

# Het transcriptieprotocol van het Gesproken Corpus van de Nederlandse Dialecten (GCND)

## I. INLEIDING

In 1957 wordt Edgard Blancquaert aan de UGent als professor Nederlandse Taalkunde opgevolgd door Willem Pée. Die was toen al geruime tijd medewerker aan de *Reeks Nederlandse Dialectatlassen*<sup>(1)</sup>, het grote project dat Blancquaert had opgezet, en het was hem opgevallen dat de aard en de kwaliteit van fonetische transcripties dikwijls afhingen van de enquêteur. Al langere tijd waren dialectologen als Blancquaert en Pée overtuigd van de meerwaarde van geluidsopnames voor de objectieve vastlegging van dialect, in vergelijking met het schriftelijke noteren ervan, maar nu waren daarvoor ook de technische mogelijkheden ter beschikking gekomen.

In 1962 slaagde Pée erin bij het FKFO<sup>(2)</sup> de nodige financiële middelen te verwerven om zijn plan ten uitvoer te leggen. Dat plan bestond erin om in zoveel mogelijk gemeenten in Nederlandstalig België en Frans-Vlaanderen geluidsfragmenten van het lokale dialect op te nemen, met een semi-professionele Revox-bandopnemer, die toen relatief goedkoop was geworden. Het FKFO financierde reis- en verblijfkosten; de UGent stond in voor de aankoop van het materiaal.<sup>(3)</sup>

<sup>(1)</sup> Zie [www.dialectzinnen.ugent.be](http://www.dialectzinnen.ugent.be)

<sup>(2)</sup> FKFO = Fonds voor Kollektief en Fundamenteel Onderzoek, later opgegaan in het Fonds voor Wetenschappelijk Onderzoek - Vlaanderen.

<sup>(3)</sup> Het Dialectenbureau in Amsterdam (nu Meertens Instituut) begon al in 1952 op instigatie van Jo Daan met het systematisch creëren van dialectische geluidsopnames (zie Rensink 1962 en <https://www.meertens.knaw.nl/soundbites/>).

De praktische uitwerking van het plan werd in de praktijk overgelaten aan medewerkers. Vooral professor Frits Vanacker, die Pée in 1971 zou opvolgen, moet hier genoemd worden, al werden er ook andere onderzoekers bij het project betrokken, zoals Johan Taeldeman en Georges De Schutter. Het idee was dus om voor elke gemeente een 45-tal minuten (de duur van één spel) zogenaamd ‘vrij gesprek’ op te nemen bij een zegsman of zegsvrouw, die aan een aantal objectieve en subjectieve criteria moest voldoen: behalve laaggeschoold, van hoge leeftijd en honkvast ook verstandig en mededeelzaam. Studenten Nederlands werden ingezet om in hun dorp of stad een geschikte zegspersoon te vinden. De student – in de jaren 60 van de vorige eeuw konden die meestal het plaatselijke dialect nog goed spreken – werd ook dikwijls ingezet als vragensteller, om ervoor te zorgen dat de zegspersoon zijn taal niet al veel zou proberen aan te passen aan de taal van de interviewer.

De gespreksonderwerpen waren in principe vrij; het spreekt natuurlijk vanzelf dat men met de criteria als hierboven op het platteland bij de boerenbevolking terecht kwam, en in de stad bij de arbeidersklasse. Zeer veel geluidsopnames gaan dus over landbouw, of over de toestand in fabrieken. Soms werd er een persoon uitgezocht die een of ander bijzonder beroep uitoefende of uitgeoefend had. Doordat de zegpersonen, die dikwijls twee wereldoorlogen hadden meegemaakt, de maatschappij drastisch hadden zien moderniseren, zijn hun levensverhalen vaak ook relevant voor historici.

In deze bijdrage lichten we de collectie en het belang ervan verder toe, alsook de manier waarop ze gedigitaliseerd en geannoteerd wordt. Het transcriptieprotocol voegen we aan het einde van dit artikel integraal toe; het kan immers een aanknopingspunt bieden voor toekomstig dialectonderzoek en de verdere ontsluiting van andere dialectcollecties in het Nederlandse taalgebied.

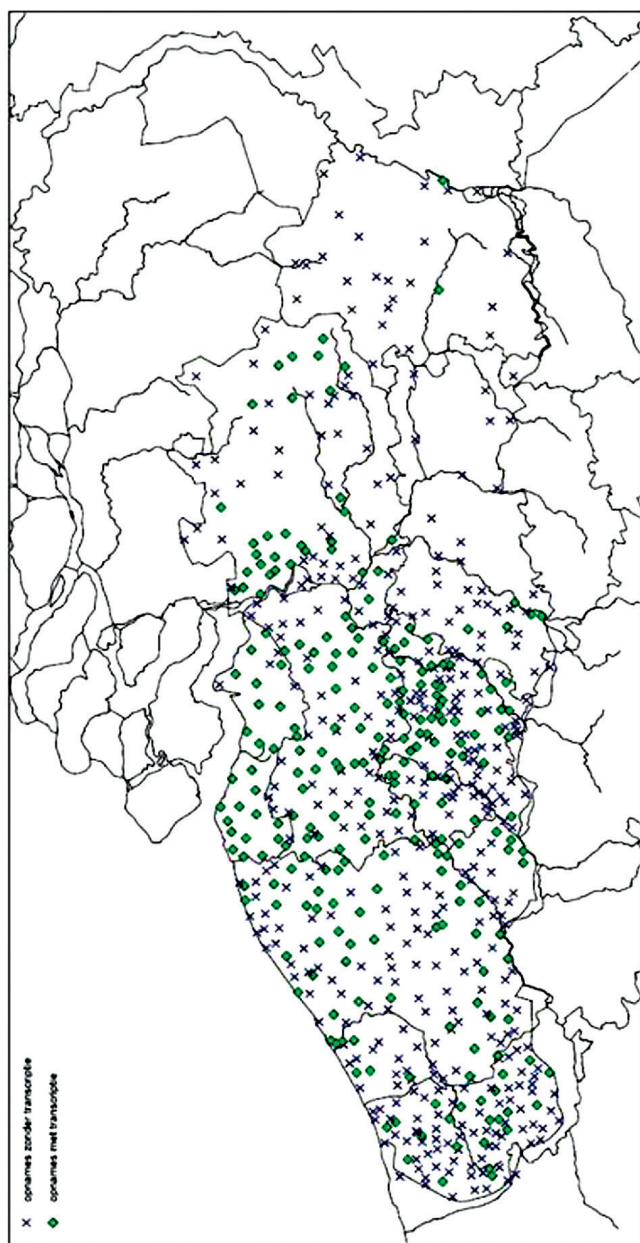


Fig 1. Verspreiding van de geluidsopnames

## 2. DE VERZAMELING 'STEMMEN UIT HET VERLEDEN'

De collectie 'Stemmen uit het verleden' bevat momenteel 783 banden voor 550 plaatsen, vooral uit het Vlaamse dialectgebied (= Frans-, West-, Oost- en Zeeuws-Vlaanderen). Naar het oosten toe (provincies Antwerpen, Vlaams-Brabant en Limburg) neemt het aantal geluidsopnames af (zie fig. 1); dat komt doordat aan de Gentse universiteit heel wat meer West- en Oost-Vlamingen studeren dan Brabanders, Antwerpenaars of Limburgers. In een aantal gevallen kon professor Vanacker een student ervan overtuigen een licentiaatsverhandeling te schrijven over de syntaxis van een bepaald dialect. In dat geval werden er voor de bestudeerde gemeente minstens drie opnames gemaakt, soms meer (zie bijvoorbeeld Van Keymeulen 1975).

De meeste banden werden opgenomen bij mensen die tussen 1885 en 1920 geboren zijn; de oudste spreker (uit Bossuit) is geboren in 1871. Voor Frans-Vlaanderen werd een speciale inspanning gedaan: daar heeft men een opname gemaakt voor elke gemeente waar nog inheemse Vlaamssprekenden aangetroffen konden worden. Achteraf zouden er ook in Zeeuws-Vlaanderen opnames bewerkt worden.

De verzameling 'Stemmen uit het verleden' is de omvangrijkste collectie geluidsfragmenten die voorhanden is voor de traditionele dialecten in Nederlandstalig België, Frans- en Zeeuws-Vlaanderen. Ze heeft in de eerste plaats een taalkundig oogmerk: het is een corpus van natuurlijke, spontane taal, wat voor sommige types van taalonderzoek een groot voordeel is. Dat is bijvoorbeeld in het bijzonder zo voor syntactisch onderzoek: doordat dialectsprekers zich van syntactische constructies vaak weinig bewust zijn en doordat dergelijke constructies dikwijls ook optioneel zijn en/of geconditioneerd door situationele context (zie Barbiers 2005), zijn syntactische constructies – in tegenstelling tot bijvoorbeeld woordenschat – lastig op te vragen via vragenlijsten.

Naast het taalkundige belang heeft de verzameling ook een grote etnologische waarde. Het is de grootste verzameling

levensverhalen van laaggeschoolde Vlamingen, waarvan de meesten tijdens hun leven heel wat hebben meegemaakt. Vele verhalen leggen getuigenis af van de grote armoede en onwetendheid waarin een groot deel van de Vlaamse bevolking zo'n honderd jaar geleden heeft moeten leven.<sup>(4)</sup> Naarmate de tijd vordert, neemt het taalkundige en etnologische belang van de verzameling dialectische geluidsopnames alleen maar toe.

### 3. DE WEBSITE WWW.DIALECTLOKET.BE

De hele verzameling werd in 2009 gedigitaliseerd<sup>(5)</sup> in het kader van een wetenschapspopulariseringsproject van de UGent, dat mede dankzij rector Paul Van Cauwenberge werd goedgekeurd. Ze werd het paradepaardje van de website [www.dialectloket.be](http://www.dialectloket.be), waarmee deugdelijke informatie over taalvariatie in het algemeen en dialecten in het bijzonder naar een groot publiek en de scholen wordt gebracht.<sup>(6)</sup> De verzameling werd ook inhoudelijk ontsloten door samenvattingen en trefwoorden, die als zoektermen gebruikt kunnen worden. Die samenvattingen en trefwoorden werden door vrijwilligers gemaakt in het kader van kleinschalige provinciale cultuurprojecten die aan de vereniging Variaties vzw. werden toegekend.<sup>(7)</sup> Op dit moment optimaliseert Variaties vzw., met de steun van de Vlaamse overheid, de bestaande trefwoorden en zoektermen, om de collectie nog beter inhoudelijk doorzoekbaar te maken.

(4) Zie Van Keymeulen e.a. (2019).

(5) De verzameling originele banden is bij het universiteitsarchief van de UGent ondergebracht; kopieën van de digitale versie ervan (+ kopieën van de fiches met metadata) zijn gedeponneerd zowel bij de Koninklijke Academie voor Nederlandse Taal en Letteren (KANTL) in Gent als bij het Meertens Instituut in Amsterdam.

(6) Voor meer informatie over de dialectologische digitaliseringsprojecten: zie Van Keymeulen (2020) in *Taal en Tongval* (ter perse). Dialectloket.be won in 2016 overigens een LIA ('Language Industry Award') in de categorie 'Best website in the language industry 2015'.

(7) *Variaties vzw. Koepelorganisatie voor Dialecten en Oraal Erfgoed in Vlaanderen* is een vereniging die een aantal dialectorganisaties en dialectliefhebbers samenbrengt.

#### 4. ANNOTATIE

Om de verzameling toegankelijk te maken voor taalkundig, historisch of volkskundig onderzoek, zijn ook transcripties – doorzoekbare teksten – van essentieel belang, zeker nu door dialectverlies steeds minder Nederlandstaligen in staat zijn de traditionele dialecten te begrijpen. Tegelijk met de opnames werden studenten of vrijwilligers ingezet om de opnames te transcriberen. Op die manier werden er 318 transcripties gemaakt (cf. Fig 1), meestal een typoscript, soms een manuscript. Die transcripties – die ook op Dialectloket te vinden zijn – zijn niet fonetisch, maar ‘woordelijk’, d.w.z. dat de dialectische uitspraak in een vernederlandste vorm is weergegeven. De transcripties werden vooral in de jaren 60 en 70 gemaakt door jobstudenten aan de hand van een aantal richtlijnen, die met wisselende ijver en kunde werden opgevolgd. Doordat de richtlijnen niet heel gedetailleerd waren en ze ook niet door iedereen op dezelfde manier werden nagevolgd, is er heel wat variatie in hoe dialectvariatie orthografisch wordt meegegeven. Dat bemoeilijkt natuurlijk taalkundig onderzoek. Daaronder zijn de transcripties momenteel nog niet digitaal doorzoekbaar; ze zijn enkel beschikbaar in de vorm van scans. Daar kan op zich met OCR – *Optical Character Recognition*<sup>(8)</sup> – een mouw aan gepast worden, maar dat is niet zo evident voor deze collectie, omdat sommige transcripties handgeschreven zijn, bij andere de inkt vervaagd is, en bij nog andere typoscripten ook nog manuele correcties werden aangebracht.

Het ontbreken van goede, digitaal doorzoekbare transcripties werd al een hele tijd als een gemis aanvoeld. Daarom nam professor Jacques Van Keymeulen in het academiejaar 2016-2017 het initiatief om een zogenaamde onderzoekslijn te starten voor bachelorstudenten, met als bedoeling digitaal doorzoekbare transcripties te maken, die gealigneerd bleven met het geluid (vergelijkbaar met ondertitels). De transcriptie gebeurde

(8) Bij OCR wordt patroonherkenning ingezet om afbeeldingen van teksten om te zetten in een opeenvolging van tekens.

manueel met het opensourceprogramma ELAN, dat het mogelijk maakt de opname in kleinere stukken te segmenteren en per 'segment' dan een transcriptie toe te voegen.

Om de transcripties zo eenvormig mogelijk te maken, werkte dr. Anne-Sophie Ghyselen – die ook de studenten opleidde – een nieuw transcriptieprotocol uit, gebaseerd op dat van de DIT-werkgroep (Dialect in Transcriptie) binnen het SAND-project.<sup>(9)</sup> Dat nieuwe protocol moest het mogelijk maken een breed gamma aan dialecten, van de westelijkste Frans-Vlaamse tot de oostelijkste Limburgse dialecten, zo uniform mogelijk orthografisch weer te geven, en moest ook toegankelijk en doorzoekbaar zijn voor zowel linguïsten (geïnteresseerd in fonetische, fonologische, morfologische, syntactische en/of lexicale dialectfenomenen), als niet-taalkundigen geïnteresseerd in de inhoud van de opnames. Daarom is ervoor geopteerd met twee transcriptielagen te werken: één die dicht aanleunt bij het dialect, de zogenaamde 'lichte' vernederlandsing, en één die dicht aanleunt bij het Standaardnederlands, de zogenaamde 'zware' vernederlandsing, waarbij de 'lichte' vernederlandsing als basis dient voor de 'zware'.

In de 'lichte' vernederlandsing worden dialectische fonemen in inhoudswoorden gerepresenteerd door de AN-pendanten ervan, en neergeschreven volgens de Nederlandse spellingregels (we schrijven dus *groot* en niet *gruët* of *grooët*); clitische vormen worden met hekje-tekens (#) van elkaar gescheiden (cf. *k#zijn#ek#ik ziek*). Er werd geopteerd voor een vernederlandsing van de fonologie in inhoudswoorden, omdat voor een uniforme en nauwkeurige weergave van de klankvariatie fonetisch schrift noodzakelijk is, wat de transcriptie te gecompliceerd en tijdrovend zou maken.<sup>(10)</sup> Bij de 'zware'

<sup>(9)</sup> Aan die werkgroep namen deel: Margreet van der Ham, Tamar Israël, Mathilde Jansen, Willy Jongenburger, Susanne van der Kleij en Gunther De Vogelaer. Guido Vanden Wyngaerd en Sjef Barbiers schreven de definitieve versie.

<sup>(10)</sup> Een groot probleem bij fonetische transcripties is dat de manier waarop fonetisch wordt getranscribeerd, sterk kan afhangen niet enkel van de kunde, maar ook van de talige achtergrond van de transcribent (cf. Blancquaert 1948, Coussé et al. 2004).

vernederlandsing wordt niet alleen de fonologie, maar ook de morfologie naar de standaardtaal toe genormaliseerd. Dialectische woordenschat en zinsbouw blijven op beide niveaus bewaard (cf. voorbeeld in Fig. 2). De ‘zware’ vernederlandsing maakt het mogelijk (semi)automatisch taalkundige annotaties toe te voegen met (deels aangepaste) computerprogramma’s die oorspronkelijk ontwikkeld werden voor de annotatie van het Standaardnederlands (zie hiervoor Farasyn et al. te verschijnen).<sup>(11)</sup> Voor meer informatie over de transcriptiemethode en een motivering van de gemaakte keuzes verwijzen we naar Ghyselen et al. (2020).

Op basis van de ervaringen van de eerste transcriptiestudenten werd het protocol verfijnd en op punt gesteld. De ervaring leerde dat studenten vaak moeite hebben om de dialectopnames woordelijk te begrijpen, niet alleen omdat het gesproken dialect vaak ver afstaat van hun eigen dagelijkse taalgebruik, maar ook omdat de studenten dikwijls weinig voeling hebben met de gespreksonderwerpen (bijvoorbeeld de traditionele landbouwtechnieken). Om zo nauwkeurig mogelijke transcripties te garanderen, werd de hulp ingeroepen van vrijwilligers, doorgaans van een relatief hoge leeftijd, die nakeken waar de transcripties aangevuld of verbeterd konden worden.

Na de eerste pogingen tot een gealigneerde transcriptie werden verschillende deelprojecten opgezet om extra opnames te kunnen transcriberen. In 2017 slaagde professor Anne Breitharth erin een FWO-navorsingskrediet te verwerven voor een transcriptieproject, waarmee 32 opnames getranscribeerd konden worden. In 2018 werd een FWO-postdoctoraal mandaat goedgekeurd voor dr. Melissa Farasyn, die zich voorneemt alle

<sup>(11)</sup> *Natural Language Processing (NLP)* veronderstelt voor onze data een aantal ingrepen om zoekfuncties mogelijk te maken: tokenisering (= het opdelen van de tekst in betekenisvolle reeksen tekens gescheiden door spaties, met inbegrip van de leestekens); lemmatisering (= het toevoegen van de woordenboekvormen aan verbogen of vervoegde woorden); POS-tagging (= het toevoegen van woordsoortinformatie); chunking (= het opdelen van de tekst in constituenten); parsing (= de syntactische opdeling van de constituenten). Met behulp van scripts kunnen die bewerkingen automatisch of semi-automatisch gebeuren.



The screenshot displays the ELAN 5.3 software interface for the file 'N72\_jeper.wav'. The top menu bar includes 'File', 'Edit', 'Annotation', 'Tier', 'Type', 'Search', 'View', 'Options', 'Window', and 'Help'. Below the menu is a toolbar with icons for grid, text, subtitles, lexicon, comments, recognizers, metadata, and controls. The main window is divided into several sections:

- Volume:** A slider set to 100, with a 'Mute' checkbox and a 'solo' radio button.
- Rate:** A slider set to 100.
- Selection:** A text box showing '00:24:29.505 - 00:24:31.882 2377' and playback controls (stop, previous, play, next, first, last, search, selection mode, loop mode).
- Waveform:** A visual representation of the audio signal with a red vertical line indicating the current selection point.
- Transcript:** A list of annotations with their corresponding time intervals and text:
  - N72\_sp1 [762]** (red text): 00:24:29.500 - 00:24:30.000, text: 'maar a#ik#ik mijn tanten horen klappen...'
  - N72\_sp1\_ndl [762]** (purple text): 00:24:30.500 - 00:24:31.500, text: 'maar als ik ik mijn tanten hoor klappen...'

Fig. 2: ELAN-scherf met een transcriptie voor Ieper (N<sup>72</sup>)

Frans-Vlaamse opnames te transcriberen als basis voor een syntactisch onderzoek. Ook de provincie Zeeland deed in 2019 een duit in het zakje, samen met de Zeeuwsche Vereeniging voor Dialectonderzoek voor de Zeeuws-Vlaamse verzameling. De provincies Oost- en West-Vlaanderen financierden in resp. 2018 en 2019 deelprojecten waarbij het transcriberen van een aantal geluidsoptnames een onderdeel was.

Door de ervaringen in deze deelprojecten kon het transcriptieproces verder op punt gesteld worden en kon ook begroot worden hoeveel tijd er nodig zou zijn om de hele collectie te transcriberen. De conclusie is hoe dan ook dat het transcriberen erg tijdsintensief is: een student heeft gemiddeld ongeveer een uur tijd nodig om 200 seconden dialectgeluid te transcriberen, en dan moet er nog eens een vrijwilliger die transcriptie nakijken, waarna een projectmedewerker de opmerkingen en correcties van de vrijwilliger digitaal doorvoert. Pogingen om het transcriptieproces te versnellen door via spraakherkenning – tests werden uitgevoerd met *automatic speech recognition* (ASR), *respeaking*, en *forced alignment* – een deel van het proces te automatiseren, leidden helaas niet tot een acceptabel resultaat (zie Ghyselen et al. 2020).

## 5. HET PROJECT GESPROKEN CORPUS VAN DE NEDERLANDSE DIALECTEN (GCDN)

In 2020 werd alle voorbereidend werk gehonoreerd door de toekenning van een grootschalig, vierjarig zogenaamd ‘middelzwaar infrastructuurproject’ door het FWO, gesteund door een consortium van instellingen (UGent, Meertens Instituut, Instituut voor de Nederlandse Taal, Universiteit Groningen en Variaties vzw.), dat onder leiding staat van Anne Breitbarth.<sup>(12)</sup>

<sup>(12)</sup> De promotoren van het project zijn Anne Breitbarth (UGent), Timothy Colleman (UGent), Katrien Depuydt (INT), Melissa Farasyn (UGent), Anne-Sophie Ghyselen (UGent), Els Lefever (UGent), Antal van den Bosch (Meertens Instituut), Jacques Van Keymeulen (UGent/Variaties vzw.) en Gertjan van Noord (Rijksuniversiteit Groningen). De praktische opleiding en

Het project sluit ook aan bij de CLARIN-doelstellingen, in die zin dat zoveel mogelijk al bestaande computersoftware gebruikt of aangepast zal worden.<sup>(13)</sup>

De bedoeling van het *Gesproken Corpus van de Nederlandse Dialecten* (GCND) is een digitaal doorzoekbaar en taalkundig geannoteerd corpus aan te leggen voor de dialecten van het zuidelijke Nederlands, d.i. Frans-Vlaanderen, Nederlandstalig België en de drie zuidelijke provincies van Nederland (Zeeland, Noord-Brabant en Limburg). Daartoe zal de hele ‘Gentse’ verzameling getranscribeerd worden, en zullen ook 38 aanvullende opnames gebeuren in Antwerpen, Vlaams-Brabant en Limburg.<sup>(14)</sup> Voor Nederland zal samengewerkt worden met het Meertens Instituut, dat over een uitgebreide collectie van vergelijkbare dialectopnames beschikt. De beoogde digitale infrastructuur moet in de toekomst volgende zaken mogelijk maken: (1) taalveranderingen geografisch in kaart brengen, (2) de functionaliteit van taalkenmerken (ook abstractere syntactische structuren) kwantitatief onderzoeken, en (3) nieuwe, voorheen onopgemerkte, en dus onbevraagde structuren opsporen.

## 6. HET TRANSCRIPTIEPROTOCOL

Hieronder volgt het transcriptieprotocol zoals dat aan studenten, stagiairs, jobstudenten en vrijwilligers wordt meegegeven. Deze versie werd het laatst bijgewerkt in juni 2020.

opvolging van de transcribenten binnen het GCND-project gebeurt door Lien Hellebaut.

<sup>(13)</sup> CLARIN (= Common Language Resources and Technology Infrastructure) is een Europese onderzoeksinfrastructuur die erop gericht is samenwerking te creëren op het gebied van computersoftware (zie <https://www.clarin.eu>).

<sup>(14)</sup> Het is waarschijnlijk nog net mogelijk om zegspersonen te vinden die een plaatselijk dialect als primaire taal hebben geleerd. Idealiter zijn die dan gesocialiseerd vóór de jaren 60 van de vorige eeuw, toen de taalverandering van dialect naar tussentaal een massaal karakter kreeg.

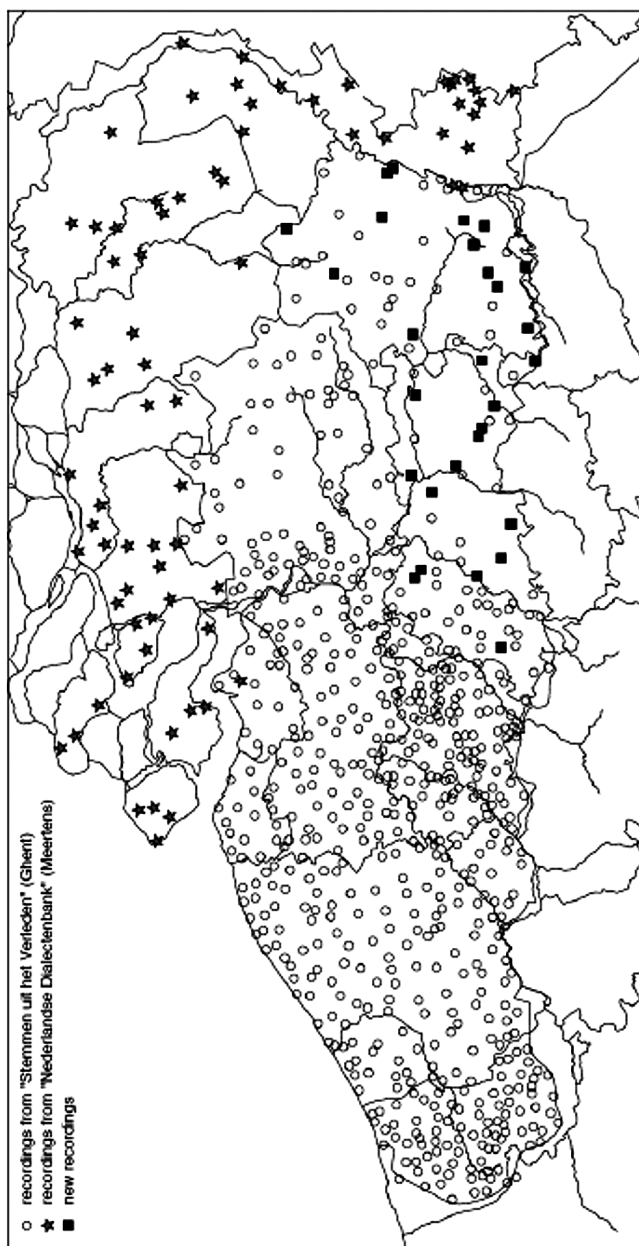


Fig. 3 Overzicht van de meetpunten voor het GCND-project

## A. Inleiding

Dit protocol is een handleiding bij het transcriberen van dialectmateriaal afkomstig uit de geluidsopnames van “Stemmen uit het verleden” (verzameling dialectbanden van UGent, [www.dialectloket.be](http://www.dialectloket.be)). Het protocol is gebaseerd op dat van de DIT-werkgroep (Dialect in Transcriptie) binnen het SAND-project.

Op de website [www.dialectloket.be](http://www.dialectloket.be) zijn niet enkel de oorspronkelijke opnames te vinden, maar ook korte samenvattingen van de banden en soms ook oude transcripties. Die transcripties zijn niet aangepast aan het protocol dat we hier beschrijven en variëren ook in accuratesse, maar ze zijn samen met de samenvattingen wel een handig hulpmiddel om moeilijk begrijpbare passages te doorgronden. Om de opnames te kunnen transcriberen, is het vaak ook noodzakelijk inhoudelijk onderzoekwerk te verrichten, bijvoorbeeld over historische gebeurtenissen, plaatsnamen of ambachten. Google is hier je vriend, alsook de wereldkennis van grootmoeder of grootvader. Het is natuurlijk niet de bedoeling dat je uren aan dit soort onderzoekwerk besteedt, maar een snelle zoekactie kan al vaak veel verhelderen.

Als je twijfelt over de correcte toepassing van de instructies in dit protocol, kijk dan eerst in de Beslissingendatabase of iemand anders al een vergelijkbaar probleem tegenkwam, en als dat niet zo is, voer je het probleem in als nieuw, te bediscussieren item in de database. Die items worden dan besproken door de projectbegeleiders. Neem niet in je eentje beslissingen. Voor de uniformiteit van de transcriptie is het van het grootste belang dat iedereen op dezelfde wijze transcribeert. Overleg is dus noodzakelijk. Een goede transcribent verdiept zich ook eerst in de grammatica van het te transcriberen dialect voor zover die beschikbaar is. Kom je tijdens het transcriberen vaak een constructie of wending tegen waarmee je niet goed weet wat te doen, ga dan op zoek in de vakliteratuur naar uitleg bij het fenomeen.

Als algemene stelregel geldt dat de data worden getranscribeerd in een vernederlandste vorm, op twee manieren: (a) sterk

naar het Nederlands toe (“zware vernederlandsing”), (b) sterk naar het dialect toe (“lichte vernederlandsing”, cf. infra). We transcriberen dus niet met fonetische tekens. Vernederlandsen betekent niet vertalen naar het Nederlands, maar wel abstractie maken van dialectische fonologie, morfologie of syntaxis. Maatstaf voor de spelling is het Groene Boekje. Als je dus *peert* of *pjèèt* hoort voor *paard*, dan transcribeer je *paard*.

Voor de transcriptie wordt het programma ELAN gebruikt. Hierin is het mogelijk meerdere lagen of *tiers* aan te maken, zodat bijvoorbeeld verschillende sprekers van elkaar onderscheiden kunnen worden. Op de lagen kun je grenzen aanbrengen, zodat de transcriptie in kleine stukjes wordt gehakt, met een maximale tijd van 10 seconden. Breng altijd bij het begin van een nieuwe zin een grens aan op de laag. Als een zin langer dan 10 seconden duurt moet je één of meer grenzen aanbrengen binnen de zin. Respecteer hierbij deelzingsgrenzen of woordgroepgrenzen.

Voor interviews waarin één spreker door één interviewer geïnterviewd wordt, gebruiken we vier lagen:

1. Interviewer (label bvb. ‘Int\_JD’) – Enkel lichte vernederlandsing
2. Informant – Lichte vernederlandsing (label = sprekercode zoals opgenomen in de database van de geluidsbestanden<sup>(15)</sup>, bvb. ‘HooIp\_I\_I\_ndl’)
3. Informant – Zware vernederlandsing (label = sprekercode + ‘\_ndl’, bvb. ‘HooIp\_I\_I\_ndl’)
4. Commentaar (label ‘com’)

Als er meerdere geïnterviewden zijn, voeg je per extra geïnterviewde een laag met lichte vernederlandsing en zware vernederlandsing toe. De twee vernederlandsingslagen zullen we in ELAN altijd met elkaar verbinden (de zware vernederlandsing – wat we de *referring tier* noemen – is afhankelijk van de lichte

<sup>(15)</sup> De sprekercode begint eerst met de Kloekecode voor de locatie van de opname, wordt eerst – indien er meerdere opnames zijn voor die locatie – gevolgd door het opnamenummer (zoals dat ook in de database opgenomen is bij ‘geluidsbestand’) en vervolgens door het label ‘1’ of ‘2’ (als er meerdere sprekers zijn).

vernederlandsing, de *parent tier*). Als er meerdere interviewers zijn, voeg je per extra interviewer een laag toe.

### *A.1. Interviewerlaag – enkel lichte vernederlandsing*

De interviewerlaag bevat de transcriptie van de taal van de interviewer(s). Die laag zal in een volgende fase NIET worden voorzien van grammaticale tags, Nederlandstalige lemma's of andere annotaties. Ze wordt enkel weergegeven in "lichte vernederlandsing" (cf. sectie B).

### *A.2. Informantlagen – lichte en zware vernederlandsing*

De informantlagen bevatten de transcriptie van de taal van de informant. Die lagen zullen in de volgende fase WEL worden voorzien van grammaticale tags, Nederlandstalige lemma's of andere annotaties. Indien er meer informanten zijn, worden per informant twee extra lagen (lichte + zware vernederlandsing) toegevoegd. De dialectische taal van de informant wordt in eerste instantie maar licht vernederlandst. Bij de 'lichte vernederlandsing' wordt enkel de dialectische fonologie vernederlandst; ze wordt dus weergegeven in Nederlandse spelling, alsof de woorden, vormen en constructies ook in het Standaardnederlands voorkomen. Bij de 'zware vernederlandsing' wordt niet alleen de fonologie, maar ook de morfologie vernederlandst. Enkel dialectische zinsbouw en dialectwoordenschat worden bewaard. Meer info over lichte en zware vernederlandsing volgt hieronder.

### *A.3. Commentaarlaag*

De commentaarlaag wordt gebruikt om opmerkingen te maken over slecht hoorbare delen van het interview of in andere gevallen waar je twijfelt over de transcriptie. Hier kun je eventueel ook resultaten van inhoudelijk onderzoekwerk kort rapporteren (bijvoorbeeld over namen, evenementen, culturele organisaties, ...).

## B. Lichte versus zware vernederlandsing

Zoals we in de inleiding al aanhaalden, werken we in dit project met vernederlandste transcripties, waarbij in variërende mate abstractie gemaakt wordt van dialectische fonologie, morfologie of syntaxis (afhankelijk van de vraag of een woord een functiewoord is of niet).

### B.1. Dialectuitspraak

Fonologische variaties op inhoudswoorden die ook in de standaardtaal bestaan, vernederlands je in zowel de lichte als de zware vernederlandsing. Je schrijft dus *boot* en niet *boët*. ‘Regelmattige’ fonologische variatie wordt sowieso vernederlandst. Bijvoorbeeld:

<i>Je hoort</i>	<i>Je schrijft (in zowel lichte als zware vernederlandsing)</i>
dinne, pit, stik	dun, put, stuk
duun, puut, stuuk	
meeln, deer, mies, brieg	molen, deur, mus, brug
derp, freet, sleetel	dorp, fruit, sleutel

### B.2. Dialectwoorden

Woorden die niet in het Standaardnederlands bestaan, spel je zowel de in de lichte als zware vernederlandsing alsof ze standaardtaalig zouden zijn. Als je *lattestoërs* (‘rolluiken’) hoort, schrijf je bijvoorbeeld *lattestoors*, ook in de zwaar vernederlandste versie (vertaal niet naar ‘rolluiken’!). Over de precieze vertaling van een woord is immers dikwijls discussie mogelijk.

Woorden die een lexicale variant (bvb. Franse versie) vormen van een woord dat in het Standaardnederlands bestaat (bvb. *een blom* i.p.v. *een bloem*, *frang* i.p.v. *frank*) – het gaat hier niet om een regelmatige fonologische variatie (cf. *boët* – *boot*, *groët* – *groot*) – noteer je in de lichte vernederlandsing ‘dialectisch’ volgens de Nederlandse spellingsregels (*blom*, *frang*). In de zware



vernederlandsing geef je het standaardtalige equivalent (*bloem*, *frank*).

Vaak is ook twijfel mogelijk over de beste manier om een dialectwoord te spellen. Volgende bronnen kunnen soelaas bieden:

- Raadpleeg eerst het WVD/WBD/WLD: [www.wvd.ugent.be](http://www.wvd.ugent.be) (zie: register), [www.e-wvd.be](http://www.e-wvd.be), [www.e-wbd.nl](http://www.e-wbd.nl), [www.e-wld.nl](http://www.e-wld.nl), [www.dialect.ruhosting.nl/wbd](http://www.dialect.ruhosting.nl/wbd), <http://www.dialect.ruhosting.nl/wld/>
- Als je het dialectwoord daar niet vindt: WNT: <http://gtb.inl.nl/search/>
- Als je het dialectwoord daar niet vindt: [www.woordenbank.be](http://www.woordenbank.be)

### B.3. Functionele elementen

Functionele elementen worden in de licht vernederlandste versie zo precies mogelijk getranscribeerd.

#### B.3.1. Een lijst van functionele elementen

We geven hier eerst – voor de duidelijkheid – een overzicht van functionele elementen.<sup>(16)</sup> Als een woord of morfeem niet tot de onderstaande elementen behoort, is het lexicaal en wordt het volgens de Standaardnederlandse orthografie getranscribeerd.

- Adposities (preposities en postposities, cf. *op*, *onder* en *naast*)
- Flexie: morfologische uitdrukking van kenmerken zoals getal, geslacht, persoon, definietheid, naamval, tijd, modus, trap (lijst is niet uitputtend).
  - Werkwoordsvervoeging (bvb. *ik speeldege* ‘ik speelde’).
    - Inclusief:
      - Prefix bij voltooide deelwoorden (*edaan* ‘gedaan’)

<sup>(16)</sup> In het lijstje maken we zoveel mogelijk gebruik van de terminologie van de Algemene Nederlandse Spraakkunst (ANS).

- Ook de ablaut die de verleden tijd van vele sterke werkwoorden markeert<sup>(17)</sup>, beschouwen we als functioneel element (bvb. *hij stoeg* ‘hij stond’).
- Voegwoordvervoeging (bvb. *dan* ‘k’ ‘dat ik’, *otn* ‘als hij’)
- Morfologie op lidwoorden, voornaamwoorden, adjectieven, nomina waarvan vermoed kan worden dat ze casus-, persoons-, en/of getalsdistincties aangeeft. (bvb. *ne vent* ‘een vent’).
- Ook sommige umlautverschijnselen<sup>(18)</sup> (bvb. *voet* vs. *vuut* ‘voeten’, *hoed* vs. *huuj* ‘hoeden’) geven getal aan. Om het eenvoudig te houden, transcriberen we umlaut enkel wanneer er zonder de umlautmarkering op geen enkele andere manier af te leiden valt dat we met een meervoudsvorm te maken hebben (bvb. *vuut*: *vuut* in lichte vernederlandsing en *voeten* in zware vernederlandsing). Andere vormen van umlaut worden niet getranscribeerd. De umlaut in bijvoorbeeld *menneke* (‘manneltje’) is op zich wel functioneel (als diminutiefmarkeerder: *man-menneke*), maar bij dat woord wordt de diminutief niet enkel door umlaut, maar ook door het diminutief *-ke* gemarkeerd. We transcriberen dus *manneke* in de lichte vernederlandsing en *manneltje* in de zware vernederlandsing. Zie hierover B.3.2. en B.3.3.

<sup>(17)</sup> Onder ablaut verstaan we “iedere vocaalwisseling binnen een paradigma of bij verwante woorden in een Indoëuropese, dus ook Germaanse taal, die in principe teruggaat op het Oerindoëuropees. (*In principe* is toegevoegd omdat een bepaald Ablautspatroom zich na de Oerindoëuropese periode analogisch kan hebben uitgebreid.) De Ablaut is, althans in oorsprong, een functionele afwisseling, d.w.z. de alternerende vocalen komen voor in vormen die semantisch en/of syntactisch verschillend zijn: vgl. got. *nima* ‘ik neem’ - *nam* ‘ik nam’ (semantisch verschil praesens - praeteritum); *nam* - *nenum* ‘wij namen’ (syntactisch verschil: sg.-plur.); *nima* - *nenum* (zowel semantisch als syntactisch verschil)” (Van Bree 2006: 146).

<sup>(18)</sup> Umlaut is een vorm van klinkerharmonie, waarbij een klinker kenmerken overneemt van de klinker in de erop volgende lettergreep in een andere vorm van een woordparadigma. Meer informatie over umlaut in de Nederlandse dialecten vind je bij Goossens (1981:20-43). Hij gaat in punt 2.4. ook in op de morfologische rol van umlaut (vormen van umlaut die functioneel relevant zijn).

- Hulpwerkwoorden (van modaliteit, aspect, tijd en passief), bvb. *hij komt ziek* ('hij wordt ziek')
- Lidwoorden, bvb. *e vent* ('een vent')
- Negatieve elementen, bvb. *noois* ('nooit')
- Voegwoorden, bvb. *dat, of, en en omdat*
- Enkelvoudige + samengestelde voorbijwoorden, zoals *er, hier, daar, ergens, overal en nergens, erbij*.
- Voornaamwoorden (pronomina)
  - Aanwijzende voornaamwoorden, bijvoorbeeld *dieje* en *dat*
  - Betrekkelijke voornaamwoorden, bijvoorbeeld *die, met wie* en *waarmee*
  - Bezittelijke voornaamwoorden, bijvoorbeeld *mijn* en *ons*
  - Onbepaalde voornaamwoorden, bijvoorbeeld *iemand* en *iets*
  - Persoonlijke voornaamwoorden, bijvoorbeeld *hij* en *ik*
  - Reciproke voornaamwoorden, bijvoorbeeld *elkaar, mekander*
  - Reflexieve voornaamwoorden, bijvoorbeeld *zich*
  - Uitroepende voornaamwoorden, bijvoorbeeld *Wat een geld! Zo stom was dat! Zulke rare mensen!*
  - Vragende voornaamwoorden, bijvoorbeeld *wie, wat, waar*

### *B.3.2. Richtlijnen voor lichte vernederlandsing van functionele elementen*

Bij functionele elementen wordt bij de lichte vernederlandsing dicht bij het dialect getranscribeerd:

- Klanken die de sprekers weglaten, transcriberen we niet. Je schrijft bijvoorbeeld *wa, maa, daa, hij kom morgen* en niet *wat, maar, daar* of *hij komt morgen*.
- Als er één of meer klanken aanwezig zijn die in de Standaardnederlandse spelling ontbreken, dan worden die klanken wel genoteerd. Je schrijft bijvoorbeeld *ik zwemmen* en niet *ik zwem*.
- Regelmatige wisselingen in het vocalisme (bvb. *moa* i.p.v. *maar*) transcribeer je niet (lichte vernederlandsing: *maa*), enkel wisselingen in het vocalisme die morfologisch functioneel zijn (bijvoorbeeld umlaut of werkwoordsvervoeging).

- Niet-standaardtalige clitische elementen, i.e. functionele elementen die samensmelten met inhoudswoorden (bvb. *tkind* voor *het kind* of *zijme* voor *zijn we*) schrijf je in de lichte vernederlandsing als één geheel, maar je markeert de verschillende onderdelen met een #-teken (bvb. *zij#me weg*). Dat is belangrijk om in een latere fase de juiste tags aan de juiste elementen van de cluster te kunnen toekennen. Wanneer een clitisch element echter ook in de standaardtaal/schrijftaal verankerd zit (bvb. de vorm *'s avonds*), dan gebruik je de standaardtalige schrijfwijze in zowel de lichte als de zware vernederlandsing (dus *'s avonds* en niet *s#avonds*).
- Een speciaal geval vormen clitische structuren als *ja'ndoet* (eigenlijk: *ja het en doet*, 'toch wel', vaak te horen in West- en Oost-Vlaanderen). Die vorm transcribeer je in lichte vernederlandsing als *ja#t#n#doet*<sup>(19)</sup> en in de zware als *ja het en doet*. Op dezelfde manier transcribeer je ook constructies als *ja##k#en#doe* (in zware vernederlandsing: *ja ik en doe*) en *gij#n#doet* (in zware vernederlandsings: *jij en doet*).
- Vaak vinden we in de dialecten in Vlaanderen ook clitisering bij subjectverdubbeling of –verdriedubbeling (*hij zei dakik te traag was; hij zei damewij te traag waren*). Hier spreken we af alle pronomina in de lichte vernederlandsing met elkaar te verbinden (*da##k#ik, da#me#wij*)<sup>(20)</sup> en die vervolgens in de zware vernederlandsing splitsen (*dat ik ik, dat we wij*).
- Zorg dat je het pronominale systeem van het te transcriberen dialect snapt. Consulteer voor de 'optimale' transcriptie van voornaamwoorden ofwel (1) De Vogelaer, Gunther (2008). *De Nederlandse en Friese subjectsmarkeerders: geografie, typologie en diachronie*. Gent: Koninklijke Academie voor Nederlandse taal- en letterkunde, (2) Goossens, Jan (1981-1994).

<sup>(19)</sup> Vaak wordt het *t*-elementje als glottisslag gerealiseerd, maar die regelmatige klankwisseling hoef je niet te markeren; we herstellen de eigenlijke 't'.

<sup>(20)</sup> Op zich zijn er ook argumenten om de *wij* – op zich een volle vorm – apart te schrijven, maar voor de eenduidigheid spreken we af bij postverbale verdubbeling de pronomina in één cluster weer te geven.

*Sprachatlas des nördlichen Rheinlands und des südöstlichen Niederlands : Fränkischer Sprachatlas.* Marburg: Elwert. Het tweede volume van die atlas gaat over subjecten en subjectverdubbeling, of (3) een grammatica van het dialect dat je transcribeert.

### *B.3.4. Richtlijnen voor zware vernederlandsing van functionele elementen*

In de zware vernederlandsing:

- Worden de standaardtalige functiewoorden weergegeven (*ge* wordt *je*, *gij* wordt *jij*, *etwie* wordt *iemand*). *Gij* vernederlands je in de zware vernederlandsing altijd tot *jij*, ook al is er op zich altijd ambiguïteit mogelijk met het standaardtalige *u*; door de vorm als *jij* te vernederlandsen, kunnen we immers de subjectsvorm van de 2<sup>e</sup> persoon enkelvoud desambigueren van de objectsvorm in dezelfde persoon (bvb. in *gij moet u niet haasten*);
- Worden enclitische elementen gesplitst en verder vernederlandsst (*kom#je#gie* wordt *kom je jij*);
- Worden de volle vormen van gereduceerde elementen genoteerd (bvb. *k#speel* wordt *ik speel* of *j#is* wordt *hij is*);
- Wordt dialectische ablaut vervangen door de pendant ervan in het AN (bvb. *miek* wordt *maakte*);
- Worden niet-standaardtalige morfologische structuren vernederlandsst (bvb. *ik werktege* > *ik werkte*). Ook congruentie van betrekkelijke voornaamwoorden (*de man dat dat zei* > *de man die dat zei*) en het genus van woorden (*de deeg* > *het deeg*) beschouwen we als ‘morfologisch’ – al is daarover discussie mogelijk – en vernederlandsen we in de zware vernederlandsing;
- Laat je niet-standaardtalige syntactische structuren intact (bvb. expletief *dat*, doorbreking werkwoordelijke eindgroepen, subjectverdubbeling, *komen* als koppelwerkwoord, *voor* + te-infinitief, *lijk/bij*+ subjectsvorm i.p.v. objectsvorm of omgekeerd, hulpwerkwoord *hebben* i.p.v. *zijn*, inversieloosheid, negatie, *als* bij comparatief i.p.v. *dan*).

## B.4. Voorbeelden

		<i>Je hoort</i>	<i>Je schrijft</i>	
			<i>Lichte vernederlandsing</i>	<i>Zware vernederlandsing</i>
Niet-functionele elementen	Fonologische variaties op woorden die ook in ST bestaan	boot schole <sup>(21)</sup> speeln	boot schole spelen	boot school spelen
	Lexicale varianten van een ST-woord	frang blom	frang blom	frank bloem
	Woorden die niet in de ST bestaan	sjchuuflette	schuiflette	schuiflet
Functionele elementen - morfologisch	Lidwoorden	tkind e kind ne vent	t#kind e kind ne vent	het kind een kind een vent
	Pronomina	gaa, gie, zie, aa, en, eur, zin, wuir, wider, wellie, us, gider, gunder, guir, zider, zuir, uir	gij, gj, zij, hij, en, haar, zijn, wuir, wijder, wijlie, uns, gijder, gunder, guir, zijder, zuir, huir	jij, zij, hij, haar, wij, ons, jullie, zij, hun

<sup>(21)</sup> Dit is een niet-suffigale sjwa. Die bewaren we in de lichte vernederlandsing, want ze zit op grens tussen fonologisch en morfologisch element.

	Clitische elementen	tisier dase komt kweetniewie daterkomt waaristje? dakik damewunder toetoet <sup>(22)</sup> tee jandoet jakendoe danze komen	t#is ier da#se komt k#weet nie wie dat er komt waar is#tje? da#k#ik da#me#wunder t#doet t#doet t#ee ja#t#n#doet ja#k#en#doe dan#ze komen	het is hier dat ze komt ik weet niet wie dat er komt waar is hij? dat ik ik dat we wij het doet het doet het heeft ja het en doet ja ik en doe dat ze komen
	Diminuering	blommeke	blommeke	bloemetje of bloempje
	Vervoeging	kzwemmen dank onk ik zijn ziek kwerktege de media weet	k#zwemmen dan#k an#k ik zijn ziek k#werktege de media weet	ik zwem dat ik als ik ik ben ziek ik werkte de media weten
	Verbuiging	ne gruuëtn twee broeks	ne groten twee broeks	een grote twee broeken
	Functionele umlaut (enkel wanneer die functioneel is, bvb. als de umlaut de enige meervouds-markeerder is)	vuut menneke	vuut manneke	voeten mannotje
	Ablaut	vrieg vraagde	vrieg vraagde	vroeg vroeg
	Genus	de deeg	de deeg	het deeg
	Congruentie betrekkelijke voornaamwoorden	het kind <u>die</u> kwam	het kind <u>die</u> kwam	het kind <u>dat</u> kwam

<sup>(22)</sup> Bijvoorbeeld in de zin: Hij komt toch niet? Toetoet!

Functionele elementen - syntactisch	Woordvolgorde	gisteren ie gingk toe duinkerke. tmakt de menschen z'adden geen geld	gisteren ij ging toe Duinkerke. het maakt de mensen ze hadden geen geld.	gisteren hij ging tot Duinkerke. het maakt de mensen ze hadden geen geld.
	Negatie	ien wist van niet.	ij#n wist van niet.	hij en wist van niets.
	Objectsvorm vs. subjectsvorm	kom je bie <u>min</u> ? zij is kleiner dan <u>mij</u> .	kom je bij <u>mijn</u> ? zij is kleiner dan <u>mij</u> .	kom je bij <u>mijn</u> ? zij is kleiner dan <u>mij</u> .
	Constructies met voegwoorden/ voorzetsels	hij deet <u>voor</u> op te vallen. hij is groter <u>als</u> ik.	hij deed#t voor op te vallen. hij is groter <u>als</u> ik.	hij deed het <u>voor</u> op te vallen. hij is groter <u>als</u> ik.

Int_VV	dialectopname Ieper. achttien november zevenenzestig.
Int_V	xxx.
No72p_I_I	serieus?
No72p_I_I_ndl	serieus?
Int_V	wel ja.
No72p_I_I	a ja.
No72p_I_I_ndl	a ja.
Int_V	uh.. oe lange woon je..
Int_V	woon je gij ier..
Int_V	e je gie al assan gewoond in Ieper?
No72p_I_I	ja ja#k.
No72p_I_I_ndl	ja ja ik.
No72p_I_I	ja. k#zijn ier geboren.
No72p_I_I_ndl	ja. ik ben hier geboren.
Int_V	a ja?
No72p_I_I	ja#k
No72p_I_I_ndl	ja ik.
No72p_I_I	k#zijn ik een echten Ieperling eni.
No72p_I_I_ndl	ik ben ik een echte Ieperling eni.
Int_V	zijn..
No72p_I_I	ma ja ze klappen al nie lijk ik zeker?
No72p_I_I_ndl	maar ja ze klappen al niet gelijk ik zeker?



No72p\_I\_I k#weten#t#ik nie.  
 No72p\_I\_I\_ndl ik weet het ik niet.  
 Int\_V zijn der nog vele die.. die zoun Iepers spreken?  
 No72p\_I\_I ah bah. ja.  
 No72p\_I\_I\_ndl ah bah ja.  
 No72p\_I\_I je moe ze weten wonen eni.  
 No72p\_I\_I\_ndl je moet ze weten wonen eni.  
 Int\_V mmm.  
 No72p\_I\_I a ja.  
 No72p\_I\_I\_ndl a ja.  
 Int\_V a ja. uh.  
 Int\_V t#schijnt da#ze in..  
 Int\_V gow.  
 Int\_V oe komt da da eigenlijk Sint-Pieters dat da t#echt Iepers is.. t#Iepers is.  
 No72p\_I\_I ma j#et er nog van Sint-Pieters wi da#t uh..  
 No72p\_I\_I\_ndl maar je hebt er nog van Sint-Pieter wi dat het uh..  
 No72p\_I\_I da#ze algelijk nog ezo nie en klappen lijk ik.  
 No72p\_I\_I\_ndl dat ze algelijk nog zo niet en klappen gelijk ik.  
 Int\_V a ja.  
 Int\_V da was ier toen de meeste beroemde parochie of zo of..  
 No72p\_I\_I ah ba k#weten ik nie oe dat dat komt.  
 No72p\_I\_I\_ndl ah ba ik weet ik niet hoe dat dat komt.  
 No72p\_I\_I ah ba al van werkmensen eni.  
 No72p\_I\_I\_ndl ah ba al van werkmensen eni.  
 Int\_V a ja.  
 No72p\_I\_I a ja.  
 No72p\_I\_I\_ndl a ja.  
 Int\_V en vo n#oorloge uhm.  
 Int\_V ja j#e gij nog die tijd gekend wel zeker?  
 No72p\_I\_I ma ja#k goddomme.

No72p\_I\_I\_ndl maar ja ik goddomme.  
 No72p\_I\_I k#en twee oorlogen meeëmaakt.  
 No72p\_I\_I\_ndl ik heb twee oorlogen meegemaakt.  
 Int\_V ah ja. en vo n#oorloge wuk spraken ze  
 toene? was#t uh?  
 No72p\_I\_I ah bah.  
 No72p\_I\_I\_ndl ah bah.  
 Int\_V was#t ook tzel.. gow.  
 Int\_V tgeen da je gie nu spreekt of?  
 No72p\_I\_I awel.  
 No72p\_I\_I\_ndl awel.  
 No72p\_I\_I ah ba ja nee.  
 No72p\_I\_I\_ndl ah ba ja nee.  
 No72p\_I\_I en j#ad er toen veel ook die t#Frans ook  
 kosten eni.  
 No72p\_I\_I\_ndl en je had er toen veel ook die het Frans  
 ook konden eni.  
 Int\_V a ja.  
 No72p\_I\_I ja.  
 No72p\_I\_I\_ndl ja.  
 Int\_V veel die Frans kosten.  
 No72p\_I\_I ja.  
 No72p\_I\_I\_ndl ja.  
 Int\_V a ja. ku#je e beetje vertellen van  
 n#oorloge?  
 No72p\_I\_I a ja. van n#oorloge.  
 No72p\_I\_I\_ndl a ja. van de oorlog.  
 No72p\_I\_I t#is te zien va wufferen oorloge.  
 No72p\_I\_I\_ndl het is te zien van welke oorlog.  
 Int\_V van n#eersten oorloge.  
 No72p\_I\_I ah van n#eers..  
 No72p\_I\_I\_ndl ah van de Eerst..  
 Int\_V ja.  
 No72p\_I\_I k#waren ik acht jaar oud.  
 No72p\_I\_I\_ndl ik was ik acht jaar oud.  
 Int\_V je was acht jaar oud?  
 No72p\_I\_I acht jaar.

No72p\_I\_I\_ndl acht jaar.

Int\_V ah..

No72p\_I\_I en k#weten ik da#m#inder evlucht en na Frankrijk eni.

No72p\_I\_I\_ndl en ik weet ik dat we wij gevlucht hebben naar Frankrijk eni.

Int\_V a ja.

No72p\_I\_I e.. en e beetje overalsten ezeten.

No72p\_I\_I\_ndl e.. en een beetje overal gezeten.

No72p\_I\_I m#en eerst ezeten in n#Normandie.

No72p\_I\_I\_ndl we hebben eerst gezeten in de Normandie.

Int\_V mmm.

No72p\_I\_I en m#en toen naa Parijs egaan.

No72p\_I\_I\_ndl en we hebben toen naar Parijs gegaan.

No72p\_I\_I en van Parijs..

No72p\_I\_I\_ndl en van Parijs

No72p\_I\_I naa Poitiers.

No72p\_I\_I\_ndl naar Poitiers.

Int\_V oh t#is al verre.

No72p\_I\_I ja ja#t.

No72p\_I\_I\_ndl ja ja het.

No72p\_I\_I en m#en in Rouen ewist ook.

No72p\_I\_I\_ndl en we hebben in Rouen geweest ook.

No72p\_I\_I m#en langst van al in Rouen ewist.

No72p\_I\_I\_ndl we hebben het langst van al in Rouen geweest.

No72p\_I\_I m#en daa vier jaar ewist.

No72p\_I\_I\_ndl we hebben daar vier jaar geweest.

No72p\_I\_I en van Rouen zij#me toen wer gekomen naar ier.

No72p\_I\_I\_ndl en van Rouen zijn we toen weergekomen naar hier.

Int\_V al te voete?

No72p\_I\_I nee goddomme.

No72p\_I\_I\_ndl nee goddomme.

Int\_V nee?

No72p\_I\_I ma ba neen.

No72p_I_I_ndl	maar ba neen.
No72p_I_I	nee.
No72p_I_I_ndl	nee.
No72p_I_I	alleszins nie.
No72p_I_I_ndl	alleszins nie.
Int_V	xxx. met d#hondenkar of xxx..
No72p_I_I	awel me n#treng eni.
No72p_I_I_ndl	awel met de trein eni.
Int_V	me n#treng.
No72p_I_I	met n#treng ja.
No72p_I_I_ndl	met de trein ja.
Int_V	aja.
Int_V	de trengs reden toen nog?
No72p_I_I	a ja#s.
No72p_I_I_ndl	a ja ze.
No72p_I_I	a ja#s.
No72p_I_I_ndl	a ja ze.
Int_V	was#t voo..
No72p_I_I	xxx m#en wijder uh..
No72p_I_I_ndl	xxx we hebben wij uh..
No72p_I_I	m#en wijder zeker als vijf jaar wegewist wi.
No72p_I_I_ndl	we hebben wij zeker als vijf jaar wegeweest wi.
No72p_I_I	ah uh..
No72p_I_I_ndl	ah uh..
No72p_I_I	met n#oorloge veertien achttiene.
No72p_I_I_ndl	met de oorlog veertien achttien.

### C. Lees- en andere tekens die geen klanken representeren

*Hoofdletters:* Omdat het bij gesproken taal – gekarakteriseerd door veel eenwoordzinnen, haperingen, hervattingen, ... – niet altijd evident is te bepalen waar een zin begint, gebruiken we in de transcripties geen hoofdletters om een zinsbegin te markeren. Wel geven we aan plaats- en eigennamen hoofdletters zoals we dat ook in de standaardtaal zouden doen.

*Leestekens:* We geven aan de hand van leestekens aan waar een zin onderbroken of beëindigd wordt. Elke annotatie moet eindigen met een punt, een vraagteken of een beletselteken. Een punt (en een vraagteken) wordt enkel voor een zinseinde gebruikt, dus niet bij letters (dus niet *A.B.N.* maar wel *ABN*) of cijfers (dus niet *1.000*, maar wel *duizend*). Verder worden geen andere leestekens gebruikt, dus geen komma's, aanhalingstekens, uitroepetekens, dubbele punten, enz. Vermijd ook zo veel mogelijk afkappingstekens (niet '*k ben*' maar *k#ben* in de lichte vernederlandsing). Alle getallen worden voluit geschreven met letters.

*Lijst van overige toegestane tekens:*

- ... = Beletselteken, aan het begin of eind van incomplete zinnen
- ??? = Geeft aan dat je dialectkennis onvoldoende is om te begrijpen wat er gezegd wordt
- xxx = Geeft een moeilijk verstaanbare constituent, deelzin of passage aan (bvb. door lawaai, over elkaar spreken, ...)
- xx = Geeft een moeilijk verstaanbaar woorddeel aan
- ggg = Geeft duidelijk hoorbare sprekergeluiden aan, zoals gelach, gehuil, gegil, gekuch, gekir etc.

Wanneer je niet begrijpt wat er gezegd wordt, noteer je '???'. Een vrijwilliger probeert die vraagtekens achteraf op te lossen. Wanneer je moeite hebt een bepaalde passage te begrijpen, is het belangrijk rekening te houden met de context. Als je denkt een bepaald woord of een bepaalde zin te horen, denk dan steeds na: kan dat in deze context? Wanneer je twijfelt, gebruik je liever '???'. Een voorbeeld: een van onze transcribenten kende het woord *hofstede* niet ('boerderij') wanneer er in een opname gesproken werd over *drie hofsteden die moesten wijken voor de aanleg van een autostrade*. De transcribent meende *hoofdstad* te horen en transcribeerde dat zo. Natuurlijk, als je even de context in rekening brengt, merk je meteen dat het wel zeer onwaarschijnlijk is dat er drie hoofdsteden zouden moeten wijken voor de aanleg van een snelweg. Om maar te zeggen: context is belangrijk; gebruik bij twijfel liever '???'.

*Diakritische tekens:* Gebruik diakritische tekens alleen waar die volgens het Groene Boekje verplicht zijn, bijvoorbeeld wel *geëindigd* of *'s avonds* maar niet *bök*. Als *één* wordt uitgesproken met een [e:] (en niet met een sjwa zoals bij het lidwoord *een*), dan schrijf je de vorm met twee accenten, net als het telwoord (als *één* dus).

Opmerking over *'s avonds/s ochtends*: omdat het afkappings-teken hier geconventionaliseerd is, hoef je geen hekje-tekens te gebruiken om de clitisering aan te duiden (dus niet *s#avonds*). Uitspraakvarianten als *snaves* of *saves* vernederlands je in zowel de lichte als zware vernederlandsing tot *'s avonds*; het gaat hier immers om uitspraakvariatie bij een inhoudswoord. Wanneer een respondent het echter over *tsaves* of *tsochtes* heeft, dan reconstrueert hij eigenlijk een deel van de oude genitiefvorm *des ochtends*. Aangezien die gedeeltelijke reconstructie van *des* veeleer functioneel dan lexicaal is, geven we het verschijnsel wel weer in de lichte vernederlandsing (met hekje-tekens):

Je hoort	Lichte vernederlandsing	Zware vernederlandsing
<i>saves</i>	<i>'s avonds</i>	<i>'s avonds</i>
<i>snavonds</i>	<i>'s avonds</i>	<i>'s avonds</i>
<i>tsaves</i>	<i>ts#avonds</i>	<i>'s avonds</i>
<i>tsochtes</i>	<i>ts#ochtends</i>	<i>'s ochtends</i>

## D. Tussenwerpsels

De tabel hieronder kan als leidraad dienen bij de spelling van tussenwerpsels.

ah	goh	oho
aha	ha, haha	poeh
ah ja	hé, hè, hei	pst
awel	ho, hu	sjt, sst
allez	hum	tut („tut tut“)
ai	jee („o jee“)	uh, uhm

au	mm-hu	uhu
bah	mmm	wauw
boe	neeë (cf. West-Vlaams)	who
bwa	nou	zuh, zulle, zunne, zun
bè (ja)	oeh, oei	wi (vnl. in West-Vlaams)
ei, eikes	oesje	zi
eni (vnl. in streek Ieper)	oh, o	zuh

### E. Versprekingen

Versprekingen noteer je gewoon zoals je ze hoort (volgens de hierboven al gespecificeerde richtlijnen); het is niet aan de transcribent te noteren wat de informant eigenlijk echt bedoelde.

### F. Dialectische plaats- en persoonsnamen

Zoals al aangegeven in punt C., geven we aan plaats- en eigennamen een hoofdletter zoals we dat ook in de standaardtaal zouden doen. Wanneer een spreker een dialectische variant gebruikt van een plaatsnaam die gerelateerd is aan de standaardtalige naam, dan behandel je die als een lexicale variant. Bijvoorbeeld:

Je hoort	Lichte vernederlandsing	Zware vernederlandsing
<i>deizel</i>	<i>Deizel</i>	<i>Dadizele</i>
<i>de madoëne</i>	<i>De Madone</i>	<i>De Madonna</i>
<i>sleine</i>	<i>Sleine</i>	<i>Sleidinge</i>

Gebruikt een respondent daarentegen een bijnaam die niets met de standaardtalige of huidige naam te maken heeft (bvb. *Sinte-Pier* voor *Nieuwvliet*), dan behoud je in zowel lichte als zware vernederlandsing de dialectische bijnaam.

## G. Veelvoorkomende transcriptiefouten

### F.1. Daar > ter

	Lichte vernederlandsing	Zware vernederlandsing
Fout	<i>is#t#er ier iemand?</i>	<i>is het er hier iemand?</i>
Juist	<i>is#ter ier iemand?</i>	<i>is er hier iemand?</i>

*Toelichting:* De *t* maakt deel uit van de zwakke vorm *ter* < *der* < *daar*, waaruit *er* ontstaan is.

### F.2. Verbindingsklanken

	Lichte vernederlandsing	Zware vernederlandsing
Fout	<i>hoe ga#g#het</i> <i>hoe ga#d#het</i>	<i>hoe gaat het?</i>
Juist	<i>hoe ga#get</i> <i>hoe ga#det</i>	<i>hoe gaat het?</i>

De *-g* is hier een verbindingsklank. Of die voor of na het clitiseringsteken moet komen, is eigenlijk een zinledige discussie; we plaatsen die hier vrij arbitrair na het hekje. Omdat het hier om functionele elementen gaat, behouden we de verbindingsklank in de lichte vernederlandsing, maar we standaardiseren ze in de zware vernederlandsing.

### F.3. Al te doorgedreven vernederlandsing

	Lichte vernederlandsing	Zware vernederlandsing
Fout	<i>ik ga ne keer iets doen</i>	<i>ik ga eens iets doen.</i>
Juist	<i>ik ga ne keer iets doen</i>	<i>ik ga een keer iets doen.</i>

*Toelichting:* Blijf bij de formulering van de respondent.

### F.4. Clitiseringsteken voor reductie binnen woord

	Lichte vernederlandsing	Zware vernederlandsing
Fout	<i>da is t#zelfde</i>	<i>dat is hetzelfde.</i>
Juist	<i>da is tzelfde</i>	<i>dat is hetzelfde.</i>



*Toelichting:* Met een # koppel je twee woorden aan elkaar. Er moeten dus ook twee woorden zijn in de zware vernederlandsing. L

### Bibliografie

- BARBIERS, Sjef (2005), De Syntactische Atlas van de Nederlandse Dialecten. In: *Taal en Tongval. Themanummer 18*. pp. 7-40.
- BLANCQUAERT, E. (1948), *Na meer dan 25 jaar Dialect-onderzoek op het Terrein*. Tongeren, Michiels. [Koninklijke Vlaamse Academie voor Taal- en Letterkunde III, 28]
- COUSSÉ, E., S. GILLIS, H. KLOOTS en M. SWERTS (2004). The influence of the labeller's regional background on phonetic transcriptions: Implications for the evaluation of spoken language resources. In: M. Lino, X. Xavier, F. Ferreira, R. Costa en R. Silva (red.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation*. Parijs: ELRA, 1447-1450.
- FARASYN, M., A-S. GHYSELEN, J. VAN KEYMEULEN en A. BREITBARTH (2020), Challenges in tagging and parsing spoken dialects of Dutch. Manuscript, UGent.
- GHYSELEN, A-S., A. BREITBARTH, M. FARASYN, J. VAN KEYMEULEN en A. VAN HESSEN (2020), Clearing the transcription hurdle in dialect corpus building: The corpus of Southern Dutch dialects as case-study. In: *Frontiers in Artificial Intelligence 3*, 1-17. <https://www.frontiersin.org/articles/10.3389/frai.2020.00010/full>
- RENSINK, W.G. (1962), Dialecten op de band. In: *Taal en Tongval 14*. blz. 184-196.
- VANACKER, V. F. en G. DE SCHUTTER (1967), Zuidnederlandse dialecten op de band. In: *Taal en Tongval 19*. blz. 35-51.
- VAN KEYMEULEN, J. (1975), *De syntaxis van het Wichels dialect*. (onuitgegeven licentiaatsverhandeling UGent).
- VAN KEYMEULEN, J. (2020), De digitale apparaten voor de dialectstudie aan de Gentse Universiteit. In: *Taal en Tongval 72:1* (ter perse); blz. 17-37.
- VAN KEYMEULEN, J., V. DE TIER, A. BREITBARTH, A. GHYSELEN en M. FARASYN (2019), Het dialectologische corpus "Stemmen uit het Verleden" van de Universiteit Gent. In: *Volkskunde 120*; blz.193-204.

Anne-Sophie GHYSELEN,  
Jacques VAN KEYMEULEN, Melissa FARASYN,  
Lien HELLEBAUT en Anne BREITBARTH