transcription time between transcribers was substantial: the professional linguists needed 50% more transcription time than the students and one of the linguists needed 20% more than the other. The students showed hardly any mutual divergence.

The assessment of actual transcription accuracy using a 'reference transcription' was beyond the scope of the pilot study, but it is not inconceivable that accuracy is correlated with transcription time. However, because of the large difference in time (and money) investment and the small difference in agreement between professional linguists and students, it was decided that transcriptions were to be made by students, if necessary corrected by professional linguists. It was also decided that the students should start from an automatically generated transcription. Apart from saving time, there are a number of advantages in not starting from scratch. Automatically generated provisional transcriptions provide a solution for cases of doubt: whenever there is doubt between two symbols, transcribers may be required to leave the symbol from the example transcription, thus improving reliability. Furthermore, they provide a way to maintain the one-to-one relationship with the orthographic transcription. For the practicability of the corpus, every orthographic word should have a phonetic counterpart. Without knowledge of the orthographic transcription it would be very hard to maintain this relationship. One solution would be to provide the orthographic transcription to the phonetic transcriber, however this would result in serious undesirable influences of the orthography.

Provisional phonetic transcriptions carry another risk: students may leave too many symbols unchanged and thus a bias could be created towards the automatic transcription. Therefore, it was to be explicitly pointed out to transcribers that they should consider the provisional transcription as no more than that - a means to save typing time and to help prevent typing mistakes. Transcribers should be encouraged to change anything that does not correspond with the speech signal.

These considerations resulting from the pilot study will be taken up again in the next sections, in which the set of symbols used in the transcriptions will be introduced (section 3), the level of detail of the broad phonetic transcription and the

specifics of the protocol will be elaborated on (section 4). Finally the transcription procedure will be discussed (section 5).

4. The CGN symbol set

The set of symbols to be used in the *broad phonetic transcription* requires a careful consideration of two issues: (1) What is the set of segments with phonemic status?, and (2) How to represent the segments?

For the identification of the segments with phoneme status, state-of-the-art publications about the phonology of Dutch were consulted, i.a. Booij (1995), Collier & Droste (1975), De Schutter (1978), Kager (1989), Trommelen & Zonneveld (1989), Heemskerk & Zonneveld (2000). The set of segments represented in Table 1 was selected and motivated in Gillis (2000a).

Insert Table 5 about here (add an IPA symbol for each CGN symbol)

In the literature a number of phonetic symbol sets have been proposed. In Dutch corpora, we find IPA, YAPA (used in FONILEX), CELEX and DISC (found in CELEX) and SAMPA. The CGN symbol set as represented in Table 5 is very close to SAMPA, though it departs from SAMPA in the following respects (see Gillis 2000a for a detailed and extensive motivation of the choices made for the CGN symbol set):

- *Vowels*: SAMPA differentiates between 'long' or 'tense' vowels on phonetic grounds, in the sense that tense and phonetically long vowels are transcribed with the diacritic ':' while the tense vowels that are only phonetically long before [r] are not transcribed with the diacritic. In the CGN set this difference is not retained and, hence, all tense vowels are transcribed without the diacritic.
- *Diphthongs*: SAMPA transcribes diphthongs as consisting of two vowels. Since Dutch only has closing diphthongs, the direction of the gliding is completely predictable, and hence the CGN set opts for a transparent transcription of the diphthongs as /E+, Y+, A+/. This also circumvents the difficult problem of monophtongization of diphthongs: for transcribers it is often very difficult – if not impossible -- to decide whether a diphthong is

rendered as a monophthong or as a genuine diphthong. The proposed transcription convention does not solve this issue and leaves it to the user of the CGN corpus to decide (most probably on the basis of more intricate acoustic analyses). This also means that a diphthong pronounced as a monophthong will not be transcribed as an 'overlong' (foreign) vowel (e.g., E:, Y:).

- *Vowel sequences*: SAMPA distinguishes six vowel sequences which are sometimes described as diphthongs, the so-called 'onechte diftongen' (Eng.: non-genuine diphthongs). In the CGN transcription conventions, these are represented as a sequence of a vowel and a glide.
- *Loan vowels*: The loan vowels, i.e. phonetically long counterparts of short or lax vowels and the nasal vowels as they occur in the French phrase 'un bon vin blanc', are represented by a sequence of symbols: the vowel followed by the diacritics ':' and '~' respectively. The transcription of these vowels will be restricted to loan words and, hence, monophthongized diphthongs are not represented as the long counterpart of a lax vowel.
- *Consonants*: the CGN symbol set coincides exactly with the SAMPA conventions. There is only one exception: the symbol /J/, lacking in SAMPA, is introduced in the CGN set. The status of this segment is relatively controversial: though it can be considered as a combinatorial variant of /n/ followed by /j/, /J/ nevertheless occurs in monomorphemes and closer scrutiny of the FONILEX lexical database reveals that the sequence /n/+/j/ does not always result in /J/.

It should be noted that in its present form, the CGN symbol set does not include a symbol for the glottal stop. Moreover well known allophones (such as the phones surfacing due to the place assimilation of nasals) can only be represented in transcriptions as long as they can be accommodated by the CGN symbol set. In this way, the borderline between the phonetic detail (not) represented in the broad phonetic transcriptions of CGN is determined by structural, phonological considerations. In the next section the set of specific instructions that the transcribers received, as they are laid down in the 'protocol' will be elaborated on and exemplified.

5. The protocol for broad phonetic transcription

The *Protocol Broad Phonetic Transcription* (Gillis 2001) is a document that aims to guide the transcribers in their task of verifying the automatically generated phonemic transcript (AT) in order to arrive at a broad phonetic transcription. The AT is automatically produced from the orthographic transcription following a procedure that will be elaborated on in section 6. The protocol provides transcribers with the CGN phoneme set and it lays down the rules by which they have to work.

The aim is to arrive at a verified broad phonetic transcription which stays within the given CGN phoneme set and which reflects phonetic processes like insertions, deletions and substitutions of segments. Gradual processes, such as degree of voice of plosives and fricatives, monophthongization of diphthongs, are not transcribed. Phenomena like nasalization and lengthening or shortening of vowels are not transcribed either.

5.1.Guidelines in the protocol

The task of the transcribers was to verify the AT and, if necessary, correct it. The protocol stipulates a number of general preliminary guidelines for this task as well as specific transcription conventions that will be presented in this section.

As to general guidelines, the protocol sets out highlighting the following:

- The attention of the transcribers was explicitly drawn to the fact that the AT was manufactured by a computer and that no human had checked it. Hence they should not expect it to be flawless. Therefore the protocol warned the transcribers against being influenced too much by the provisional AT, as it was no more than an aid to save them time.
- Transcribers were instructed to listen as often as felt necessary (but not endlessly) to stretches of about half a second of speech, overlapping in time. They were explicitly instructed to listen across segment boundaries as well and if a segment boundary was placed within a sound, they had to prepare a bug report so that (eventually) the boundary could be moved or removed. Also other errors had to reported.

- Transcribers were instructed to have the transcriptions reflect phonetic processes (substitutions, deletions and insertions of phonemes) resulting in a segment represented by a CGN symbol. In other words, transcribers were instructed to leave the symbol provided in the AT untouched when a sound was heard that was outside the CGN symbol set. For instance, in the first example the first vowel in 'margarine' may be heard as a long lax vowel (influenced by the following vocalized /r/). Instead of using the length symbol ':', the [A] from the AT is to be preferred. The diphthong /E+/ in the second example may be heard as a nasalized /E/. But since nasalized vowels were to be transcribed as such only in loan words the [E+] provided in the AT should be retained.

(1)

Orthographic transcription	Two possible transcriptions	Preferred transcription
Margarine (Eng : margarine)	mA:G@rin@	mAG@rin@
we zijn weer thuis (Eng : we're home again)	w@ zE~n wer tY+s	w@ zE+n wer tY+s

- Transcribers are asked to leave the symbol in the AT untouched in cases of doubt. For instance, in the first example in (2) the transcriber may be in doubt whether he hears a voiced or a voiceless coronal fricative. In that case, the guideline is to leave the symbol in the AT, viz. [z], untouched. In the second example, the /r/ may be hardly perceivable, and, again, the guideline is to leave the symbol [r] present in the AT as it is.

(2)

Orthographic transcription	Two possible transcriptions	Broad phonetic transcription
we **z**ijn weer thuis (Eng : we're home again)	w@ **z**E+n wer tY+s w@ **s**E+n wer tY+s	w@ **z**E+n wer tY+s
oma wo**r**dt negentig (Eng : grandma is getting	oma wO**r**t nex@nt@x oma wOt nex@nt@x	oma wO**r**t nex@nt@x

ninety)		

- Transcribers were instructed to keep the one-to-one correspondence at the word level between the orthographic and the phonetic annotation layers. In the data structures of CGN the word is the central unit, so it was of crucial importance to have this principle also reflected in the phonetic transcription layer. The one-to-one principle causes difficulties because continuous speech is not a sequence of separate words but a sequence of sounds. Problems emerge with phenomena like cross word degemination and cross word linking sounds. Specific notational conventions were developed for such cases, that will be discussed in section 5.2.

5.2. Special transcription conventions

In the broad phonetic transcriptions of the CGN a number of special conventions were introduced to deal with (1) identical segments across word boundaries, (2) consonant insertions at word boundaries, and (3) unintelligible speech.

Geminates are a clear example of identical segments across word boundaries. For instance, the word sequence “om meer” (Eng : for more) is very likely to be pronounced as [Omer], but to restore the one-to-one correspondence with the orthographical layer it is noted as [Om_mer], thus using the underscore as a convention for signaling identical segments or a sequence of segments across word boundaries. As the examples in (3) show, the underscore is also used for shared sequences of segments, like [hE**b@_b@**rE+kt] in which it cannot be decided if the [b@] is the second syllable of [hE**b@]** or the first syllable of [**b@**rE+kt], hence it is left undecided instead of forcing a random choice in such cases.

(3)

Orthographic transcription	**Broad phonetic transcription**
hij vraagt o**m m**eer.	hE+ vraxt o**m_m**er
(Eng : he asks for more)	
vee**l l**iever.	ve**l_l**iv@r
(Eng : much better / rather)	
ik vin**d d**at ook.	Ik fIn**d_d**Ad ok

(Eng : that's what I think too)	
de naam die op de lij**st st**ond.	d@ nam di Ob d@ lE+**st_st**Ont
(Eng: the name that was on the list)	
zij he**bbe**n **be**reikt.	zE+ hE**b@_b@**rE+kt
(Eng : they have reached/accomplished)	

In connected speech, cross-word linking sounds often occur, which do not occur when the words are pronounced in isolation. These linking sounds are actually not a genuine part of the first or the second word. They are surrounded by hyphens in the CGN transcriptions, as shown in the examples in (4).

(4)

Orthographic transcription	**Broad phonetic transcription**
Speeld**e hij** toen al?	speld**@-n-E+** tun Al
(Eng: did he play already at that time?)	
ik d**oe aa**n sport.	Ig d**u-w-a**n spOrt
(Eng : I exercise)	
waarom zeg je het dan?	warOm zEx j@-n-@t_tAn
(Eng: why do you say it then?	
is ie klaar?	Is-t-i klar
(Eng: is he ready?)	

The consonants most often inserted are [w], [j], [n] and [t]. The glides are usually inserted intervocalically, while the insertion of the nasal often occurs between a word ending in [@] followed by a words starting with [@]. Insertion of /t/ often occurs before the word <ie>, which is a reduced form of <hij> (Eng. <he>).
In general, such linking sounds are preceded and followed by a hyphen. In that way, the consonant is linked to both the preceding and the following word and the one-to-one correspondence with the orthography is retained.

A third special symbol in the broad phonetic transcription are the brackets []. These are used whenever a sound or several sounds or even a word is unintelligible (see examples in (5).

(5)

What you hear (? = unintelligible)	**Broad phonetic transcription**
hij leest ?endertig bladzijden.	hE+ lest []@ndErt@x blAtsE+d@
hij leest ? bladzijden.	hE+ lest [] blAtsE+d@
hij leest achten?tig bladzijden.	hE+ lest Axt@n[]t@x blAtsE+d@
hij leest ? ?	hE+ lest [] []
(Eng : he reads thirty eight pages)	

The last special symbol to be introduced is the hash (#), which is used to indicate that nonlinguistic material like coughing, laughing, crying, etc is audible (examples in (6).

(6)

Example
ev@ mE+N kel sxrap@ # zo dAd Iz bet@r
(Eng : let me just clear my throat # that's better)

5.3. Important phonetic processes

In addition to formulating the ground rules for the verification of the automatic transcript, the protocol also draws the attention of the transcribers to some very important phonetic processes that they should not overlook. The following processes (as described in the phonological literature, such as Booij (1995)) are mentioned and exemplified:

- *Final devoicing* : the voiced obstruents at the end of a word become voiceless, unless assimilation of voice across word boundaries takes place (examples in (7)).

(7)

Orthographic transcription	Broad phonetic transcription
krab	krAp
bord	bort
vlag	vlAx

absurd	ApsYrt
admiraal	Atmiral

- *Assimilation of voice (cross word and word internally)*: very often assimilation occurs between two obstruents: one of the consonants changes voice or disappears altogether due to degemination.

(8)

Orthographic transcription	Broad phonetic transcription
krabt (Eng : scratches)	krApt
klapband (Eng : blowout)	klAbAnt
zakdoek (Eng : handkerchief)	zAgduk
opvallend (Eng: striking)	OpfAl@nt
op zaterdag. (Eng: on Saturday)	Op sat@rdAx
handvat (Eng : handle)	hAntfAt
ik word ziek. (Eng : I'm becoming ill)	Ik wOrt sik
lief ding (Eng : sweet thing)	liv dIN
misdaad (Eng : crime)	mIzdat

- *Assimilation of place - Nasal assimilation*: Assimilation occurs very often when a nasal, especially [n] is followed by an obstruent or another nasal (examples in (9)).

(9)

Orthographic transcription	Broad phonetic transcription
onbepaald (Eng :undertermined)	Omb@palt
in bad (Eng : in the bathtub)	Im bAt
ongecontroleerd (Eng : uncontrolled)	ONG@kOntrolert
onmiddellijk (Eng : immediately)	OmId@l@k
in memoriam (Eng: idem)	Im_memorijAm

- *Assimilation of place - Palatalization*: When [s] or [n] is followed by [j], assimilation can occur (examples in (10)).

(10)

Orthographic transcription	Broad phonetic transcription
kusje (Eng : little kiss)	kYS@
ik mis je (Eng : I miss you)	Ik mIS_S@
oranje (Eng : orange)	orAJ@
Ben je thuis (Eng : are you at home)	bEJ_J@ tY+s

- *Assimilation of voice before a vowel*: A word final obstruent can be voiced when the following word begins with a vowel (examples in (11)).

(11)

Orthographic transcription	Broad phonetic transcription
op aanvraag.	Ob anvrax
dat is zo.	dAd Is_so
ik ook.	Ig ok
net als iedereen.	nEt Alz id@ren
of is dat niet zo?	Ov Iz dAt nit so
ik lach altijd.	Ik lAG AltE+t

- *Consonant insertion*: When two vowels follow one another – across word boundaries or word internally - a consonant is often inserted (examples in (12)).

(12)

Orthographic transcription	Broad phonetic transcription
bio (Eng: bio)	bijo
televisie-interview. (Eng: television interview)	tIl@vizijInt@rvju
duet (Eng : duet)	dywEt
doe het snel. (Eng : do it quickly)	du-w-@t snEl

- *Schwa insertion*: The sequence of a liquid and a (heterorganic) obstruent or nasal, is often broken up by an inserted schwa (examples in (13)).

(13)

Orthographic transcription	Broad phonetic transcription
melk (Eng : milk)	mEl@k
werk (Eng : work)	wEr@k
erg (Eng : bad)	Er@x
Delft	dEl@ft
arm (Eng : poor)	Ar@m

- *Final n-deletion*: The (word)final [n] is often not pronounced especially when preceded by a schwa (examples in (14)).

(14)

Orthographic transcription	Broad phonetic transcription
tegen (Eng : against)	teG@
kopen (Eng : to buy)	kop@
molen (Eng : mill)	mol@

- *Word internal degemination*: A cluster of two identical consonants in the orthographic transcription is usually pronounced as if there were just one consonant (examples in (15)).

(15)

Orthographic transcription	Broad phonetic transcription
rustte (Eng : rested)	rYst@
onnodig (Eng :unnecessary)	Onod@x

6. Transcription procedure

The entire Spoken Dutch Corpus is enriched with an automatically generated broad phonetic (or phonemic) transcription. For a selection of ten percent of the data, i.e. the

so-called 'core corpus', the transcription was verified manually. This section describes the procedures that were followed to obtain both types of transcriptions.

6.1. Automatically generated transcription

The automatically generated transcription consisted of a concatenation of the canonical phonemic transcriptions drawn from the CGN lexicon. In this lexicon, all so-called obligatory word-internal processes are applied, whereas optional word-internal processes are not. Each word in the orthographic transcription was looked up in the lexicon and replaced by its corresponding phonetic transcription.[16] The basic sources for the transcriptions in the CGN lexicon were the lexical databases CELEX for the Northern Dutch data and FONILEX for Southern Dutch data.[17] For word forms not present in one of these databases, a special procedure was devised.

For out of vocabulary items, a grapheme-to-phoneme conversion was used to generate a transcription. For Southern Dutch data the memory-based learning system TiMBL (Tilburg Memory-based learner, Daelemans et al. 1998) was used to train a grapheme-to-phoneme converter with FONILEX as training source. TiMBL was trained using the basic IB1 algorithm and default parameter settings. A windowing technique was used that took three graphemes preceding the target grapheme and three graphemes following the target grapheme into account. The resulting converter was used for all Flemish transcriptions, i.e. the system was not retrained once additional CGN data became available. Out of vocabulary items in Northern Dutch

---

[16] Initially a more diversified strategy was envisaged: since FONILEX contains transcriptions in three different speech styles (from 'sloppy' to very careful pronunciation), the idea was to adapt the initial transcription to the speech styl of each fragment. Thus in a casual conversation, the 'sloppy' transcription would be used while in a formal lecture, the 'very careful' transcripion would be used. This idea proved impractical and in addition, the different speech styles are only represented in FONILEX and not in CELEX.

[17] For the Northern Dutch data, remaining gaps after the look-up procedure were filled by obtaining transcriptions from several other lexica, viz. the CELEX English database for English words and Onomastica for proper nouns. The phonetic representations obtained from different sources appear to vary with respect to the application of the process of /n/-deletion after schwa, resulting in inconsistency in this respect in the provisional transcriptions.

data were obtained by means of the rule-based grapheme-to-phoneme converter FONPARS (Kerkhoff et al. 1984).

Before entering the manual verification process, the transcriptions underwent post-processing. In the course of time, errors in the CGN lexicon that were discovered during transcription were corrected and stored in a separate lexicon. These transcriptions prevailed over the transcriptions from other sources.

Words marked with a "*" in the orthographic transcription[18] were often not present in the CGN lexicon. To retain the one-to-one correspondence with the orthography, they were either reproduced literally from the orthography (Flanders) or obtained by means of grapheme-to-phoneme conversion (Netherlands).

6.2. Transcribers

In The Netherlands, the transcribers were recruited among linguistics students. In Flanders, however, it proved very difficult to find students who were prepared to engage themselves for an extended period of time. As a result the bulk of the Flemish data were transcribed by an experienced linguist. This condition is a change for the better for the reliability of the data: with only one person working on the transcriptions, the circumstances for creating reliability are ameliorated , though intra-subject variability can not be ruled out and the procedure bears the risk of introducing a transcriber bias.

In the Netherlands, it was easy to find students who were prepared to work for a longer period of time. Prior to their CGN work, transcribers had little or no transcription experience. Most of them did, however, have basic phonetic knowledge. Before transcribers were engaged, a listening test was administered in order to assess their listening skills. Next they received a training that consisted of refreshing their knowledge of relevant phonetic terminology and concepts and a familiarization with the transcription protocol and the transcription tool. Initially students were hired for a trial period of 24 to 40 hours and if judged fit for the job, they were hired for at least

[18] An asterisk is used for foreign words, dialect words, dialectically pronounced words, words that are broken off and mispronounced words, and the like (see the protocol for orthographic transcription).

half a year. Most transcribers worked 12 hours a week, and were paid €10,89 gross per hour. For reasons of efficiency and consistency, we aimed at engaging transcribers as long as possible. The training period takes up a lot of time: transcribers needed approximately two months to develop a reasonable transcription quality at a reasonable pace.

Transcribers were expected to carry out their work in one and the same room in order to enable mutual communication about the transcriptions. During the recruitment period, no attention was paid to the place of origin of transcribers, nor to the place where they were raised. However, transcribers who originated from a certain part of the country were regularly asked for their opinions by other transcribers if speech samples originated from that same part of the country.

6.3. Transcription process

Transcribers worked in the label modus of the interactive speech processing tool Praat, on a Windows platform. They were provided with the wave form of the speech signal and a number of tiers, representing the individual speakers in the speech sample. An "unknown" tier was used for stretches of speech of which the speaker could not be identified. The tiers contained the automatically generated transcription, including segment boundaries. The transcribers used acoustic information; visual information was at the most used for confirmation. Navigating through the signal and listening to parts of the signal was enabled through key combinations or mouse clicking. Transcribers were instructed to work in a window of two seconds in order to give them a sense of the degree of precision they were expected to deliver. They were to listen as often as they thought necessary (but not endlessly) to stretches of about half a second of speech, overlapping in time, and modify the automatically generated transcription until it reflected what had been said. They were explicitly instructed to listen across segment boundaries as well.

Transcribers were expected to listen to the signal in chronological order (and not speaker after speaker) to help them understand the meaning of what was said. Ideally, a phonetic transcription should not take context into account, but it often proves to be fairly difficult to categorize a (spontaneous) speech sound out of the

blue. And if a speech sound could equally well be described with two different symbols, it is useful to stick to the canonical form for the sake of reliability

In Flanders, 100% of the data were rechecked by the experienced linguist. For budgetary reasons, only 50% of the Northern Dutch transcription data were checked by a second transcriber. The whole transcription process was monitored by a professional phonetician, who also trained the students.

On hindsight, the pilot study reported on in section 2 (see Table 1) provided a fairly good estimate of the time required for producing the broad phonetic transcriptions. In Table 6 we present an overview of the verification time devoted by the Dutch students ('first transcriber' or 'first pass') and the time devoted by the second transcriber ('second pass') in checking the transcriptions. The figures are split up by category in the corpus ("category"). For each category the length of the sound files that were checked is indicated ("length of fragments") in hours:minutes:seconds for the first and the second pass separately, since – as indicated before – the entire core corpus went through a first pass and only part of the material through a second pass. Next the ratio of the length of the speech fragments and the time the students spent verifying the automatically produced transcription against the actual sound material is indicated ("Ratio first pass"). The ratio should be read as a time factor: the duration of the sound fragments should be multiplied by the ratio in order to arrive at the actual transcription or verification time. The next columns specify analogous information for the second pass ("Length fragments second pass" and "Ratio second pass"), i.e. the time the second transcriber spent verifying the output of the first transcriber(s).

Table 6: Transcription time required for the Northern Dutch material according to category

Category	Length of fragments first pass	Ratio first pass	Length of fragments second pass	Ratio second pass
1. conversations ('face-to-face')	9:53:39	52.23	9:02:05	29.96

2. interviews	2:16:10	67.27	1:04:21	31.62
3. telephone conversations	14:42:40	43.15	9:03:09	27.27
4. business transactions	2:14:33	47.71	1:36:27	25.87
5. broadcasted interviews and discussions	6:12:22	43.14	3:33:49	36.95
6. discussions, debates,meetings (non-broadcasted)	2:17:22	50.81	1:35:50	23.06
7. lectures	2:43:36	32.76	2:29:17	23.31
9. spontaneous commentary	1:57:43	51.67	0:25:37	41.38
10. newsreports, current affairs programmes	2:12:47	41.23	1:24:19	28.46
11. news broadcast	2:26:28	46.31	0:21:36	19.44
12. commentary	2:24:31	40.45	1:35:41	29.05
13. lectures, speeches	2:12:29	47.78	0:54:42	20.29
14. read aloud text	8:51:45	24.58	0:12:02	10.80
Total	69:48:56		33:18:55	
Mean		45.32		26.73

*** Editors: please fill in the appropriate terminology used for the various categories

60 hours 16minutes and 05 seconds were verified by the first transcribers (first pass) who spent approximately 2602 hours on this task. Approximately half of that amount of transcriptions was verified again by a second transcriber (second pass): 33 hours 18 minutes and 55 seconds of recordings were checked by the second transcriber and this took up approximately 905 hours. This means that for the phonetic transcription, a mean ratio of 45.32 was achieved (one minute of sound material required 45 minutes of transcription time); the second pass achieved on average a ratio of 26.73, which adds another half hour to the transcription time of a minute speech. Thus if two passes are implemented, a time investment of about one hour and fifteen minutes is required for every minute of speech, solely and exclusively for the transcription and verification tasks.

The data show a fair variation (range 24.58 – 67.27 for the first pass and range 10.80 – 41.38 for the second) between text types: read aloud text is relatively fast to transcribe, while spontaneous speech, interviews and the like require much more time.

The data also show a reasonable difference between the first and the second pass: the second one requires somewhat less than half of the time of the first pass for lectures and read aloud text, to almost 80% of the time of the first pass for interviews and discussions.

6.4. Application of the protocol

Using protocol is indispensable for the uniformity of transcriptions, but a protocol cannot foresee every possible transcription problem. During the first months of the transcription process, a number of problems turned up that had not surfaced in the pilot transcriptions. The protocol was adjusted to account for the small gaps. In addition a few interpretation problems turned up:

- A first problem related to the transcription of voice. The continuous nature of the feature voice makes it difficult to categorize sounds that are “somewhere in the middle” between voiced and voiceless. Transcribers will typically appreciate the border differently. Moreover, the distinction between voiced and voiceless is very complex: Ernestus (2000) distinguishes no less than 8 acoustic cues for the distinction. Because obtaining consistency in the transcriptions was primordial,

the choice for the transcribers was made somewhat less difficult by instructing them to transcribe voice if they heard any voice and to ignore other cues. In case of doubt, they were to leave the symbol provided in the automatically generated transcription.

- A second problem (especially for Northern Dutch data) related to the transcription of /r/, /n/ and to a lesser extent /l/. In syllable-final position these sonorants are frequently only perceptible in the preceding vowel, and can hardly be identified as separately localizable sounds. Because of the restricted set of CGN symbols, in which rhoticized, nasalized or lateralized vowels are not represented, transcribers could not use the latter. We opted for transcription of both the vowel and the /r/, /n/ or /l/ to reduce the loss of information.
- A third interpretation problem related to the transcription of unreleased plosives. Plosives are regularly left unreleased, especially in sequences of plosives like [pd] in <opdracht> (Eng. assignment). Without the release, it is difficult to decide if a plosive is present at all in the signal, let alone to decide on its voice characteristics. For this reason, transcribers were instructed to leave the symbol pregenerated in the transcription whenever traces of the plosive were heard.

6.5. Post processing

After manual transcription and verification , an automatic check was performed with a software tool that verified i.a. the following aspects of the transcriptions:

- the use of illegal symbols
- the illegal use of special conventions (like underscores and hyphens)
- incorrect syntax of speaker names
- incorrect file format
- superfluous whitespace
- unlikely high speech rates
- consistency with the orthographic transcription (by comparing the number of tiers, the number of chunks per tier, the number of words per chunk and the number of symbols per word)

- consistency of segment boundaries with the orthographic transcription

Superfluous white spaces were corrected automatically. Inconsistencies in the location of segment boundaries were corrected interactively with a software tool and the other errors were corrected manually, either in the orthography or in the broad phonetic transcription. Very high speech rates proved to be a useful check, though spotting it did not always lead to correction: in spontaneous speech very high speech rates do occur, for instance when a sequence of short words is uttered.

6.6. Automatic transcriptions

Ninety percent of the Spoken Dutch Corpus is transcribed automatically, without any human corrections. Because we assume that human consensus transcriptions are the best possible transcriptions, we intended to generate automatic transcriptions that resemble human consensus transcriptions as closely as possible. For the time being, the procedure is equal to the generation of the automatically generated provisional transcriptions.

7. Quality monitoring and control

Bearing in mind that human transcriptions are susceptible to unreliability and subjective bias, a number of precautions were taken to try to minimize these risks. For example, Dutch transcribers were supervised by a phonetician who monitored the transcription process closely, especially during the training period: recurring errors were spotted and discussed with the transcribers and remediation was attempted. As mentioned above, Dutch students were only hired if they were willing to participate in the project for at least 12 hours a week for a period of at least half a year, for the sake of efficiency and consistency. Moreover, they were required to work in the same room to be able to consult with each other.

As mentioned before, the first transcription cycle in Flanders was carried out by a trained phonetician, who also corrected all of the data. For half of the transcribed Dutch data (with a priority of spontaneous speech over other speech components) a second transcriber corrected the work of the first one. The data were post processed with a software tool (a Perl script) that performed several formal checks.

Notwithstanding the care that was taken to arrive at a high quality transcription, there will undoubtedly remain inconsistencies, disputable transcriptions and even blunt errors. Thus in this section we take up the question which is crucial for future users of the CGN to what extent our precautions have paid off. Because the independent quality check performed by BAS (Salverda Bird, Hajiç & Höge 2001) did not include an evaluation of the broad phonetic transcriptions (which were not yet available at that time), it was decided to perform an initial quality check internally (Binnenpoorte, Goddijn & Cucchiarini 2003, Goddijn & Binnenpoorte 2003). An attempt was made to assess the achieved transcription quality in the Northern Dutch CGN data by means of a comparison with a consensus transcription. Measuring within-transcriber agreement was beyond the scope of the study. In the next paragraphs, a summary of this study is provided .

7.1. Consensus transcriptions

In order to assess the quality of the phonetic transcriptions, a reference is needed for the sake of comparison. Ideally, this reference should represent the perfect transcription. It is, however, generally acknowledged that there is no absolute truth of the matter as to what phones a speaker produced in an utterance (Cucchiarini & Strik 2003). In other words: the perfect transcription does not exist. We may, however, try to approach it by having two or more experienced transcribers make a joint transcription in which they have to agree on every symbol, resulting in a consensus transcription. This is the procedure that was followed in our assessment of the broad phonetic CGN transcriptions.

Four CGN transcribers, each having more than five months transcription experience, transcribed 16 minutes of speech according to the CGN protocol. They started from automatically generated transcriptions in order to optimally match the real CGN conditions. The selected speech material varied with respect to speech style and speaker, thus constituting a plausible sample of the Northern Dutch part of the CGN. The material consisted of 16 different samples representing four speech styles in increasing order of spontaneity: read aloud text (RS), prepared lectures (LC), interviews (IN) and spontaneous conversations (SC). The same material was

transcribed by two experienced linguists, who started from scratch, and who had to reach a consensus on the transcription.

Next the consensus transcription and the transcription of the four CGN transcribers were aligned. The alignment was performed by the program Align (Cucchiarini 1993), a dynamic programming algorithm that returns the number of deletions, insertions and substitutions but also calculates a distance measure between two transcriptions taking into account and weighing differences in terms of articulatory features such as place and manner of articulation, voice, lip rounding, etc.

7.2 Results

Percentages agreement (solely based on number of deletions, insertions and substitutions) between students and consensus ranged from 85.3% for SC to 93.9% for RS. Inter-transcriber agreement ranged from 85.7% to 96.3%, the lower figure representing agreement between two students for SC and the higher figure agreement between two different transcribers for RS.

On average 43% of all observed substitutions in relation to the consensus transcription involved only one feature, resulting in a small distance. The most salient substitution pattern appeared to involve the feature voice: in RS the five most frequent substitutions were of this type. In the more spontaneous speech varieties vowel reduction also played a role, as well as actual transcriptions for what was deemed "unintelligible" in the orthographic transcription. The research was further limited to voice substitutions.

Most voice substitutions involved a voiceless obstruent in the reference transcription which was transcribed as its voiced variant in the students' transcriptions. From personal communication with transcribers it appeared that they were inclined to transcribe the voiced variant whenever a plosive or a fricative was unclear or 'soft', although they were instructed to base their choice only on the feature voice and not on other cues (see above). Table 7 shows that of the segments that were most liable to substitution (/x-G/, /t-d/, /f-v/, /s-z/, /k-g/), 78% to 95% of them (depending on phoneme and speech style) were transcribed in accordance with the reference transcription. For 84% up to 95% of these substitutions, the previous and

subsequent segments were voiced. Apparently, transcribers found it more difficult to establish voicelessness in an all-voiced context.

Table 7: Percentage unvoiced-voiced substitutions of all occurrences of the unvoiced segment in RT

%	x,G	t,d	f,v	s,z	k,g
RS	19	6	11	5	11
LC	13	10	5	8	11
IN	22	11	8	10	17
SC	20	10	21	13	22

A special case is the /x-G/ substitution, which represented the most frequent voice substitution. In the symbol set a distinction is made between /x/ and /G/, the first symbol representing the voiceless dorsal fricative and the second symbol its voiced variant. The distinction is justified by the fact that in some southern parts of the Netherlands, as well as in Flanders, a distinction between the sounds is made and experienced in spoken language. In the rest of the Netherlands however, both speech sounds are used, dependent on idiosyncracy and context, but there is no awareness of the distinction. It proved to be very difficult to make our transcribers aware of the distinction, especially since they tend to confuse it with another difference in pronunciation of /x/ between the southern and northern part of the Netherlands, which is much more conspicuous to them (i.e. the more velar pronunciation of this speech sound in the south). Three of the four transcribers had trouble in discriminating these speech sounds: the /x/-/G/-confusions were to be attributed almost entirely to these three.

References

Binnenpoorte, D., Goddijn, S., & Cucchiarini, C. (2003). *How to improve human and machine transcriptions of spontaneous speech*. Paper presented at the ISCA and IEEE workshop on spontaneous speech processing and recognition, Tokyo, Japan.

Booij, G. (1995). *The phonology of Dutch*. Oxford: Clarendon.

Bouma, G., & Schuurman, I. (1998). *De positie van het Nederlands in Taal- en*

Spraaktechnologie (Report for the Nederlandse Taalunie).
Bouma, G., & Schuurman, I. (2000). De digitale infrastructuur van het Nederlands. *Nederlandse Taalkunde, 5*, 90-94.
Collier, R., & Droste, F. (1982). *Fonetiek en fonologie*. Leuven: ACCO.
Cremelie, N. (2000). *Heuristische zoekstrategie en automatische aanmaak van lexica voor continue spraakherkenning*. Unpublished PhD, Universiteit Gent, Gent.
Cucchiarini, C. (1993). *Phonetic transcription: A methodological and empirical study*. Unpublished PhD, University of Nijmegen, Nijmegen.
Cucchiarini, C., Binnenpoorte, D., & Goddijn, S. (2001). *How to combine efficiency and good transcription quality*. Paper presented at the Eurospeech, Aalborg, Denmark.
Cucchiarini, C., & Strik, H. (2003). *Automatic phonetic transcription: An overview*. Paper presented at the ICPhS, Barcelona, Spain.
Daelemans, W., & Strik, H. (2002). *Actieplan voor het Nederlands in de taal- en spraaktechnologie: Prioriteiten voor basisvoorzieningen* (Report for the Nederlandse Taalunie).
Daelemans, W., Zavrel, J., Van der Sloot, K., & Van den Bosch, A. (1998). *TiMBL: Tilburg Memory Based Learner, version 1.0, reference manual* (Technical Report ILK-9803, ILK, Tilburg University).
De Schutter, G. (1978). *Aspekten van de Nederlandse klankstruktuur [= Antwerp papers in linguistics 15]*. Wilrijk: University of Antwerp.
Den Os, E. (1994). Transliteration of the Dutch Speech Styles Corpus. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam, 18*, 87-94.
Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch: A corpus-based study of the phonology-phonetics interface*. Utrecht: LOT.
Gibbon, D., Moore, R., & Winski, R. (1997). *Handbook of standards and resources for spoken language systems*. Berlin: Mouton de Gruyter.
Gillis, S. (2000a). *Motivering fonemische transcriptie* (CGN Internal Document).
Gillis, S. (2000b). *Fonemische proeftranscripties Vlaanderen* (CGN Internal Document).
Gillis, S. (2001). *Protocol voor brede fonetiche transcriptie* (CGN Internal document).
Goddijn, S., & Binnenpoorte, D. (2003). *Assessing manually corrected broad phonetic transcriptions in the Spoken Dutch Corpus*. Paper presented at the ICPhS, Barcelona, Spain.
Heemskerk, J., & Zonneveld, W. (2000). *Uitspraakwoordenboek*. Utrecht: Spectrum.
Kager, R. (1989). *A metrical theory of stress and destressing in English and Dutch*. Unpublished PhD, Rijksuniversiteit Utrecht, Utrecht.
Kerkhoff, J., & Wester, J. (1987). *Fonpars user manual. Part I: Rule format* (Final Documentation of the ESPRIT-project 860, Section 3.2, Code: NU-Cb/1, 12-02-87).
Kerkhoff, J., Wester, J., & Boves, L. (1984). A compiler for implementing the linguistic phase of a text-to-speech conversion system. In H. Bennis & W. van Lessen Kloeke (Eds.), *Linguistics in the Netherlands* (pp. 111-117). Dordrecht: Foris.
MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Hillsdale: Erlbaum.
Piepenbrock, R. (1999). *Nederlandse gesproken corpora: Een inventarisatie* (CGN Internal Document).
Pols, L., & Van Son, R. (2002). Accessing the IFA-corpus. In N. Volskaya & N.

Svetozarova (Eds.), *Book in honor of the 70-th anniversary of Prof. L.V. Bondarko* (pp. 316-320). St. Petersburg: University of St. Petersburg.

Salverda, R., Bird, S., Hajic, J., & Höge, H. (2001). *Mid term evaluation Spoken Dutch Corpus project*.

Schuurman, I. (1997). ANNO: A multi-functional Flemish text corpus. In J. Landsbergen & J. Odijk & K. van Deemter & G. Veldhuijzen van Zanten (Eds.), *Computational Linguistics in the Netherlands 1996. Papers from the Seventh CLIN Meeting* (pp. 161-176). Eindhoven: IPO, Technische Universiteit Eindhoven.

Shriberg, L., & Lof, L. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics, 5*.

Trommelen, M., & Zonneveld, W. (1989). *Klemtoon en metrische fonologie*. Muiderberg: Coutinho.

Van Son, R., Binnenpoorte, D., Van den Heuvel, H., & Pols, L. (2001). The IFA corpus: A phonemically segmented Dutch open source speech database. *Proceedings of Eurospeech 2001, 3*, 2051-2054.

Van Son, R., & Pols, L. (2001). The IFA Corpus: A phonemically segmented Dutch "Open Source" speech database. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam, 24*, 15-26.

Van Son, R., & Pols, L. (2001). Structure and access of the open source IFA Corpus. *Proceedings of the IRCS workshop on Linguistic Databases, Philadelphia*, 245-253.

Vieregge, W. (1985). *Transcriptie van de spraak: Theoretische en praktische aspecten van de symboolfonetiek*. Dordrecht: Foris.

Table captions

Table 1: Average HT and AT transcription times and time factors

Table 2: Comparison of AGT with AT and HT transcriptions (percentage of total number of symbols in transcriptions)

Table 3: Comparison of AT and HT transcriptions (percentages of total number of segments)

Table 4: Substitution patterns

Table 5: The CGN set of phonetic symbols

Table 6: Transcription time required for the Northern Dutch material according to category

Table 7: Percentage unvoiced-voiced substitutions of all occurrences of the unvoiced phoneme in RT