Analysis of spatial variation with app-based crowdsourced audio data

Marie-José Kolly[1], *Adrian Leemann*[2], *Florian Matter*[3]

[1] LIMSI-CNRS Orsay / Department of Comparative Linguistics, University of Zurich
[2] Department of Theoretical and Applied Linguistics, University of Cambridge
[3] Department of General Linguistics, University of Bern

marie-jose.kolly@uzh.ch, al764@cam.ac.uk, florianmatter@gmail.com

10.21437/Interspeech.2015-393

Abstract

Crowdsourcing technologies such as the use of web or smartphone applications enable the collection of large linguistic corpora. Such corpora allow for an analysis of various linguistic features. In the present contribution we analyzed data of >1500 speakers that were crowdsourced with the smartphone application 'Dialäkt Äpp' to examine one such feature: the articulation of /kh/ as [k͡x] in Swiss German. We compared our results with findings from historical dialectological surveys. Results showed evidence for sound change in progress in Northern Switzerland with [k͡x] diffusing towards Zurich but retreating from the Northeast. We discuss the findings against the backdrop of methodological caveats of crowdsourcing audio data through smartphone applications.

Index Terms: crowdsourcing, iOS application, sound change, dialectology, Swiss German

1. Introduction

Swiss German (hereafter SwG) dialects are the prestige varieties in German-speaking Switzerland [1, 2, 3]; they are spoken by 66% of the Swiss population [4] and belong to the High Alemannic varieties [5]. There is considerable linguistic variation both within and between the various dialects. Extensive documentation of many aspects of SwG exists, which is, however, mainly based on traditional dialectological approaches [6, 7, 8]. One prominent representative of such traditional approaches is the *Atlas of German-speaking Switzerland* [6], which documented dialectal variation in 566 SwG localities in over 2000 maps. Data for the *Atlas* were collected in the 1940s in the field, where dialectologists typically surveyed one or two elderly informants per locality. Hence this historical data reflects dialect use from about 100 years ago [9]. More recent, large-scale documentation exists in the form of dialect data that has been crowdsourced with smartphone applications. In Switzerland, this has been done with 'Dialäkt Äpp' (hereafter DÄ) [10].

DÄ was launched in March 2013 [10]. The app has two main functionalities: (a) it predicts the users' SwG dialects by letting them indicate their dialectal pronunciation of 16 words; (b) it allows users to record these words and a short dialogue in their own dialect. These recordings become available on an interactive map for users to listen to (cf. [11] for a detailed description of the methods used to create DÄ, the functionalities it provides, and the data it crowdsources). DÄ was a success, being the number one downloaded free iOS app in Switzerland after its release [12]. It received major media attention [13] and, so far, has >85000 downloads. The crowdsourced corpus contains (a) (written) choices of pronunciations for 16 words for each user who participated in the dialect prediction functionality and (b) audio data for the same 16 words and, in addition, recordings of a short dialogue for users who did the recording. The corpus contains data of over 58000 subjects for (a) and over 3000 subjects for (b) [11].

Data crowdsourced through DÄ has been the basis for a number of small-scale research projects. Data from part (a) of the corpus have been used (i) to create contemporary dialect maps and compare them to the historical maps published in the *Atlas* [14], and (ii) to document phenomena of linguistic change in the canton (i.e. administrative subdivision) of Bern [15]. Data from part (b) of the corpus have been used (i) to compare dialects at the acoustic phonetic level, where preliminary results show differences in speaking rate between the dialects of Bern and Zurich [16]; (ii) to train an automatic speech recognition algorithm for DÄ's follow-up app, 'Voice Äpp' [17, 18, 19], which predicts users' dialect based on automatic speech recognition; (iii) as stimuli for experiments on the perception of speakers' age [20], of young speakers' gender [21] and on the pleasantness of vocalized and lateral /l/ in the SwG word *Kelle* 'ladle' [22]. Currently, a handful of smartphone applications crowdsource linguistic data. [23, 24] present applications for documenting endangered languages; [25, 26] developed applications to collect speech for the training of acoustic models.

In the present contribution, crowdsourced audio data from DÄ's recording function were used for a spatial analysis of the pronunciation of /kh/ as [k͡x] – a relatively recent feature of SwG [7]: going on a brief historical tangent, Germanic voiceless plosives changed to affricates and fricatives in the 8th century [27]. In SwG, word-initial [k͡x] later shifted to /x̯/ [28]. Word-initial [kh] and [k͡x] are now mainly found in loanwords and morphologically complex forms as, for example, the past participle of *ha* 'to have': *gha*. Like other past participles it is formed with the prefix /k-/ – and /kh/ is pronounced as [k͡x] in some SwG areas. This linguistic innovation was mapped by [29], using historical data collected for the *Atlas*.

As Germanic plosives and affricates have undergone sound change quite extensively in the past [27, 28] we expect the complementary distribution of [kh] and [k͡x] in *gha* to have evolved over the past 100 years. We therefore conducted a diachronic analysis by comparing contemporary DÄ data with the historical *Atlas* data mapped in [29]. For regions where the DÄ data differs from the historical data, we expect younger speakers to differ from older speakers, which would be evidence for sound change in apparent time [30].

Copyright © 2015 ISCA

2. Methods

2.1. Speakers

We analyzed a total of 1531 speakers who participated in the dialogue recordings. The speakers self-declared their dialect by selecting the *Atlas* localities that best correspond to their dialects [11]. Speakers stem from 357 different localities (retrieved as of 29.01.2014). We further elicited gender (49% m, 51% f) and age (<20: n=499; 20–30: n=366; 30–40: n=225; 40–50: n=234; >50: n=207).

2.2. Material

Data were provided by DÄ users who recorded tokens of *gha* by adapting the Standard German dialogue part shown in Figure 1 (bottom right) to their dialect. The text includes the past participle in "[...] hattest du aber schwer zu tragen *gehabt*" '[...] you have had a heavy [load] to carry'. Two realizations of the syllable onset are possible: [kh] or [k͡x].

2.3. Procedure

The interface of the recording functionality prompts users to choose their dialect from a list of localities used in the *Atlas*, and to indicate age and gender (Figure 1, top left) before they proceed to a screen with recording instructions (Figure 1, top right) that read: "Please record your voice in as quiet an environment as possible. Keep an approximate distance of about 15 cm between your device and your lips. Please articulate the text loudly and clearly in your own dialectal pronunciation". For the recording of the dialogue, additional information is provided (Figure 1, bottom left). This screen reads: "In the following, you can record the dialogue 'Gespräch am Neujahrstag' in your dialect. Please adapt words and pronunciation to your own dialect where necessary […] The text is presented in short paragraphs. After each reading, you will be able to listen to, and, where necessary, re-record the passage." The user then proceeds to do the recording (Figure 1, bottom right). The text is an excerpt of [31], which was originally designed to feature a high number of variables that potentially reveal between-dialect differences.

Recordings were analyzed auditorily by a phonetically trained researcher (third author). An acoustic analysis of the material was deemed unnecessary, since the focus was on the variation between [kh] and [k͡x], and not on acoustic features.

To compare DÄ data with historical *Atlas* data, we used data present in [29]. This comparison data is structured somewhat differently, given that [29] does not display raw data but used four categories of subset data instead: in the first category were localities in which all tokens of *gha* contained [k͡x], in the second category >33.3% of tokens contained [k͡x], in the third category <33.3% of tokens contained [k͡x] and in the final fourth category no tokens contained [k͡x] (or no data was available). To conduct descriptive statistics and, more importantly, to create maps of spatial variation, we applied the following conversion: (1) 100%, (2) 66.6% ((100+33.3)/2), (3) 16.6% (=(33.3+0)/2), (4) 0%. The maps created based on DÄ data show the proportion of speakers per locality that used [k͡x] whereas the *Atlas*-based maps – based on data published in [29] – show the proportion of [k͡x] in all of the *gha* tokens collected from one or more informants in a locality.

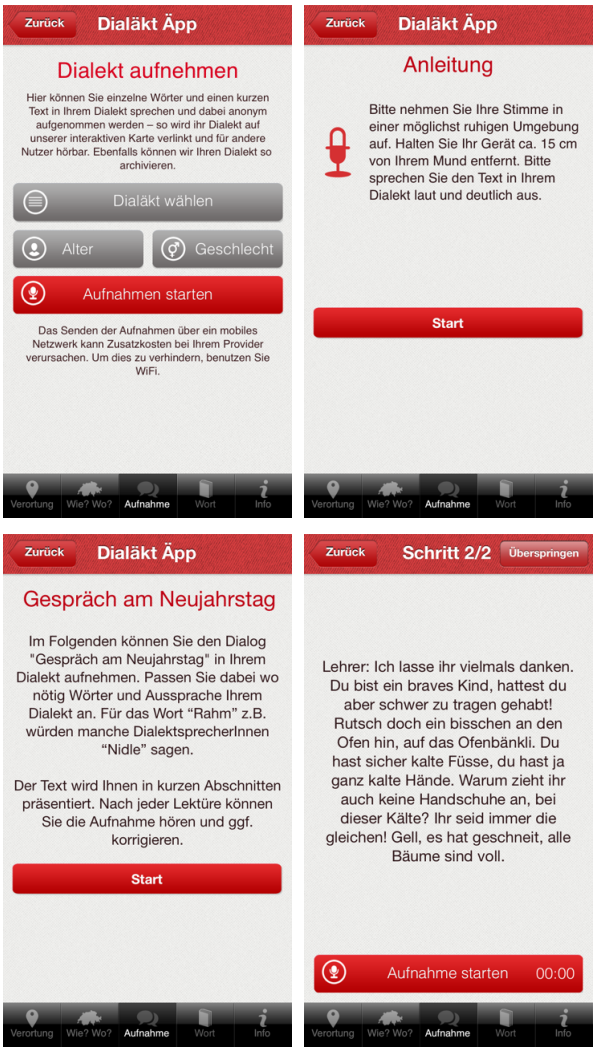


Figure 1: *Screens that prompt users to indicate metadata (top left), instruct them on how to perform the recordings (top right), show the reading instructions (bottom left) and illustrate the dialogue (bottom right).*

2.4. Visualization

Using [32], the coordinates for every locality were retrieved and combined with the information present in the metadata. We used a Google Javascript API [33], which enabled a customization of Google Maps elements and allowed for the creation of heatmaps based on custom data. As the variants of this binary variable are complementary, we graphically illustrate the spatial distribution of proportions of [k͡x] only. Red areas in the maps indicate high, yellow areas moderate, and green areas low proportions of [k͡x]. Areas without [k͡x] remain uncolored.

3. Results

3.1. Descriptive statistics

Out of 1531 tokens of *gha* in the DÄ corpus, 202 (13.2%) contained [k͡x]. In the *Atlas* data, 22.2% of the *gha*-tokens contained [k͡x] (cf. 2.3). As for the cantonal distributions, high

proportions of [k͡x] were found in the districts of Wallis and Glarus, while the *Atlas* additionally showed high proportions for St. Gallen, Appenzell and Graubünden. Relative proportions of [k͡x] per canton are presented in Table 1.

Table 1. *Percentage of [k͡x] per canton.*

Canton	DÄ	*Atlas*	Canton	DÄ	*Atlas*
Wallis	87.1% (n=31)	47.8% (n=30)	Zürich	6.6% (n=378)	0.3% (n=55)
Glarus	84.2% (n=19)	75.8% (n=11)	Luzern	4.5% (n=111)	25.3% (n=31)
St. Gallen	33.9% (n=124)	47.4% (n=44)	Zug	3.1% (n=32)	0% (n=7)
Appenzell	33.3% (n=6)	68.1% (n=12)	Aargau	2.6% (n=152)	0% (n=40)
Graubünden	26.9% (n=52)	47.7% (n=36)	Basel	2.0% (n=99)	0.1% (n=18)
Bern	17.2% (n=267)	20.7% (n=78)	Unterwalden	0.0% (n=23)	0% (n=12)
Fribourg	15.4% (n=13)	14.6% (n=8)	Schaffhausen	0.0% (n=18)	8.3% (n=8)
Solothurn	10.5% (n=76)	3.0% (n=22)	Schwyz	0.0% (n=44)	10.3% (n=13)
Thurgau	10.1% (n=79)	10.9% (n=22)			

Table 1 gives the general picture for the spatial distribution of [k͡x]. While it is true as a rule of thumb that each canton has its own SwG dialect, linguistic differences do not strictly correspond to these administrative borders [1, 34, 35]. We therefore created feature maps for the subsequent analyses.

3.2. Spatial distribution of [k͡x]

The maps in Figure 2 show areas where [k͡x] was (top) and still is (bottom) more dominant than in others: Wallis (as indicated with the purple dot (1)), the Bernese Oberland (2), Glarus (3), Graubünden (4), Appenzell and St. Gallen (5).

In the DÄ data, tokens of [k͡x] are found in a large green area around the Northern lake of Zurich and in other localities throughout the Central Plateau (3). The atlas does not show this trend. However, we find a larger area with [k͡x] tokens in St. Gallen and Appenzell (5) in the *Atlas* data than in DÄ data.

3.3. Areas of sound change

The *Atlas*-based map in Figure 2 (top) shows a similar pattern to the DÄ-based map (bottom). Two important differences between the maps stand out: in the Northeastern cantons of St. Gallen and Appenzell, there seems to be a higher proportion of [k͡x] in the historical data whereas around the Northern lake of Zurich we find higher proportions in contemporary data.

As to Northeastern Switzerland, close-up visualizations of the *Atlas*-based (Figure 3, left) and the DÄ-based (Figure 3, right) maps suggest a retreat of [k͡x] over the past decades. In the *Atlas* data, [k͡x] is predominant in the localities shown in Figure 3 (81.5%). In the same area, we found much smaller proportions in the DÄ data, with 28 [k͡x] of 123 *gha* tokens (22.8%). As for Zurich, Figure 4 suggests an innovation of [k͡x]: in the *Atlas* data (left), [k͡x] is not attested at all in the localities shown on the map (0%) whereas DÄ data (right) showed 23 [k͡x] of 306 *gha* tokens (7.5%).

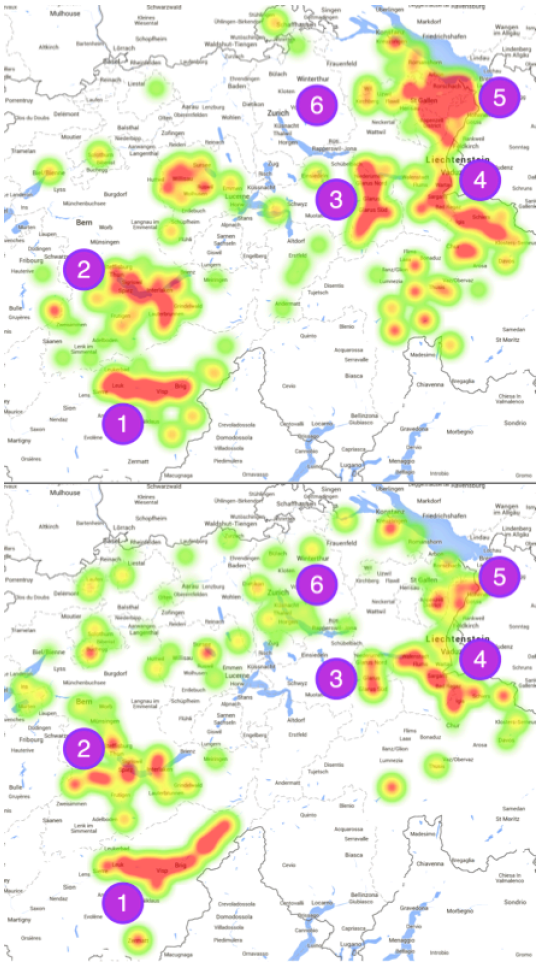


Figure 2: *Distribution of [k͡x] based on* Atlas *(top) and DÄ (bottom) data.*

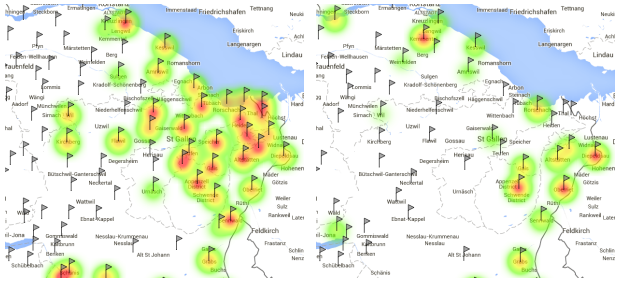
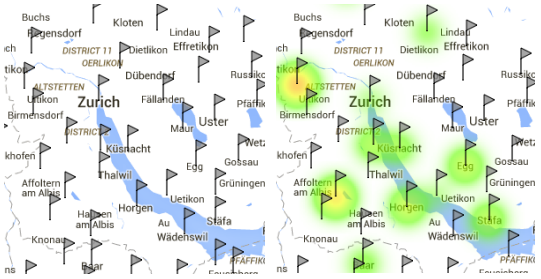


Figure 3: *[k͡x] in Northeastern Switzerland according to* Atlas *(left) and DÄ data (right). Flags stand for sampled localities.*


Zurich
Kloten
Dietlikon
Effretikon
Dübendorf
Fällanden
Uster
Küsnacht
Egg
Gossau
Thalwil
Horgen
Uetikon
Au
Wädenswil
Stäfa
Meilen
Rüschlikon
Gattikon
Adliswil
Langnau am Albis
Hausen am Albis
Affoltern am Albis
Knonau
Buchs
Regensdorf
Lindau
Russikon
Pfäffikon
Wetzikon
Grüningen
Birmensdorf
Urdorf
Oberrieden
Kilchberg
Rüschlikon
Thalwil
Zollikon
Erlenbach
Herrliberg
Feldmeilen
Uetikon am See
Männedorf
Stäfa
Hombrechtikon
Meilen
Küsnacht
Zollikerberg
Forch
Pfaffikon
Wetzikon
Hinwil
Bubikon
Rüti
Dürnten
Bäretswil
Seegräben
Wetzikon
Fehraltorf
Illnau-Effretikon
Lindau
Hittnau
Pfäffikon
Russikon
Kyburg
Weisslingen
Turbenthal
Wila
Sternenberg
Bauma
Fischenthal
Wald
Rüti
Dürnten
Bubikon
Hinwil
Gossau
Grüningen
Egg
Mönchaltorf
Volketswil
Schwerzenbach
Fällanden
Dübendorf
Wangen-Brüttisellen
Opfikon
Glattbrugg
Kloten
Bassersdorf
Nürensdorf
Embrach
Lufingen
Rorbas
Freienstein-Teufen
Bülach
Eglisau
Hochfelden
Höri
Bachenbülach
Winkel
Oberglatt
Niederglatt
Rümlang
Regensdorf
Buchs
Dielsdorf
Niederhasli
Niederweningen
Schöfflisdorf
Stadel
Weiach
Glattfelden
Zweidlen
Wasterkingen
Rafz
Wil
Hüntwangen
Wil
Stadel
Niederweningen
Schöfflisdorf
Dielsdorf
Niederhasli

3.4. Effect of age

Figure 5 shows the proportions of [k͡x] per age group in Northeastern Switzerland (left) and the area around the Northern lake of Zurich (right) for DÄ data. The bar charts show a tendency for the use of [k͡x] to increase with age in Northeastern Switzerland and to decrease with age around Zurich. To test for the effect of *age*, we constructed, for each area, a generalized linear model (response variable: [k͡x] counts, factors: *age group, locality*). We performed a standard likelihood ratio test comparing this model with a reduced model without the factor *age*. The effect of *age* was not significant at an α-level of .05 in both instances.

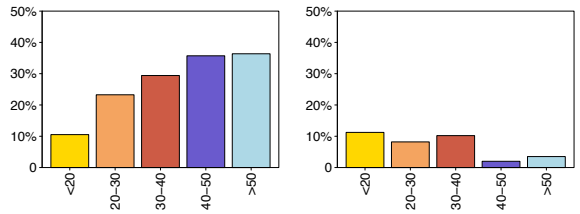


Figure 5: *Percentage of [k͡x] per age group in Northeastern Switzerland (left; n=123), and around Zurich (right; n=306).*

4. Discussion and conclusion

Our results showed that the spatial distribution of [k͡x] and [kh] in *gha* was relatively similar in the historical and contemporary data. This confirms findings by [14, 36, 37] who reported that the spatial distribution of phonological features is relatively stable, whereas lexical features, for example, are more likely to undergo linguistic change.

Two areas in Northern Switzerland, however, showed tendencies for sound change in apparent time [30]: (1) In Northeastern St. Gallen and Appenzell we observed lower proportions of [k͡x] in contemporary DÄ data than in historical *Atlas* data. The contemporary data showed a tendency for the proportion of [k͡x] to increase with speakers' age, which leads to the hypothesis that [k͡x] may be on retreat in this area. (2) In the Zurich area we observed the opposite pattern: higher proportions of [k͡x] in the contemporary and lower proportions in the historical data. The contemporary DÄ data showed a tendency for the proportion of [k͡x] to decrease with speakers' age. This suggests that the feature may be spreading towards the area around the Northern lake of Zurich.

An explanation for the sound change taking place around the Northern lake of Zurich, and in the cantons of St. Gallen, and Appenzell, might be found when looking at the geographical reality of these areas: they lie in the Swiss midlands, where they are surrounded by influential dialect regions. Regions where results revealed high proportions of [k͡x] (cf. 3.2) lie in more mountainous and less accessible areas, where less dialect contact is expected [1]. One possible hypothesis for the sound change taking place in two different directions – with a diffusion towards Zurich and the Central Plateau and a retreat from Northeastern Switzerland – may be a wave-like process where the [k͡x] innovation has taken place earlier in the Northeast and is now progressing further towards the West of the Central Plateau, leaving the Northeast behind. However, the following methodological considerations should make it clear that such explanations for the patterns of change reported in this paper are hypothetical.

When comparing results based on DÄ data with those based on *Atlas* data, it is critical to highlight the different methodological framework that underlies the two approaches [9, 11]: (1) Data for the *Atlas* was collected in writing using a direct method, where dialectologists surveyed informants in the field; audio data for DÄ were collected using the indirect crowdsourcing method with little control over data elicitation. (2) Elderly informants were chosen as speakers for the *Atlas*. These speakers were elicited in the 1940s. For DÄ, speakers of all age groups were able to record themselves in 2013/14 (needless to say that the mobile app targeted more younger speakers, age mean=30.5, median=26, sd=15.6). Hence DÄ data reflects the SwG linguistic situation of about 100 years later than that given in the *Atlas*. (3) The *Atlas* features informants who – allegedly – spent their entire life in the respective locality. In the DÄ corpus, we find speakers from all walks of life, conceivably featuring wide mobility. (4) The *Atlas* surveyed one to two speakers per locality while DÄ elicited speech from a high number of speakers, which may reflect a more objective picture of the SwG population (e.g. average number of speakers: mean=4, median=2).

An audio corpus which is fully retrieved through a crowdsourcing mobile app has pitfalls: (1) for the recordings of *gha*, DÄ users are presented – and possibly primed – with the written Standard German form <gehabt>, which may prompt them to produce <geh-> as [kh] more often. If this had influenced speakers' pronunciation, we would have expected a higher proportion of [k͡x] than the one found in the present study. (2) Several speakers skipped the part in question or did not record the full dialogue altogether (n=17). Such cases were discarded. (3) For some recordings, the database featured entries where speech production was deemed overly unnatural (Standard German words (n=40), extensive pausing, disfluencies, and stuttering (n=10)), where the speaker submitted multiple recordings (n=52) [see 38, 39], or where the speaker recorded nonsense (n=27). Such recordings were discarded as well. (4) Some of the metadata indicated by users were also problematic: obvious wrong age (n=5) and gender (n=8) selections were corrected, and recordings with an obvious mismatch between the dialect indicated by the user and the speech in the recording (n=12) were dismissed.

How can we evaluate the validity of our crowdsourced data? In the current case, one may apply a more traditional dialectological approach to gather data on the pronunciation of *gha* in SwG and compare these results to the ones obtained through crowdsourcing by DÄ [40, 14]. One could, hypothetically, collect data in a rapid anonymous study in several SwG localities – analogous to the framework used in [41] – and compare the outcome to the results described in the present paper. Previous research has shown that results crowdsourced with DÄ strongly correlate with those collected through more traditional methods [14].

5. Acknowledgments

Warm thanks go out to 65 backers who sponsored DÄ through crowdfunding.

6. References

[1] R. Hotzenköcherle, *Die Sprachlandschaften der deutschen Schweiz.* Ed. by N. Bigler, R. Schläpfer, Aarau: Sauerländer, 1984.

[2] I. Werlen, "Zur Sprachsituation in der Schweiz mit besonderer Berücksichtigung der Diglossie in der Deutschschweiz," *Bulletin VALS-ASLA (Vereinigung für angewandte Linguistik in der Schweiz)*, no. 79, pp. 1–30, 2004.

[3] H. Christen, "Was Dialektbezeichnungen und Dialektattribuierungen über alltagsweltliche Konzeptualisierungen sprachlicher Heterogenität verraten," in C. Anders, M. Hundt, and A. Lasch (Eds.), "*Perceptual dialectology*" *– Neue Wege der Dialektologie*. Berlin/New York: de Gruyter, pp. 269–290, 2010.

[4] Bundesamt für Statistik, "Üblicherweise zu Hause gesprochene Sprachen," http://www.bfs.admin.ch/, 2013.

[5] P. Wiesinger "Die Einteilung der deutschen Dialekte," in W. Besch, W. Knoop, U. Putschke, and H. E. Wieland (Eds.), *Dialektologie – Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Berlin & New York: de Gruyter, pp. 807–900, 1983.

[6] SDS = *Sprachatlas der deutschen Schweiz*. Bern (I–IV), Basel: Francke (VII–VIII), 1962–2003.

[7] *Schweizerisches Idiotikon Idiotikon*. Frauenfeld: Huber, 1881–.

[8] A. Bachmann (Ed.), *Beiträge zur schweizerdeutschen Grammatik* (BSG), 20 vols., Frauenfeld: Huber, 1919–1941.

[9] R. Hotzenköcherle, *Einführung in den Sprachatlas der deutschen Schweiz*, *Band A – Zur Methodologie der Kleinraumatlanten*, Bern: Franke, 1962.

[10] A. Leemann and M.-J. Kolly, *Dialäkt Äpp*, https://itunes.apple.com/ch/app/dialakt-app/id606559705?mt=8, 2013.

[11] M.-J. Kolly and A. Leemann, "Dialäkt Äpp: Communicating dialectology to the public - crowdsourcing dialects from the public," in A. Leemann, M.-J. Kolly, S. Schmid, and V. Dellwo (Eds.), *Trends in Phonetics and Phonology. Studies from German-speaking Europe*. Bern: Peter Lang, in press.

[12] App Annie. http://www.appannie.com (accessed 27.02.2015), 2015.

[13] https://sites.google.com/site/dialaektaepp/public-media (accessed 26.03.2015).

[14] A. Leemann, M.-J. Kolly, R. Purves, D. Britain, and E. Glaser, "Crowdsourcing language change with smartphone apps," submitted to *PLoS ONE*.

[15] F. Hasler, *Die Berner Dialektlandschaft im Wandel. Eine vergleichende Untersuchung sprachgeografischer Strukturen*. Master Thesis, Department of General Linguistics, University of Bern, 2013.

[16] A. Leemann, M.-J. Kolly, and V .Dellwo. "Crowdsourcing regional variation in speaking rate through the iOS app 'Dialäkt Äpp'," *Proceedings of Speech Prosody*, 2014.

[17] J.-P. Goldman, A .Leemann, M.-J. Kolly, I. Hove, I. Almajai, V. Dellwo, and S. Moran. "A Crowdsourcing Smartphone Application for Swiss German – Putting language documentation in the hands of the users," *Proceedings of LREC*, 2014.

[18] M.-J. Kolly, A. Leemann, V. Dellwo, J.-P Goldman, I. Hove, and I. Almajai, "Voice Äpp – A smartphone application for crowdsourcing Swiss German dialect data," *Proceedings of Digital Humanities* 2014.

[19] A. Leemann, M.-J. Kolly, J.-P. Goldman, V. Dellwo, I. Hove, I. Almajai, S. Grimm, S. Robert, and D. Wanitsch, "Voice Äpp: a mobile app for crowdsourcing Swiss German dialect data," *Proceedings of Interspeech*, 2015.

[20] N. Renning, "Einfluss von Hintergrundmusik auf die Wahrnehmung des Alters durch die Stimme," term paper, Department of Comparative Linguistics, University of Zurich, 2013.

[21] N. Renning, "Hörerwahrnehmung des Geschlechts von Kinderstimmen aufgrund der Herkunft," term paper, Department of Comparative Linguistics, University of Zurich, 2013.

[22] A. Leemann, M.-J. Kolly, S. Grimm, S. Robert, S. Elspass, R. Möller, J. Fleischer, R. Kehrein, *Grüezi, Moin, Servus*, https://itunes.apple.com/ch/app/gruezi-moin-servus/id982633800?mt=8, 2015.

[23] *Ma Iwaidja*, https://itunes.apple.com/au/app/ma-iwaidja/id557824618?mt=8, 2012.

[24] F. R. Hanke and S. Byrd, "Large-scale text collection for unwritten languages," *International Joint Conference on Natural Language Processing*, pp. 1134-1138, 2013.

[25] T. Hughes, K. Nakajima, L. Ha, A. Vasu, P. Moreno, and M. LeBeau, "Building transcribed speech corpora quickly and cheaply for many languages," *Proceedings of Interspeech*, pp. 1914-1917, 2010.

[26] N. de Vries, M. H. Davel, J. Badenhorst, W. D. Basson, F. de Wet, E. Barnard, and A. de Waal, "A smartphone-based ASR data collection tool for under-resourced languages," *Speech Communication*, 56, pp. 119-131, 2014.

[27] J. Grimm, *Deutsche Grammatik*. Göttingen: Dieterichsch, 1999 [1819–1931].

[28] J. Fleischer and S. Schmid, "Zurich German," in *Journal of the International Phonetic Association*, vol. 36, no. 2, pp. 243–253, 2006.

[29] J. Bleiker, *Zur Morphologie und Sprachgeographie der Verben "haben, sein, tun" im Schweizerdeutschen.* Frauenfeld: Huber, 1969.

[30] W. Labov, *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1979.

[31] Phonogrammarchiv [Universität Zürich], *Der sprechende Atlas. "Gespräch am Neujahrstag" in 24 Dialekten.* Text by Eugen Dieth. Frauenfeld: Huber, 2000.

[32] Google, *Geocoding API,* http://developers.google.com/maps/documentation/geocoding/ (accessed 10.09.2014).

[33] Google, *Javascript API* http://developers.google.com/maps/documentation/javascript/, (accessed 10.09.2014).

[34] P. Auer, "The construction of linguistic borders and the linguistic construction of borders," in M. Filppula, J. Klemola, M. Palander and E. Penttilä, *Dialects Across Borders*. Amsterdam: Benjamins, pp. 3–30, 2005.

[35] C. Woolhiser. "Political borders and dialect divergence/convergence in Europe," in P. Auer, F. Hinskens and P. Kerswill (Eds.), *Dialect Change. Convergence and Divergence in European Languages*. Cambridge: Cambridge University Press, pp. 236–262, 2008.

[36] H. Christen, "Convergence and divergence in the Swiss German dialects," *Folia Linguistica*, vol. 32, pp. 53–67, 1998.

[37] B. Juska-Bacher, "Wortgeografischer Wandel im Schweizerdeutschen – Sommersprossen, Küchenzwiebel und Schmetterling 70 Jahre nach dem SDS," *Linguistik online*, vol. 42, pp. 19–42, 2010.

[38] M. H. Birnbaum, "Human research and data collection via the Internet," *Annual Review of Psychology*, vol. 55, pp. 803–832, 2004.

[39] U. D. Reips, "Standards for Internet-based experimenting," *Experimental psychology*, vol. 49, no. 4, p. 243, 2002.

[40] B. Juska-Bacher, C. Biemann, and U. Quasthoff, "Webbasierte linguistische Forschung: Möglichkeiten und Begrenzungen beim Umgang mit Massendaten," *Linguistik online*, vol. 61, pp. 7–29, 2012.

[41] A. Leemann, M.-J. Kolly, I. Werlen, D. Britain, and D. Studer-Joho, "The diffusion of /l/-vocalization in Swiss German," *Language Variation and Change,* vol. 26.2, pp. 191–218, 2014.