

# Reeks Nederlandse Dialectatlassen

## Content

File(s)	Content
info.pdf	This file.
Heeringa_2001.pdf, Heeringa_2004 pp. 214-226.pdf	Publications about this data set
RNDstats.xlsx	Table that gives for each language/local variety the number of available word transcriptions
RNDascii.xlsx, RNDxampa.xlsx, RNDutf8.xlsx	RND transcriptions in original coding, X_SAMPA and UTF-8.

## Data set

The Reeks Nederlandse Dialectatlassen (RND) is a series of atlases covering the Dutch dialect area. The Dutch dialect area comprises the Netherlands, the northern part of Belgium, a smaller northwestern part of France and the German county Bentheim.

In the RND, the same 141 sentences are translated and transcribed in phonetic script for each dialect. Since digitizing the phonetic texts is time-consuming, we selected only 125 words from the text (for some local dialects 166 words). These words were digitized for a subset of 361 local dialects. Around 2001 Standard Dutch and Standard German were added.

For more information the user of this data set is referred to two publications:

- Wilbert Heeringa (2001). De selectie en digitalisatie van dialecten en woorden uit de Reeks Nederlandse Dialectatlassen. *TABU, Bulletin voor Taalwetenschap*, volume 31, number 1/2, pp. 61-103.
- Wilbert Heeringa (2004). Measuring dialect pronunciation differences using Levenshtein distance. Doctoral dissertation University of Groningen.

Where differences exist between the first and the second publication, the second publication overrules the first one.

At the time the first paper was written, the data set consisted of a small set of 350 local dialects which were spoken in 347 locations. At the time the second paper was written, the data set consisted of 360 local dialects that were spoken in 357 locations. Around 2008 the local dialect of Amsterdam as spoken in the Jordaan was added.

For the two standard languages and for 137 local dialects transcriptions are available for a set of 166 words. For the remaining 226 local dialects transcriptions are available for a smaller set of 125 words. The Excel tables that contain the data are composed of 363 rows (two standard languages and 361 local dialects) and 166 columns. Each column contains the transcriptions of a word. If only 125 words are available for a dialect, in the cells of the missing (166-125=) 41 words a '|' is found, indicating that no transcription is available. In the file `RNDstats.xlsx` it is indicated whether a

local dialect contains the larger of smaller set of word transcriptions, and the number of available word transcriptions is given.

## **Codings**

The RND sentences were originally transcribed in a dialect of the International Phonetic Alphabet. When digitizing the 166/125 words of the 361 local dialects, a dialect of X-SAMPA was used. The file `RNDascii.xlsx` is an Excel spreadsheet that contains the data in this 'original' coding. For an explanation of the coding see the Appendix (in Dutch) below. The file `RNDxsampa.xlsx` contains the data in the X-SAMPA coding. The file `RNDutf8.xlsx` contains the data being coded in Unicode standard UTF-8.

Not all diacritics used in RND IPA could be mapped to an exact equivalent in X-SAMPA and Unicode. For example, the strike-through tilde accent indicating a sound being 'half nasalized' is not available in X-SAMPA/IPA. In those cases the closest X-SAMPA/IPA equivalent was used or the diacritic was left out (for example, the diacritics indicating fortis and lenis). For a comparison between the original coding and X-SAMPA see 'Diacritische tekens' in the Appendix.

Regardless the coding, missing transcriptions are indicated by a '|'.

## **Inconsistencies**

In section 9.1.4 some inconsistencies in the RND material are discussed. The sections 9.1.4.1-9.1.4.5 describe how those inconsistencies were processed when applying dialectometric analyses to the data. For the sake of completeness, we point out that the adaptations mentioned in these sections are not implemented in the RND data available in this repository.

## **Permission**

The data is made available with permission of De Sikkel publishing house. We are grateful to the publisher for this permission. In 2000, publisher De Sikkel and the educational department of Standaard Uitgeverij were taken over by Groupe De Boeck and brought together under the name Uitgeverij De Boeck. That was in turn taken over in June 2016 by publishing house Van In, see: <https://www.vanin.be/nl>.

## **Acknowledgements**

We thank Saakje van Dellen, Rogier Nieuweboer, Marcus Bergman and Johan Dijkhuis who digitized the greater part of the selected RND transcriptions which are, as a fruit of their labor, available in this repository.

## Appendix

### Klinkers

ASCII	X-SAMPA	uitspraak	opmerking
i	i	als in Ned. bieden	
I	I	als in Ned. bidden	
e.	e:\	als in Ned. beek	
e:/I.	e:/I:\	als in Ned. zeer	
E	E	als in Ned. bek	
Q	{	als in Eng. bad	
a.	a:\	als in Ned. baak	
A	A	als in Ned. bak	
D	Q	als in Eng. law	
)	0	als in Ned. ros	
0	U	als in Ned. Tom/ton/tong	
o.	o:\	als in Ned. boos	
o:/0.	o:/U:\	als in Ned. boor	
u	u	als in Ned. boek	
y	y	als in Ned. nu	
0	2	als in Ned. beuk	
0:/U.	2:/Y:\	als in Ned. beur	
Y	9	als in Ned. freule	
Y	9	als in Ned. lui	eerste lid tweeklank
U	Y	als in Ned. bus	
@	@	als in Ned. de	

## Medeklinkers

ASCII	X-SAMPA	uitspraak	opmerking
c	c	als in Ned. rietje	
p	p	als in Ned. paard	
b	b	als in Ned. bal	
t	t	als in Ned. tak	
d	d	als in Ned. dak	
k	k	als in Ned. kar	
K	g	als in Eng. goal	
m	m	als in Ned. maart	
n	n	als in Ned. nacht	
N	N	als in Ned. tingeling	
J	J	als in Ned. oranje	
l	l	als in Ned. les	
r	r	als in Ned. raar	trillingen d.m.v. tongpunt
R	R\	als in Ned. raar	trillingen d.m.v. huig
f	f	als in Ned. fout	
v	v	als in Ned. vat	
W	v\ of P	als in Ned. water	labiodendaal
w	w	als in Eng. wow	bilabiaal
s	s	als in Ned. som	
z	z	als in Ned. zout	
S	S	als in Ned. kruisje	
Z	Z	als in Ned. gelei	
j	j	als in Ned. jaar	
x	x	als in Ned. lachen	
g	G	als in Ned. geven	
G	x	als in Ned. wasgoed	stemloze fricatief
h	h	als in Ned. hoed	
?	?		glottisslag

## Diacritische tekens

ASCII	X-SAMPA	uitspraak
.	:\	halflang
:	:	lang
`	_h	aspiratie
,	= of _=	vocalisering waardoor l, m, n, N en r vaak syllabische waarde verkrijgen
1	_0	ronding van niet-geronde of overronding van geronde
2	_0	halve ronding
3	_c	ontronding of minder dan normale ronding van geronde
4	_c	halve ontronding
{	_r	meer gesloten
}	_o	meer open
>	_+	meer naar voren
<	_-	meer naar achteren
5	_v	geheel of gedeeltelijk stemhebbend
6	_0	geheel of gedeeltelijk stemloos
~	~ of _~	nasalering
7	~ of _~	halve nasalering
8	_j	palatalisering die tevens een indruk van mouillering geeft
^	_F	sleeptoon of tweetoppig zwak gesneden accent of tweetoppig dalend accent
\	_R	stoottoon of scherp gesneden accent
"		leniskarakter
!		fortiskarakter
'	"	beklemtoneering lettergreep
9	_X	zachtere uitspraak

Voor het gebruik van { en } dient als basis de volgorde i, e, I, E, Q in de niet geronde palatale reeks; y, 0, Y in de geronde; u, o, O, ) in de gutterale; U, @, a, A in de mediale; deze tekens kunnen ook bij consonanten worden gebruikt, b.v. bij de k of t die op weg zouden zijn om door een volgende j te worden geassimileerd.