Frans Hinskens*, Stefan Grondelaers and David van Leeuwen

Sprekend Nederland, a multi-purpose collection of Dutch speech

<https://doi.org/10.1515/lingvan-2019-0024>

Received March 22, 2019; accepted December 7, 2019

Abstract: This contribution sketches the aims, participant recruitment and methodology of the *Sprekend Nederland* project, the amount and types of data, the research possibilities they offer and suggestions for similar future projects.

Keywords: crowd sourced data; language variation and change; perception and attitudes; speech technology; computational linguistics

1 Introduction

A century ago, only a few Dutch people spoke Standard Dutch; most people spoke a dialect of Dutch or one of the regional languages. Standard Dutch was the language of the societal upper-crust and for the rest mainly the language for written communication (cf. Hinskens and Taeldeman 2013; Marynissen and Janssens 2013 for ample references). Today, in most regions the traditional dialects are gradually falling into disuse, but they leave traces behind in regional Standard Dutch. In a large part of the language area it is no longer entirely taboo to diphthongize the tense mid vowels (originally a dialect feature) in spoken Standard Dutch. Future Standard Dutch will doubtlessly differ from today's. Judgements of what is 'good Dutch' will equally change, although maybe with a phase difference. Until the second half of the 19th century, subject *u* ('you', polite) as in *Vindt u dat juist?*, 'Do you think that is right?', was as disputed as subject *hun*, dative 'them' (*Hun vinden dat prima*, 'Them find it great') is nowadays.

The study of processes of language change and the associated judgements and prejudices can profit tremendously from large amounts of comparable, well documented language use by male and female speakers and listeners from different parts of the language area, with a range of different educational and cultural backgrounds. For the Netherlands this has recently been collected in *Sprekend Nederland* ('Speaking Netherlands'), a large-scale crowd-sourcing project for speech and the perception and evaluation of speech, initiated and designed by scientific journalists of the national public broadcast organization, NTR, and a small group of researchers (van Leeuwen et al. 2016; Hinskens et al. 2019). With the aid of a free app for mobile phone and tablet, samples of speech were systematically collected from over 10,000 participants, some 3,000 of whom supplied data regarding their socio-biographical background. The participants supplied 1,552,683 answers to questions concerning other participants and their speech.

Below we will sketch the aims, participant recruitment and methodology of the project (Section 2), the amount and types of data (Section 3), the research possibilities they offer (Section 4) and suggestions for similar future projects (Section 5).

2 Aims, approach and method

NTR planned to devote a series of entertaining TV and radio shows and productions for social media to variation in spoken modern Standard Dutch and its perception and evaluation. To gauge the dimensions

*Corresponding author: Frans Hinskens, Meertens Instituut (KNAW), Amsterdam, Netherlands; and Department of Linguistics/ Centre for Language Studies, Radboud Universiteit, Nijmegen, Netherlands, E-mail: Frans.Hinskens@meertens.knaw.nl

Stefan Grondelaers: Radboud Universiteit Faculteit der Letteren, Centre for Language Studies, Nijmegen, Netherlands

David van Leeuwen: Radboud Universiteit, Nijmegen, Netherlands

of variation in modern spoken Dutch, NTR intended to collect data with a free app, to be developed in consultation with linguists. An underlying objective was to counter misconceptions and prejudices regarding regional, social and ethnic accents with empirically supported reflections.

The linguists aimed at the digital documentation of modern Standard Dutch spoken by as many different speakers of as many (geographically, socially, stylistically, ethnically) different varieties as possible, as well as the collection of data regarding the perception and evaluation of these varieties. The data needed to be suitable for the sociolinguistic investigation of the linguistic and extra-linguistic distribution of the variation, and of perception, attitudes, stereotypes and prejudices. Besides, the data needed to lend themselves for technological and computational research in the domains of acoustic phonetics, speech recognition, automatic transcription and automatic accent recognition.

NTR was the initiator of the project, the sponsor of the app, and the maker of the media productions (for radio, TV, and internet activities) which were intended to highlight the project nation-wide, both to recruit participants and to summarize some first findings for a wide audience.

The academics made an inventory and a selection of research questions, developed the design for the data collection, composed the stimuli,¹ monitored the progress of the data collection and carried out first analyses of aspects of the data.

During the preparations, NTR stressed the importance of an appealing app, as it would further the spread and hence the participation. The main interest of the academics was the collection, via the app, of both scripted and relatively spontaneous speech (unscripted monologues on common themes such as e.g., holidays and music) from the participants, and of data regarding the perception and evaluation of speech of other participants. To that end, the app needed to enable listening to other participants' speech. For the research aims, there was moreover the need to gather specific metadata regarding the participants' socio-biographical backgrounds.

For the participants the app had to be free, entertaining and simple to download and install; furthermore, it needed to guide the participants smoothly through a series of interactions (questions, tasks). The participants needed to be informed about the aim of the project and about privacy protection (participation was anonymous); they needed to be asked permission for the use of their speech and data for scientific research. During the development of the app we consulted the data collection manager of the Meertens Institute in Amsterdam for advice regarding legal aspects of privacy protection and we also remained in continuous discussion with the CLS ethics Institutional Review Board. For more information about the ethical considerations and decisions made see van Leeuwen et al. (2016).

The recruitment of participants was done through traditional media (radio, TV and newspapers), internet, social media, and word of mouth. Via the app the participants were asked to read aloud sentences and brief lists of isolated words, to name objects shown in pictures (equally displayed on the screen), to provide short descriptions of, for example, one's favorite holiday destination and the like. Moreover, the participants were asked to answer questions about themselves and about other participants and their speech. All tasks and questions were balanced around seven themes such as e.g., living, holidays and music.

The personal questions concerned aspects of the participant's socio-biographical background and the answers were stored under the tab 'Mijn profiel' ('My profile', bottom part of the screen; Figure 1). The questions concern the participant's origin, residence, education, occupation, commuting behavior, gender, age, whether or not the participant has children, as well as his/her linguistic repertoire. There were also questions regarding political and religious convictions, attitudes towards specific regional and ethnic groups and the like.

The participants were asked to rate other speakers on semantic differentials for qualities such as intelligence and reliability, as well as for their suitability for certain roles (as e.g. neighbor) and functions (such as newscaster).

'Rewards' for the participants were interactive language games (tongue-twisters, jokes and riddles, under the tab 'Bonusmateriaal'; 'Bonus material' Figure 1) and feedback on their own speech from other participants. In order not to put off potential participants, negative judgements were not mentioned.

¹ All this after consultation of other researchers (of linguistics, social geography, social psychology, sociology and communication studies).

Figure 1: The screen for the start of the theme *Wonen*, ‘Living’.

3 Data

Using the app, data were collected in the period 1 December 2015 to 31 December 2016. The majority of the data was collected in the first half of this period, in which there were several nation-wide media events about the app and the project in general. The effect of the media events could clearly be seen in the average daily rate at which recordings were made. This rate would shoot up to levels as high as 10 recordings per minute, and in the following days this would then exponentially decrease at a rate of approximately 5% per day (van Leeuwen et al. 2016).

Despite the attempt to make the app as attractive as possible, the data collection suffered high dropout rates of participants. Even though almost 18 thousand people registered (approximately one promille of the population), i.e., filled in an email-address for the purpose of keeping participants apart, only 56% of these recorded at least a single utterance. A higher percentage (73%) gave at least one answer to a question, which is a better yield but still shows that there is a great reluctance amongst the general audience to give personal information. In the end we recorded over 290 thousand utterances from participants, totaling 528 hours of speech data. The average number of recordings made per participant is 29.2, and the average duration of a recording is 6.5 seconds. Participants gave over 17 million answers to questions, an average of 12.3 per answering participant, of which 9% were metadata questions (age, sex, origin, sociological attitudes) and 91% were judgement questions about other participants based on their speech.

The distribution of participants was quite representative in terms of the province of origin (Lohfink 2017), but much less so in terms of age, over-representing the ages under 30 and under-representing ages over 50. The participant's gender was skewed towards female (56%), and the level of education was skewed toward higher education, with about $\frac{2}{3}$ reporting university or higher vocational education where about $\frac{1}{3}$ would be more representative of the Dutch population. The English Dialects App showed very similar biases; cf. Leemann et al. (2018).

The app was structured in terms of themes (e.g., living, or birthday), a coherent section in the app with an introduction movie featuring a celebrity, and recording prompts and questions relating to the theme. This theming had an effect on the moment the participants stopped using the app; there was a tendency to lay the app aside after a theme had been completed, never to return. Because the metadata questions were posed at fixed moments within a theme, the distribution of the number of answers per question is very uneven, with much lower numbers at each new theme (van Leeuwen et al. 2016).

One of the major experimental design challenges was how to couple speakers to listeners, i.e., how to decide which recordings of which participants to present to the current participant for judgements. The original idea was to sample speakers based on their geographical distance to the listener, filling logarithmically spaced distance bins equally. This sampling assumes that accent perception ability wears off with distance, and ignores variation in population density. However, this gave rise to computational difficulties in the first days of the deployment of the app, and this policy had to be changed to a random speaker-listener coupling. Still, this policy caused early participants to be selected as speaker more often than later participants. Because one of the incentives of participating was to receive attitude judgements from others as feedback, the policy was changed later to select speakers with fewer attitude judgements.

Most of the recordings involved read speech, with prompts from one of six categories, where each of the categories was designed with different research questions in mind. These were augmented with two categories aimed at more spontaneous speech: one was a picture-naming task and one a description task, the latter related to the theme. In choosing speech material to present to a listening participant for judgement, all categories were sampled randomly.

The data are presently hosted by the Nederlands Instituut voor Beeld en Geluid ('Dutch Institute for Vision and Sound'), the national archive of the public radio and TV, in Hilversum.

4 Some of the first studies

Preliminary findings have been summarized by the researchers and journalists involved in non-technical language and disseminated through NTR's *Kennis van Nu* TV show, Facebook and related social media. Recently, exploratory studies of several aspects have been carried out with parts of the collected data. In the following subsections, three of these studies will be succinctly presented.

4.1 Accent recognition

A first step towards automatic accent recognition on *Sprekend Nederland* (henceforth SN) data was a study carried out in a Master's thesis (Lohfink 2017), using an early release of SN (March 2016). A selection of the

then-available recorded data was made by the requirement for speakers to have made a recording of at least 10 seconds, and having supplied location information in the metadata questions. There were several questions related to location and accent, but due to the skewedness in the distribution of answers per question, only the answers to the question “where have you lived the largest part of your life” were used, as the answer to this question had the highest response rate.

Since the location answers were given as a point location on a pannable/zoomable map, these needed to be converted to an assumed accent in order to be able to carry out accent recognition research. In a first attempt the locations were binned in one of the 12 provinces of the Netherlands, but automatic accent recognition results were very poor for these classes. Since provinces are mostly political and governmental constructs, we cannot really assume that accents form according to these boundaries. Therefore, a k-means clustering approach was used to cluster all of the participant’s locations into 12 clusters.² These clusters could typically be associated with a larger city that was located close to the center of the cluster. A manual correction was made by adding a cluster in Zeeland, which did not have a cluster center, and removing one in Flevoland, which was also motivated by the fact that this province did not really exist yet when a dialect map of the Netherlands was published by Daan and Blok (1969).

With these clusters as categories, it was possible to train a system for automatically recognizing accents. The system used i-vectors (Dehak et al. 2009) as a 400-dimensional dense vector representation of a speech utterance and a 5 layer Deep Neural Network as classifier. After optimizing the performance on the full training set in a cross-validation setup, the system’s performance could be compared to the average performance of other SN participants. One of the questions about a spoken utterance was “Where do you think the speaker of this utterance comes from?”, where the answer was given using the same pannable/zoomable map interface. For 897 speakers their location was thus guessed by other participants, and these locations were then discretized as one of the classes defined in the clustering method described above. The overall performance for the participants was a mean recall of 18.1% and a mean precision³ of 21.5%. The best automatic system, when tested on the same 897 speakers, achieved a comparable recall of 16.9%, but a somewhat lower precision of 15.6%. All of these figures are not very high (a perfect system would score 100% on both recall and precision), indicating that this accent recognition task is very hard, or that the accents exposed by the speakers of this subset of SN are not very different from Standard Dutch. The latter might be an effect of the fact that participants of SN were aware that their speech was being judged, and in fear of receiving stereotypical “bad judgements” participants might have tended to hide their everyday accent and expose a more Standard Dutch accent.

4.2 Four wide-spread phonetic variables in four regions. Production, perception and evaluation

Among the scripted speech elicited with the app, there were ten sentences (containing 43 relevant words) and 44 isolated words which served to elicit realizations of five linguistic variables in controlled linguistic contexts. These variables crucially determine the linguistic make-up of regional and ethnic accents across the Netherlands. Three of these variables concern the phonetic realization of specific segments, one concerns the ending /ən/, which is usually pronounced n-lessly.

Smidt (2017) carried out a small-scale pilot study of the realization of these four variable phenomena in a sample of 31 speakers, and of the perception and evaluation of the speakers and their speech by other participants. All speakers (15 men, 16 women) were between 20 and 40 years of age and all had had higher vocational or academic training. They were evenly distributed over the four regions⁴ Noord-Holland, Groningen, Twente and Limburg. Cf. Map 1.

² Here, K-means associates the participant’s locations according to their shortest distance to any of $k = 12$ centers, which are the “center of gravity” of the locations associated to this center.

³ The recall is the fraction of a class in the database that was recognized as such, the precision is the fraction of a recognized class that has the same label in the database.

⁴ At the time, the sample contained only 3 male participants from Groningen who met all criteria.

Map 1: The four regions represented by the 31 speakers in the pilot study.

The data for two of the phenomena display interesting patterns. One concerns the realization of lexical word-final /ən/ (in e.g. *regen*, ‘rain’, *laken*, ‘sheet’, *varken*, ‘pig’). There is a statistical interaction between the region of origin and the sex of the speakers ($F(3) = 4,755$, $p = 0.010$). See Figure 2.

Among the Groningen speakers, the males appear to realize final /n/ more often than the females; the Twente pattern is the exact reverse. Whatever the explanation of this patterning (say, e.g., in terms of indexicality), insofar as the finding can be generalized, it will be clear that among speakers with a high educational level the phenomenon will probably dwindle away in Groningen and survive in Twente. The variation in the realization of morphemic -ən (e.g. in plurals) shows no clear pattern whatsoever.

For 26 of the speakers mentioned above, selected read sentences were listened to by 145 other participants; on average there were 12.4 listeners per speaker. From the listeners who had themselves answered the question concerning their region of origin, 33 came from the northeast, 51 from the northwest, 28 from the southeast and 13 from the southwest. After listening to the speech fragments, the listeners answered the

Figure 2: The phonetic realization of lexical /n##/ per region and gender.

Table 1: Outcomes of the factor analysis (Varimax rotation).

	Personal engagement	Personal and societal engagement	Societal engagement
Intelligent	0.152	0.759	0.539
Highly educated	0.058	0.933	−0.041
Honest	−0.158	0.880	−0.096
Reliable	0.171	0.872	−0.237
Fun	0.712	0.548	0.075
Warm	0.079	0.722	−0.452
Trendy	0.490	0.612	−0.102
Modern	−0.053	0.813	−0.051
Neighbor	0.961	0.005	0.179
Colleague/classmate	0.880	−0.096	0.355
Friend	0.940	0.046	−0.200
Boss	0.896	0.036	0.021
Family-in-law	0.921	−0.059	−0.245
Love partner	0.758	−0.161	−0.569
Newscaster	0.486	0.230	−0.245
Teacher	−0.240	−0.645	0.326
Mayor	0.024	−0.347	0.845
Advertiser	0.057	−0.052	0.925
Quizmaster	0.386	0.088	−0.491
Money maker	−0.441	0.710	0.136

Per factor the speaker qualities with the highest loading are presented in bold italics.

question where they thought the speaker comes from. Of these, 55.4% of the localizations appeared to be wrong. The speakers from Noord-Holland were best recognized as such (58.8%), but the listeners from this region (northwest) were themselves the worst recognizers; the listeners from the northeast turned out to be the best recognizers (40.6% correct).

Thirty of the speakers in the sample were evaluated by 132 other participants on a number of qualities and for their suitability for certain roles and functions. Factor analysis shows that more than three quarters ($R^2 = 77.16\%$) of the variance in the scores on 20 scales can be explained by three factors. See Table 1.

The first factor mainly concerns personal engagement (hence the low loadings for ‘newscaster’ and ‘quizmaster’), and the third factor societal engagement; there is a second factor, concerning personal aspects with an immediate societal relevance.

Remarkably, for the second factor, i.e. on the scales ‘intelligent’, ‘highly educated’, ‘honest’, ‘reliable’, ‘warm’, ‘trendy’, ‘modern’ and ‘money maker’, the female speakers were rated significantly lower than the males ($F(1) = 11.166$, $p = 0.025$) and this pattern is independent of the region, in other words regional coloring, however strongly, of Standard Dutch speech does not play a role. The first and third factor did not show any relationship with the speakers’ sex or region of origin.

4.3 Sprekend Nederland data and speaker evaluation studies

The difference between the studies outlined in Sections 4.1 and 4.2. and the ones in the present section is that the latter do not investigate SN-materials as such, but exploit them as a privileged source of high quality stimulus speech in speaker evaluation experiments, such as the matched- or verbal-guise paradigm, which are used to investigate the prestige of language or accent varieties. Compared to earlier corpora like the Teacher Corpus included in the Spoken Dutch Corpus (on which Grondelaers et al. 2010, Grondelaers et al. 2015; Grondelaers and van Hout 2010; and are based), the SN-corpus has a much richer gamut of stimuli on offer, containing different types of read aloud and spontaneous speech with various degrees of scripting. Most beneficial for Grondelaers’ research into the perceptual correlates of standard language dynamics in Dutch, is a set of ten sentences constructed to contain the five phonetic variables which are crucial to the quality of regional and ethnic accents across the Netherlands (see Section 4.2 above). Another pivotal advantage of the publicly sourced SN materials is their accent strength bandwidth, which is much larger than in the Teacher

corpus (which contains the speech of teachers, who typically have only mild regional or social accents and try to reduce it as much as possible – Grondelaers et al. 2015). Building on SN-speech, we were able to carry out two new studies which nuance the dominant view on standard language dynamics in Dutch. Whereas previous studies invariably confirmed the uncontested dominance of the Randstad accent, the new study reported in Grondelaers et al. (2019) suggests that the regional or ethnic origin of an accent is less important nowadays than its strength: allegedly low prestige accents such as the southern Limburg accent lose some of their stigma when they are comparatively milder. And the Moroccan accent of Dutch may not be prestigious, but it is deemed dynamic and cool, which may explain why endogenous youths and Surinamese rappers are increasingly adopting it (Grondelaers and van Gent 2019).

5 Strengths and weaknesses of the database and suggestions for future work

Thanks to a fruitful cooperation with the national broadcaster NTR, we have obtained a large research corpus of unprecedented richness. NTR allowed us (the researchers) substantial freedom in crucial decisions on the experimental design of the app, viz. the stimulus materials and the survey questions. In view of the impressive media attention NTR was able to generate for the project, the app was distributed on a much larger scale than would have been feasible for any scientific consortium, and participation, accordingly, was overwhelming. The eventual participant sample we were able to recruit represents most of the demographic groups in Dutch society to a sufficiently large extent to allow for balanced sampling.

A downside of collaborating with media partners is their strict production goals (a mixture of science and entertainment) and even stricter deadlines. As a consequence, the researchers did not have full control over the design of the app, and no control whatsoever over participant recruitment, which resulted in underrepresentation in some cells: few participants have provided all the required meta-data, and there are proportionally fewer female participants, and participants with a migrant background or lower education in the corpus. Somewhat more problematic than such imbalances is the fact that language documentation apps like SN attract speakers with some confidence in the prestige and/or “quality” of their Dutch; in addition, contributors will typically make an effort to sound as standard as possible (see Grondelaers et al. 2019: 222), as a result of which their speech is less identifiable in geographical terms.

In spite of these drawbacks, the SN-corpus is arguably the richest speech corpus available in Dutch, if not in terms of size, then certainly in terms of the variety of its speech materials, its metadata, and the unique availability of diverse types of evaluations. Taken together, these advantages enable various brands of language researchers to significantly extend their knowledge of (standard) language dynamics in present-day Netherlandic Dutch.

To begin with, the unexpectedly high recording quality of the majority of its speech samples makes the SN-corpus suited for all types of acoustic analysis – including forensic applications – which require clean speech. For the same reason, the corpus can assist speech technologists in improving their recognition software: the scripted speech in the corpus represents an evident training set for the classification of the topic-controlled spontaneous materials.

In addition, the SN-corpus can assist in the development of automatic tools for the identification of regional and ethnic accents (Bahari et al. 2013; Hanani et al. 2013; van Leeuwen and Orr 2016). Again, the availability of meta-data pertaining to the geographic origin and mother tongue of the speakers can help correct and finetune the algorithms.

On a more theoretical level, the micro-stratification and the availability in the corpus of different types of speech per speaker (from read-aloud to completely unscripted) enables sociolinguists to investigate on a hitherto unprecedented scale the persistence of dialect and other non-standard features in Netherlandic spoken Standard Dutch (Hinskens and Taeldeman 2013); the wide age range in our participant sample (from 18 to 80 years of age) allows for a crucial apparent time dimension in the research. The presence of prestige and attractiveness evaluations in the database enables us to investigate the indexicality (social meaning) of emergent speech variants, which has been found to co-determine their production success. On the basis of two types

of prestige evaluations in the SN-corpus, it is possible to investigate whether speakers who are deemed more traditionally superior stick to standard forms, whereas speakers who are deemed more dynamically prestigious produce proportionally more “cool” non-standard forms (as maintained in, for instance, Grondelaers and van Hout 2016).

The unique availability in the corpus of perception and production data pertaining to the same speakers, however, also allows for a more ambitious way to study the impact of prestige on speech production: a learning algorithm can be trained to automatically classify transcribed speech in terms of its degree of traditional and/or dynamic prestige, as measured on linguistic (lexical, morphosyntactic) and phonetic (segmental and suprasegmental) features.

Apart from linguistic research, the SN-corpus is eminently suited for language education purposes, especially acquisition programs “extra muros”, which rarely confront aspiring students with the manifold surface manifestations of modern spoken Standard Dutch. The SN-corpus can bridge the wide gap between the homogeneity illusion incorporated in many teaching methods and the linguistic reality which, even in a country like the Netherlands, with its largely successful standardization, is becoming increasingly variable.

References

- Bahari, Mohamad, Rahim Saeidi, Hugo van Hamme & David van Leeuwen. 2013. Accent recognition using i-vector, Gaussian mean supervector and Gaussian posterior probability supervector for spontaneous telephone speech. *Proc. ICASSP*, 7344–7348.
- Daan, Jo & Dirk Blok. 1969. *Van randstad tot landrand: Toelichting bij de kaart Dialecten en naamkunde*. Amsterdam: Noord-Hollandsche uitgevers maatschappij.
- Dehak, Najim, Patrick Kenny, Réda Dehak, Pierre Dumouchel & Pierre Ouellet. 2009. Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing* 19(4). 788–798.
- Grondelaers, Stefan & Paul van Gent. 2019. How deep is dynamism? Revisiting the evaluation of Moroccan-flavoured Netherlandic Dutch. In the theme issue *Implicitness and experimental methods in language variation research of Linguistics Vanguard* 5(1). 20180011, ISSN (Online) 2199-174X, DOI: <https://doi.org/10.1515/lingvan-2018-0011>.
- Grondelaers, Stefan & Roeland van Hout. 2010. Is Standard Dutch with a regional accent standard or not? Evidence from native speakers' attitudes. *Language Variation and Change* 22. 221–239.
- Grondelaers, Stefan & Roeland van Hout. 2016. How (in)coherent can standard languages be? A perceptual perspective on co-variation. (Theme issue *Coherence, covariation and bricolage. Various approaches to the systematicity of language variation*. In F. Hinskens & G. Guy (ed.), *Lingua*, 172–173, 62–71.
- Grondelaers, Stefan, Roeland van Hout & Mieke Steegs. 2010. Evaluating regional accent variation in Standard Dutch. *Journal of Language and Social Psychology* 29. 101–116.
- Grondelaers, Stefan, Roeland van Hout & Sander van der Harst. 2015. Subjective accent strength perceptions are not only a function of objective accent strength. Evidence from Netherlandic Standard Dutch. *Speech Communication* 74. 1–11.
- Grondelaers, Stefan, Roeland van Hout & Paul van Gent. 2019. Re-evaluating the prestige of regional accents of Netherlandic Standard Dutch. The role of accent strength and speaker gender. *Journal of Language and Social Psychology* 38. 215–236.
- Hanani, Abualsoud, Martin Russell & Michael Carey. 2013. Human and computer recognition of regional accents and ethnic groups from British English speech. *Computer Speech & Language* 27(1). 59–74.
- Hinskens, Frans & Johan Taeldeman. 2013. Introduction to the volume. In Frans Hinskens & Johan Taeldeman (eds.), *Language and Space: Dutch*, 1–12. Berlin: De Gruyter.
- Hinskens, Frans, Stefan Grondelaers, David van Leeuwen & Martijn Wieling. 2019. *Sprekend Nederland*. Een nieuwe, voor meerdere doeleinden te gebruiken dataverzameling van en over gesproken Standaardnederlands. Submitted for publication in *Nederlandse Taalkunde*.
- Leemann, Adrian, Marie-José Kolly & David Britain. 2018. The English Dialects App: The creation of a crowdsourced dialect corpus. *Ampersand* 5. 1–17.
- Lohfink, Georg. 2017. *The ‘Sprekend Nederland’ project applied to accent location*. MA-thesis, University of Utrecht.
- Marynissen, Ann & Guy Janssens. 2013. A regional history of Dutch. In Frans Hinskens & Johan Taeldeman (eds.), *Language and Space: Dutch*, 81–100. Berlin: De Gruyter.
- Smidt, Eva. 2017. *The production, perception and evaluation of regional accents in present-day standard Dutch in the Netherlands. A study based on crowdsourced data through ‘Sprekend Nederland’*. MA-thesis, University of Groningen.
- van Leeuwen, David & Rosemary Orr. 2016. The ‘Sprekend Nederland’ project and its application to accent location, *Proc. Odyssey 2016: The Speaker and Language Recognition Workshop, ISCA, Bilbao*, 101–108.
- van Leeuwen, David, Frans Hinskens, Borja Martinovic, Arjan van Hessen, Stefan Grondelaers & Rosemary Orr. 2016. *Sprekend Nederland: A heterogeneous speech data collection*. *Computational Linguistics in the Netherlands Journal* 6. 21–38.