Abstract 9

10 Increased geographical mobility motivates dialectologists to consider the exposure of survey participants to linguistic 11 variation. Changing mobility patterns (e.g. longer-distance commuting; easier relocation to distant places for work, study or 12 marriage) have caused linguistic connections to become much more diverse, contributing to an acceleration of dialect change. 13 To assess the impact of individual mobility on this change, we propose the Linguistic Mobility Index (LMI) framework, which 14 estimates long-term exposure to dialectal variation based on episodes of linguistic biography. An LMI of a survey participant 15 comprises combinations of influencing factors, such as dialects of parents, long-term partners, places lived, place of work and 16 education. The linguistic effects of these factors are represented by linguistic distances to the survey participant in question, 17 and the effects are cumulated into an LMI in a weighted manner, according to the relationship the factor embodies and the 18 intensity of the participant's exposure to the factor. LMI is conceptualised and evaluated based on 500 speakers from 125 19 localities in the Swiss German Dialects Across Time and Space (SDATS) corpus. Four LMI prototypes are constructed, 20 employing different theoretical considerations and combinations of influencing factors to simulate the availability of metadata 21 in other studies, thereby assessing the generalisability of the framework. Using mixed-effects modelling, we evaluate the utility 22 of the LMI prototypes as predictors of dialect change between historic and contemporary linguistic data of Swiss German. The 23 LMI prototypes successfully show that higher exposure to dialectal variation contributes to more dialect change and that its 24 effect is stronger than some of those sociodemographic variables often tested for similar effects (e.g. sex and educational 25 background). The success of the four prototypes justifies the potential implementation of the LMI framework in other studies, 26 including those with a limited amount of metadata, for which we also provide further guidance in the contribution.

Introduction

In this paper, we argue that quantifying exposure to other dialects at the speaker level may provide researchers with a new tool for investigating language variation and change. Increased mobility jeopardises the validity of region being the primary determinant of linguistic variation [1], due to mobility-induced dialect change. Mobility leads to a potential increase in contact, exposing individuals to linguistic variation, and the intensity of this exposure plays a key role in language change [2]. Due to this increasing exposure to linguistic variation, it is indispensable to address the mobility of participants in linguistic surveys [3,4]. This paper introduces the framework of the Linguistic Mobility Index (LMI), a tool to estimate individuals' exposure to potential linguistic influences through examining long-term mobility patterns in their linguistic biographies. Our

27

35 focus is on dialectology and sociolinguistics, but we also encourage the application of the linguistic mobility framework in 36 other fields within linguistics.

37 Mobile behaviour results in potential contact with peers in different localities and long-term exposure to linguistic 38 variation (e.g. through regular contact, such as via commuting or relocation). Previous dialectological and sociolinguistic 39 studies have researched mobility and exposure to contact in relation to different linguistic aspects (e.g. [1-3,5-9]). We view 40 linguistic mobility as reflecting the combined potential effects of the places visited or contacts made (i.e. contact with peers 41 from these places) on an individual's dialect. Thus, a linguistically mobile person is characterised by activities that bring them 42 into contact with linguistically different localities, such as multiple relocations throughout their lives, routinely commuting for 43 study, work or other regular activities, or having familiar ties in such localities. In contrast, a linguistically non-mobile person 44 would gain most of their linguistic influences from the inhabitants of the locality in which they grew up.

45 The constant increase in geographical mobility over the last century has strengthened the potential of diverse linguistic 46 contacts to impact dialects [10-12]. This linguistic change caused by the intensification of dialect contact among speakers from 47 a larger number of places has been framed in dialectology as a function of the general mobility patterns of the population, 48 conceptualised, for example, as the wave model of language change [13] and the linguistic gravity model [9]. However, little 49 quantitative research has been conducted on the effects of mobility at the speaker level, partly because surveys traditionally 50 focused on capturing variation elicited from NORMs and NORFs, cf. [14]. Of the available studies on geographical mobility 51 of individuals, Chambers [1] tested his Regionality Index (RI) on three Canadian city-wide lexical databases, Beaman [5] 52 studied the role of relocation in the attrition of Swabian German through observing dialect change within the lifespan of 53 individuals, while Bowie [6] and Regan [8] observed the retention of phonological forms and, respectively, change in the 54 perceived socioeconomic status of words, based on the number of years spent away from a reference locality. Moreover, 55 Chambers' RI [1] was applied in studies to account for the extent to which individuals could represent local communities [15-56 17].

In terms of quantifying exposure to the linguistic effects of contact, research has focused on the relation between language change and the most important linguistic influences in individuals' lives. An individual's linguistic inheritance is viewed as coming from the dialects of their parents and primary caregivers, who have a foundational influence on their dialect, especially in early childhood before large-scale exposure to older peers (e.g. [18–21]). These foundations are then strongly shaped by other intense contacts, including relatives; peers during childhood and adolescence [21], such as at school [22,23];

partners [24,25]; and contacts within the workplace and other communities as adults [7,26]. These factors are important to take
 into consideration when studying linguistic mobility.

To date, quantitative measures of the linguistic exposure of individuals to other dialects have not been determined based on aggregating biographical information and other influential factors from survey metadata; instead, quantification of linguistic mobility has been rather elementary, such as focusing on flat rates of time spent away from a reference locality [5,6,8]. The LMI framework addresses this research gap by systematically constructing an index from multiple linguistically influential factors that can be extracted from survey metadata. Thus, LMI integrates linguistic biography data into a single value representing exposure to linguistic variation.

70 In this paper we implement the LMI framework using data from 500 survey participants recorded in the SDATS 71 corpus (*Swiss German Dialects Across Time and Space* [27]). Four LMI prototypes are constructed which simulate the 72 availability of metadata in other dialect surveys, thereby testing the generalisability and flexibility of the LMI framework.

73 We evaluate the usefulness of the implemented LMI by testing its relation to dialect change and controlling for 74 variation in sociodemographic variables frequently used to assess language change. Other studies have included in such tasks, 75 depending on the design of the linguistic survey, for instance, gender (e.g. [28,29]), educational background (e.g. [30]), and 76 urbanity and social networks (e.g. [31-35]). Our evaluation tests the performance of the four LMI prototypes as predictors in 77 mixed-effects models. A dialect change rate is calculated based on ten lexical variables, using historical linguistic data from 78 the SDS (Sprachatlas der deutschen Schweiz [36]) and contemporary data from the SDATS project. We expect the model 79 results to confirm that linguistically mobile speakers have higher rates of dialect change, while non-mobile speakers have lower 80 rates.

81 The remainder of this paper is structured as follows. After outlining the general construction of LMI, we present the 82 implementation of LMI for the SDATS corpus and evaluate the four LMI prototypes in mixed-effects models. In addition, we 83 briefly address the question of regionality and urbanity regarding the effects of mobility. After presenting and discussing the 84 results of the models, we indicate the limitations of LMI and provide recommendations for the application of LMI in other 85 studies.

86 Methods

87 Introducing the Linguistic Mobility Index

88 The core concept of LMI is the aggregation of the linguistic effects of linguistic connections in a participant's 89 biography. These linguistic connections (henceforth referred to as 'factors') may be people and groups that the participant has 90 been exposed to, and which have had linguistic effects on them. LMI estimates a summary of these linguistic effects 91 accumulated throughout the life of the participant, and the effects are weighted according to the intensity of the contact and the 92 participant's relation to the factor. Linguistic surveys with a controlled selection of participants usually elicit some metadata 93 about their participants and the influential factors in question. Usually, surveys assign each participant to a reference locality 94 or linguistic variety, and a location could also be assigned to the factors. The potential linguistic effect is determined based on 95 these assigned locations through the calculation of linguistic distance. The intensity of the participant's exposure to these 96 locations, and their relation, is expressed based on information elicited from survey questionnaires or other metadata (e.g. 97 through what kind of personal relation they were exposed to the location, or how much time they spent there). In comparison 98 to RI, which is used for measuring the extent to which an individual has been exposed to a reference locality, LMI measures 99 the potential effect of the linguistic variation pertaining to places encountered by the participant outside the reference locality. 100 A generic solution to calculate the exposure of an individual to one linguistically influential factor is based on the 101 following steps. 102 Calculating linguistic distance: The basis of LMI is linguistic distance, which is a quantitative estimate of 103 the difference between the linguistic variety pertaining to the factor and the participant's own variety. 104 Weighting based on the intensity of the exposure: Fine-tuning the estimation of the factor's effect based 105 on the available metadata in the linguistic biography (e.g. age at the time of contact, duration and frequency 106 of the contact).

Weighting based on the relationship to the factor: Estimating the possible role of the factor in the
 individual's life and thereby its long-term influence, in a categorical manner.

109 LMI is then created as an aggregation of the exposure to the factors considered. The larger the resulting LMI, the more 110 long-term exposure the participant had to dialectal variation.

111 Implementing the LMI framework on Swiss German dialect data

112Before embarking on the implementation of the LMI framework, it is important to take note of the specific linguistic113situation of Swiss German dialects. In the diglossic context of Switzerland, dialects enjoy high prestige compared to Standard

German (e.g. [37,38]) and do not have a strong orientation to social class. Swiss German speakers (practically all German speakers born or raised in Switzerland) use their own dialects colloquially, with the lingua franca between speakers of various dialects rarely being the standard language, although 'levelling' is occurring in Swiss German dialects [19,39]. Use of Standard German is expected only in a few official situations, but Swiss German speakers are constantly exposed to Standard German from an early age, such as in school and via the media. The diverse topography and the (historical) administrative structure of Switzerland cause specific mobility flows and cultural orientations, contributing to the regional diversity of the dialects. Thus, spatial patterns in linguistic mobility and, correspondingly, in dialect change are expected.

121 The LMI framework was implemented using information elicited in the SDATS metadata about various long-term 122 linguistic influences. SDATS speakers were recorded in 2020-2021 across 125 localities in German-speaking Switzerland, 123 which form a subset of the SDS survey localities [40]. From each of these reference localities, we used data from four 124 participants: two older (65+ years old) and two younger speakers (20-35 years old), with one male and one female speaker in 125 both age cohorts. For every speaker, one parent came from the region of the reference locality, and the speakers themselves 126 grew up in and lived most of their lives there. Further, their daily travel time was required to not exceed the Swiss average of 127 approximately two hours. After a dialect interview, the speakers filled out an unsupervised online metadata questionnaire 128 consisting of over 300 items (cf. [41] for details). The SDATS participants were properly instructed and indicated their consent 129 to participate by signing an appropriate consent form, approved by the Legal Services Office of the University of Bern. The 130 participants explicitly consented to the anonymous analysis of the data they provided and were informed about the applicability 131 of the Data Privacy Act of the Canton of Berne (Datenschutzgesetz des Kantons Bern - KDSG; BSG 152.04, from 19.02.1986). 132 This procedure of collecting and analysing anonymous user data conforms to the regulations of the Bern cantonal ethics 133 committee (https://www.gsi.be.ch/de/start/ueber-uns/kommissionen-gsi/ethikkommission.html) and the accompanying federal 134 act on research involving human beings in Switzerland (https://www.fedlex.admin.ch/eli/cc/2013/617/en). For this reason, we 135 did not seek further ethical approval from cantonal or federal institutional bodies.

Through the example of Speaker A, we will demonstrate the construction of LMI based on SDATS metadata. Figure 1 also illustrates the steps followed in constructing LMI. Information on the data preparation, construction of LMI and data modelling can be accessed in the Supplementary Material, along with the corresponding R source code in **S1_Appendix** and **S2 Appendix**, at https://osf.io/hfbpk/. 140 Implementing the LMI framework started by determining the relevant linguistically influential factors that could be 141 retrieved based on the speakers' biographical data. We constructed the following factors: mother, father, long-term partner 142 (henceforth: partner), residence where a speaker has lived outside the reference locality (henceforth: external residence), place 143 of ongoing education and current workplace. Each factor contributes to LMI, in the manner detailed below.





145 Figure 1 Example calculation of LMI components, demonstrating the three main steps of the process: linguistic

146 **distance calculation, exposure weights and relational weights.** The linguistic distance values are multiplied by the exposure

147 weights and by the relational weights, and then the results are summed.

149 Linguistic distance

The potential linguistic effect of a factor on a speaker is quantified by calculating a linguistic distance. As the first step in this quantification, a locality is assigned to the factor through geocoding. A linguistic distance is then calculated between the reference locality of the speaker and the locality assigned to the factor.

We assigned an SDS survey locality (*n*=565) to each factor in the following way. In R [42] (version 4.0.4.), the locality names associated with the factors, as recorded in the metadata questionnaire, were matched to geographical coordinates by means of geocoding, using the *tidygeocoder* package [43] (version 1.0.3.). Then, solving point-in-polygon problems using the Voronoi polygons of SDS localities, the closest SDS survey locality was assigned to the factor with the help of the *sf* package [44] (version 1.0-2.). Speaker A named Bern as the locality of his workplace. Geocoding returned the geographic coordinates of the capital city, Bern. The point-in-polygon routine then determined that the coordinates are found in the Voronoi polygon of the SDS locality 'Bern'.

160 As the next step, linguistic distances were determined between the assigned locality and the reference locality of the 161 speaker. These linguistic distances were calculated based on Goebl's Relative Identity Value RIV_{ik} [45], using a portion of the 162 SDS variables (289 linguistic variables: 107 phonetic, 118 morphosyntactic and 64 lexical variables), which were digitised by 163 Scherrer and his colleagues [46], and spanning from 0 to 1 (i.e. 0=linguistically identical and 1=total linguistic discrepancy). 164 Based on this data, a linguistic distance matrix was set up, consisting of linguistic distances for each pair of SDS localities, 165 calculated as follows [47]. For each linguistic variable, variant categories were constructed based on phonetic similarity. Within 166 these variant categories, a further distinction was made between subvariants. The linguistic distance between a locality pair is 167 the proportion of variables (among those *n* variables that had data for both localities) that differ regarding the variant categories. 168 If the variant categories match for a variable, but the subvariants differ, the linguistic distance still grows by a smaller amount. 169 At variable level this can be written as

170
$$D_{ij}^{ling} = \frac{\Sigma D_Q}{n}$$

171 where D_Q is the number of differing variables regarding localities *i* and *j*. For the workplace of Speaker A, the linguistic 172 distance between Bern and Brig (the reference locality of Speaker A) amounts to 0.46, which means that about half of the 173 linguistic features considered are different across the two localities. The linguistic effect of the reference locality received a flat rate of 0 because it is known that the speakers grew up in the reference locality and, lacking additional information, we infer the peer effect to be local in the first 10–20 years of language acquisition. The same is the case for places that are officially not German speaking, as we do not consider their linguistic impact on the speaker's dialect. Places in Germany and Austria, representing Standard German, received a flat rate of 0.5 as linguistic distance, regardless of possible local dialectal influence. As a matter of fact, there are fewer than 200 places in Austria and Germany among the 3000+ places geolocated.

180 **Exposure weights**

181 Using exposure weights, which quantify the intensity of the speaker's contact with the factors, we fine-tuned the 182 degree to which the effect of a factor is considered in LMI. The factors may have various associated parameters in the metadata, 183 which can be compiled into weights to quantify the factors' linguistic influence on the speaker. For each factor, linguistic 184 distance is multiplied by the exposure weight value (ranging from 0 to 1). For Speaker A's workplace, for example, the exposure 185 weighting involves the proportion of Swiss German and Standard German used at work, and the years he has been working 186 there, and was implemented as a multiplication (Figure 1). Exposure to Standard German among the calculation of exposure 187 weights also used the flat rate of 0.5. Exposure weights are specific to the kinds of factors included and regulate the effect of 188 these factors across speakers. The calculation of the exposure weights for each factor are explained in detail in S1 Appendix.

189 **Relational weights**

Linguistic exposure also depends on the nature of the relation an individual has to the locality (e.g. through a certain contact person or life situation). We can assume more intensive contact with the local variety through a person that is most likely from the locality in question (e.g. a partner) than through a factor more vaguely connected to the locality (e.g. studies, workplace) through which the speaker potentially meets a more mixed linguistic community.

194 *Relational weights* are independent of metadata items and of the possibility to calculate exposure weights. For a 195 specific factor, every speaker receives the same relational weight as a multiplier to differentiate the assumed effect, for example, 196 of the workplace from the effect of other factors, such as parents. In the case of Speaker A, after multiplying the linguistic 197 distance pertaining to his workplace with the exposure weight as the quantifier of the intensity of his contact with the workplace 198 (Figure 1A), this was also multiplied by the relational weight associated with workplace as a factor, 0.47 (Figure 1B). 199 To assign objective weights for the current study, we implemented a modelling procedure. We set up four mixed-200 effects models for estimating relational weights for specific factors, using their corresponding β -coefficients as relational 201 weights:

Relational weight of 'place of education' for those currently in education (n=119);

203 Relational weight of 'partner' for those with a partner identified in the metadata (n=351); 204 Relational weight of 'workplace' for those that indicated a workplace (n=306); and 205 Relational weights of 'mother', 'father' and 'external residence' in a general model (n=500). 206 The models use the unweighted linguistic distances as fixed effects alongside the control variables age, sex and 207 educational background, all of which were considered in the participant selection of the SDATS survey (henceforth: survey 208 design variables). The models predict a dialect change rate based on ten lexical variables (also used for the evaluation of LMI, 209 thus detailed further). Speaker identifier and the linguistic variable are used as random intercepts in the models. In terms of the 210 missing values we used bootstrapped regression imputation to impute values for the partner's and the workplace's linguistic 211 distance, with the help of the *mice* package [48] (version 3.14.0). For the details of this modelling procedure, see S1 Appendix. 212 The resulting relational weights are highlighted in Figure 1B.

213 Setting up the LMI prototypes

We constructed four LMI prototypes based on the factors, exposure weights and relational weights. The prototypes are related to different theoretical considerations and simulate possible scenarios of metadata availability. We assigned a name and an abbreviation to each prototype. Table 1 presents the components involved in each LMI prototype along with the weights associated with the factors involved.

218 Minimal prototype (LMI^A)

The minimal prototype simulates a survey where only a little background information is available about the speakers. We included the origin of the father and the origin of the mother as factors, considering that many surveys include this information to account for authenticity. The factors are weighted only by relational weights.

222 Cumulative prototype (LMI^B)

Surveys often elicit various pieces of background information (e.g. origin of parents, partners and peers, former places of residence, places of education and work) without gathering detailed information about the possible effect of these localities (e.g. [49]). The cumulative prototype sums up the linguistic distances from all factors in a linear manner, disregarding exposure weights and using only the relational weights. Maximising the effect of each factor (i.e. each locality) considered, this prototype basically cumulates influential factors in a linguistic biography, assigning dialectal locations to them.

228 Comprehensive prototype (LMI^C)

All the aforementioned factors' effects (i.e. mother, father, partner, external residence, workplace and place of ongoing education) are included in LMI^C, using exposure weights and the relational weights associated with them. The parameters of this prototype correspond to a scenario where a larger range of metadata is elicited from which one could estimate exposure weights for some of the components.

233 Cohort-based prototype (LMI^D)

234 The cohort-based prototype is tailored for the specific situation in SDATS. It accounts for the increased mobility of 235 the last 50 years through assessing the history of potential exposure in the two age cohorts differently. The younger SDATS 236 cohort (20-35 y.o. in 2020) has grown up in different circumstances regarding dialect acquisition and change compared to the 237 older cohort (60–80 v.o. in 2020) [19]. With societal changes, the increasing likelihood of geographical mobility and the access 238 to a wider variety of media, the younger cohort has potentially more intense exposure to other dialects and to the standard 239 language than the older cohort did at the same age. As we will show later, age is the strongest design variable in our survey to 240 explain language change: younger speakers tend to show more dialect change. This is also because more time has elapsed 241 between the SDS and the younger speakers' dialect acquisition, allowing more time for dialects to change. That is, the younger 242 cohort has a different baseline against which their dialect might change over their lifespan [21]. Although dialect-related 243 identities and attitudes may also change, in the case of Swiss German, the SDATS data shows that the strength of identity is 244 maintained [50].

The relational weights are constant across the LMI prototypes. For LMI^D, however, we have implemented age-cohortbased relational weights for the 'workplace' and 'external residence' factors (**S1 and S2_Appendix**), incorporating information that age provides about the speakers and overcoming the difference in the metadata available about the two cohorts. This allows the younger cohort to gain higher LMI values than in the other LMI prototypes.

Factor	LMI ^A	LMI ^B	LMI ^C	LMI ^D
Mother's origin	Ling. Dist.* 0.6351	Ling. Dist. * 0.6351	Ling. Dist. * 0.6351	Ling. Dist. * 0.6351
Father's origin	Ling. Dist. * 0.317	Ling. Dist. * 0.317	Ling. Dist. * 0.317	Ling. Dist. * 0.317
Partner's origin	-	Ling. Dist. * 0.3461	Ling. Dist. * weight _{exposure} * 0.3461	Ling. Dist. * weight _{exposure} * 0.3461
External residence	-	Σ Ling. Dist. * 0.0319	Σ Ling. Dist. * weight _{exposure} * 0.0319	Older cohort: Σ Ling. Dist. * weightexposure * 0.0206 Younger cohort: Σ Ling. Dist. * weightexposure * 0.1663
Workplace	-	Ling. Dist. * 0.4728	Ling. Dist. * weight _{exposure} * 0.4728	Older cohort: Ling. Dist. * weightexposure * -0.8728 Younger cohort: Ling. Dist. * weightexposure * 0.6061
Place of education	-	Ling. Dist. * 0.2023	Ling. Dist. * weight _{exposure} * 0.2023	Ling. Dist. * weight _{exposure} * 0.2023

249 Table 1 Composition of the four LMI prototypes.

250

Relational weights are indicated numerically.

251

Table 1 summarises the components included in the four LMI prototypes and the corresponding weighting.

252 Evaluating LMI as a predictor of dialect change

We will evaluate LMI by testing the performance of the four prototypes as predictors of dialect change. Figure 2A–D shows the relation of the four LMI prototypes to age, with the shape of the symbols representing sex and colours representing educational background. Speakers without tertiary education appear to be less mobile, while a number of younger speakers with no or ongoing higher-level education form a cluster on the left side of Figure 2A–D. For comparison, Figure 2E presents the relation of age, sex, educational background and the dialect change rate used for the evaluation. All panels in Figure 2 show that younger speakers have, on average, a slightly higher LMI and dialect change rate, respectively.







Gender

- female
- male

Educational background

- •
- with tertiary education without tertiary education •

Figure 2 Distribution of LMI and dialect change with regards to age. In Panels (A)-(D) the distribution of standardised
 LMI⁴, LMI^B, LMI^C and LMI^D values are shown (y-axes) against the age of the speakers (x-axes). Panel (E) shows the
 distribution of standardised dialect change rates, against the age of the speakers. Each point represents a speaker (n=500).
 Point colour represents educational level and shape represents sex. The blue concentrical lines show the density of speakers.

- 264 To test the utility of LMI as a predictor of language change, the four LMI prototypes were modelled as fixed effects 265 in logistic mixed-effects regression models. We expect the model results to confirm, while controlling for the survey design 266 variables, that linguistically mobile speakers have higher rates of dialect change and less mobile speakers have lower rates. The 267 outcome variable explained in the models is change rate in lexical variables, calculated based on ten items (Table 2) recorded 268 at 125 survey localities which were included in both SDS and in SDATS, approximately 70 years apart. The ten lexical items 269 were chosen to represent maximal expected variation in terms of dialect change and different word frequencies. Lexical items, 270 also including those investigated by [19], were chosen specifically to capture substantial dialect change as the lexical level 271 changes faster compared to grammatical linguistic levels [51]. Table 2 shows the rates of change in the 500 speakers and the 272 word frequency in Switzerland in 2019 based on Google Books Ngrams [52]. Most change occurred in the word 'freckles' (Std. 273 Germ.: 'Sommersprossen') followed by 'butterfly' (Std. Germ.: 'Schmetterling'), while other lexical items show a smaller, 274 comparable rate of change. The goal of the modelling for this evaluation was neither to find the perfect model for predicting 275 dialect change nor to explain language change at the level of individual SDATS speakers or linguistic items. The evaluation 276 also tests whether LMI has validity for explaining dialect change that may not have happened within the lifetime of the speakers, 277 but which is partly hereditary.
- 278 Table 2 The ten lexical items used in the evaluation study.

SDS	Standard German	Dialectal variants	English	Proportion of	Frequency
map				change from SDS	in Switzerland
number					in 2019
V 179	Butter	Butter, Schmalz, Anke etc.	butter	33.2%	0.001'425%
VI 237	Schmetterling	Summervogel, Fifolter, Schmätterling etc.	butterfly	48.4%	0.000'316%
V 212	Bonbon	Zuckerli, Täfeli, Tröpsli, Bombom, Guetsch etc.	(hard) candy	29%	0.000'073%
IV 17	Wange	Wang, Wang(j)i, Backe (with or without fricative second consonant)	cheek	36.4%	0.003'38%

IV 43	Sommersprossen	Laubfläcke, Summersprosse, Merzetupf, Merzedräck etc.	freckles	72.8%	0.000'234%
IV 71	Schluckauf	Hitzgi, Gluggsi, Hixer, Hösch etc.	hiccup	41.8%	0.000'065%
V 21	Kuss	Kuss, Schmutz, Müntschi, Muntsi etc.	kiss	28%	0.004'093%
VI 179	Zwiebel	Zibele, Zwible, Bele, Böl(l)e etc.	onion	28%	0.000'55%
VI 40	Pfütze	Glunte, Gumpe, Gülle, Gudle, Glungge, Lache, Pütze etc.	puddle	42%	0.000'231%
V 139, V 140	Taschentuch	Nastuech, Naselumpe, Schnupftuech, Fazeneetli etc.	tissue (hanky)	28.2%	0.000'787%

279

The data on word frequency in Switzerland in 2019 comes from Google Books Ngrams, based on the Standard German

version of the words.

281 **Predictors of dialect change**

282 Dialect change rates show differences with regard to age, sex and educational background. We test these empirical observations, 283 often used in sociolinguistics for dialect change modelling, in simple linear regression models and we control for them in the 284 mixed-effect models of the evaluation. Due to the differing baselines of dialect change, we expect that any predictor of language 285 change would deliver more noisy results for the younger cohort, which makes it crucial to control for age cohort in our study. 286 The baseline for dialect change would also be expected to vary in space, making the spatial origin of individuals an important 287 predictor of dialect change. However, in this evaluation, we avoided including spatial variation directly in the modelling due 288 to the following reasons. On the one hand, accounting for spatial variation in Switzerland may not bring additional value to 289 establishing LMI as a general framework for language change studies, which may include scenarios where spatial variation in 290 the speaker sample is irrelevant. On the other hand, four speakers in 125 survey localities is not an optimal number for 291 accounting for categorical effects in tests that assume normality. To characterise the spatial distribution of the dialect change 292 rate and linguistic mobility, we tested the similarity of localities with regard to the values (using the Kruskal-Wallis test) and 293 the clustering of values in space (measuring the spatial autocorrelation using Moran's I). Additionally, we tested the effect of 294 urbanity, operationalised based on the population of the SDATS survey sites in 2018.

295 Mixed-effects modelling

The mixed-effects modelling was implemented in *R* using the *lme4* package [53] (version 1.1-27.1). Each observation in the dataset represents a combination of a speaker (n=500) and an item (n=10), amounting to 4983 observations after removing invalid or missing answers. As fixed effects, each model includes one of the four LMI prototypes, together with the survey design variables (Table 3). All fixed effects were z-standardised (Equation 2) to facilitate the interpretation of the model results [54].

301 [Equation 2 about here]

$$\frac{x - \mu_x}{2\sigma_x}$$

Binary fixed effects were contrast coded to zeroes and ones in the order of expected dialect change (z-standardised to -0.5 and 0.5). The standardised LMI values range between -1.62 and 2.14. Due to the presence of pseudoreplication (all speakers answered each of the ten questions), we included speaker and item as random effects in the mixed-effects models. This also controlled for the differences in the frequency with which words occur, as more frequent forms are expected to be more resistant to change [55]. Collinearity was tested through the analysis of variance inflation factors (VIFs), using the *car* package [56] (version 3.0-10.), resulting in values only slightly larger than 1; thus, collinearity problems were not expected. For more details, consult S2 Appendix.

- 507 consult 52_Appendix.
- 310 Table 3 The variables entered in the mixed-effects models of the evaluation

Outcome variable	
Dialect change in item	Change (1), or no change (0) in comparison to
	SDS data in the same locality
Fixed effects	
Linguistic Mobility Index (LMI)	Continuous variable – One of the four LMI
	prototypes, z-standardised (the four LMI
	prototypes LMI ^A , LMI ^B , LMI ^C and LMI ^D are
	used in separate models)
Age cohort	Binary variable – 'older' > 45 years (-0.5);
-	'younger' < 45 years old (0.5)
Sex	Binary variable – female (-0.5) and male (0.5)
Highest completed education	Binary variable – with tertiary education
	background (-0.5) or without (0.5)
Random effects	
Speaker	n = 500
Item	n = 10

311

312 **Results**

313 In this section we report the outputs of the mixed-effects models described above, in order to evaluate the LMI 314 framework as a useful heuristic for language variation and change studies. In addition, to contextualise the correspondence of 315 linguistic mobility and dialect change, we characterise the spatial patterns of dialect change rate and the LMI prototypes. First, 316 however, we briefly explore simple linear regression models of the survey design variables, LMI prototypes and dialect change. 317 Age cohorts and sex are statistically significant predictors of dialect change, while the difference relating to education is not 318 (see also Figure 3). The younger cohort shows more change ($\mu = 47.03\%$, SD = 18.75%) than the older cohort ($\mu = 30.8\%$, SD319 = 17.18%), with slightly more change occurring among males ($\mu = 40.77\%$, SD = 20.16%) than among females ($\mu = 37.07\%$, 320 SD = 19.13%). In terms of the four LMI prototypes (scaled and centred), although all four show significant predictive power, 321 the correlation coefficients determine that LMI^B (the cumulative prototype, $R^2 = 0.1073$) and LMI^D (the cohort-based prototype, 322 $R^2 = 0.1014$) are better as sole predictors of the dialect change rate than the other prototypes (Figure 4).





Figure 3 Dialect change rate for the ten lexical items by age cohort, sex and educational background



325

Figure 4 The relation of the four LMI prototypes to the dialect change rate. LMI values and dialect change rates are standardised. The panels also show the numerical results of the linear regression models. Linear (red) and second-order polynomial regression lines (green) show the major trends. The slope of the lines shows the positive correlation.

In the mixed-effects models, fixed effects prove to be significant predictors, except for educational background (Table 4). The LMI prototypes' predictive power in increasing order of their *z*-values is A (minimal), C (comprehensive), B (cumulative), D (cohort-based). Their highly significant effects are coupled with low standard error (*SE*). The smallest *AIC* (Akaike Information Criterion), characterising model quality, also belongs to LMI^D, which makes it the best model out of the four. Slope estimates and *SE* are on the log-odds scale and must be exponentiated for a more accessible interpretation. For example, two standard deviations of increase in LMI^B, $e^{0.4395} = 1.5519$ and the corresponding *SE*, $e^{0.0784} = 1.08155$ mean a 55.19% (± 8.15%) increase in odds for dialect change.

336 Table 4 Output of the mixed-effects models involving the survey design variables of the SDATS.

LMI ^A	Model quality	<i>AIC</i> 6129.63	Random effects	RandomSD_{speaker}effects0.4809			
	Fixed effects	estimate	SE	z-value	р		

	LMI	0.3486	0.0755	4.6152	<0.001
	Age cohort	0.7691	0.0771	9.9754	<0.001
	Sex	0.2022	0.0777	2.6025	0.0093
	Education	0.0901	0.0811	1.1105	0.2668
	Madal anality	AIC	Random	$SD_{speaker}$	SD_{item}
	Niodel quality	6119.76	effects	0.4663	0.6129
	Fixed effects	estimate	SE	z-value	р
LMI ^B	LMI	0.4395	0.0784	5.6041	<0.001
	Age cohort	0.6681	0.0789	8.4704	<0.001
	Sex	0.1919	0.0769	2.4942	0.0126
	Education	0.1374	0.0806	1.7055	0.0881
	Model quality	AIC	Random	$SD_{speaker}$	SD_{item}
		C1 20 10	66 4	0 4670	0 (1)
	1 0	6120.18	effects	0.4672	0.613
	Fixed effects	estimate	SE SE	0.4672 <i>z-value</i>	0.613
LMI ^C	Fixed effects LMI	6120.18 estimate 0.4203	<i>SE</i> 0.0755	0.4672 <i>z-value</i> 5.5694	0.613 p <0.001
LMI ^C	Fixed effects LMI Age cohort	6120.18 <i>estimate</i> 0.4203 0.7596	SE 0.0755 0.0764	0.4672 <i>z-value</i> 5.5694 9.9386	0.613 <u>p</u> <0.001 <0.001
LMI ^C	Fixed effects LMI Age cohort Sex	6120.18 estimate 0.4203 0.7596 0.1797	<i>SE</i> 0.0755 0.0764 0.077	0.4672 <i>z-value</i> 5.5694 9.9386 2.3325	0.613 <u>p</u> <0.001 <0.001 0.0197
LMI ^C	Fixed effectsLMIAge cohortSexEducation	estimate 0.4203 0.7596 0.1797 0.1228	SE 0.0755 0.0764 0.077 0.0804	0.4672 z-value 5.5694 9.9386 2.3325 1.5264	0.613 <u>p</u> <0.001 <0.001 0.0197 0.1269
LMI ^C	Fixed effects LMI Age cohort Sex Education Model quality	6120.18 estimate 0.4203 0.7596 0.1797 0.1228 AIC	SE 0.0755 0.0764 0.077 0.0804 Random	0.4672 <i>z-value</i> 5.5694 9.9386 2.3325 1.5264 <i>SD</i> _{speaker}	0.613 <u>p</u> <0.001 <0.001 0.0197 0.1269 SD _{item}
LMI ^C	Fixed effects LMI Age cohort Sex Education Model quality	6120.18 estimate 0.4203 0.7596 0.1797 0.1228 AIC 6115.63	effects SE 0.0755 0.0764 0.077 0.0804 Random effects	0.4672 <i>z-value</i> 5.5694 9.9386 2.3325 1.5264 <i>SD</i> _{speaker} 0.4602	0.613 p <0.001 <0.001 0.0197 0.1269 SD _{item} 0.6129
LMI ^C	Fixed effects LMI Age cohort Sex Education Model quality Fixed effects	6120.18 estimate 0.4203 0.7596 0.1797 0.1228 AIC 6115.63 estimate	effects SE 0.0755 0.0764 0.077 0.0804 Random effects SE	0.4672 <i>z-value</i> 5.5694 9.9386 2.3325 1.5264 <i>SD</i> _{speaker} 0.4602 <i>z-value</i>	0.613 <u>p</u> <0.001 <0.001 0.0197 0.1269 <u>SD_{item}</u> 0.6129 <u>p</u>
LMI ^C	Fixed effects LMI Age cohort Sex Education Model quality Fixed effects LMI	6120.18 estimate 0.4203 0.7596 0.1797 0.1228 AIC 6115.63 estimate 0.4602	effects SE 0.0755 0.0764 0.077 0.0804 Random effects SE 0.077	0.4672 <i>z-value</i> 5.5694 9.9386 2.3325 1.5264 <i>SD</i> _{speaker} 0.4602 <i>z-value</i> 5.9714	p <0.001 <0.001 0.0197 0.1269 SD_item 0.6129 p <0.001
LMI ^C	Fixed effectsLMIAge cohortSexEducationModel qualityFixed effectsLMIAge cohort	6120.18 estimate 0.4203 0.7596 0.1797 0.1228 AIC 6115.63 estimate 0.4602 0.6934	effects SE 0.0755 0.0764 0.077 0.0804 Random effects SE 0.077 0.0804	0.4672 <i>z-value</i> 5.5694 9.9386 2.3325 1.5264 <i>SD</i> _{speaker} 0.4602 <i>z-value</i> 5.9714 8.9693	p <0.001 <0.001 0.0197 0.1269 SD _{item} 0.6129 <0.001 <0.001 <0.001
LMI ^C	Fixed effectsLMIAge cohortSexEducationModel qualityFixed effectsLMIAge cohortSex	6120.18 estimate 0.4203 0.7596 0.1797 0.1228 AIC 6115.63 estimate 0.4602 0.6934 0.1859	effects SE 0.0755 0.0764 0.077 0.0804 Random effects SE 0.077 0.073 0.0766	0.4672 z-value 5.5694 9.9386 2.3325 1.5264 SD _{speaker} 0.4602 z-value 5.9714 8.9693 2.4261	p <0.001 <0.001 0.0197 0.1269 SD _{item} 0.6129 p <0.001 <0.001 <0.001 <0.001 <0.001 <0.001 0.0153

337 For each model variant, containing one of LMI^A, LMI^B, LMI^C or LMI^D, fixed-effect coefficients are shown in yellow, while

the standard deviations of random effects are shown in green and the AIC of the model in blue.

339

340 In the case of each LMI prototype, the predictive power of age cohort shows a decisive effect. They are significant, 341 with estimates larger than LMI's estimates and a SE similar to LMI's SE values in each case. Age cohort reaches the highest 342 estimate and z-value in the case of LMI^A and LMI^C where, in turn, LMI estimates are lower compared to LMI^B and LMI^D. The 343 effects of sex are significant and show similar slope estimates across the prototypes. Relative to the slope estimates of age 344 cohort and LMI, SE values of sex are higher. Educational background does not have a significant effect for any LMI prototype 345 (although it is almost significant in LMI^B), and its SE values are high. Comparing the different predictors, LMI emerges as a 346 better predictor than sex and education in every case. The effect of age cohort dominates over the other variables, while the 347 effect of sex is larger than that of education.

A heuristic explanation of the importance of the random effects is provided by comparing their standard deviation (*SD*) to the estimate of a fixed effect. If the *SD* of the random effects is larger than the LMI prototypes' slope estimates, this means that the speaker and item effects are larger than the effect of LMI. Note that *SD*_{item} is practically the same for every model as the fixed effects are related to the speaker rather than the item.

Adding an interaction term between age cohorts and LMI means a small, but significant, increase for LMI^A (Table 5). This fact and the negative slope estimate indicate that the difference in LMI's effect is smaller than expected within the younger cohort compared to that in the older cohort. Not observing an interaction between these variables in the other models tells us that the effect of LMI does not depend on age cohort and therefore LMI is a predictor of similar worth in both age cohorts. Submodel ranking, using the *MuMIn* package [57] (version 1.43.17.), detailed in **S2_Appendix**, shows all fixed effects as significant contributors to the model quality.

358 Table 5 Addition of an interaction term between age cohort and the minimal LMI protot	type.
---	-------

	Model quelity	AIC	Random	$SD_{speaker}$	SD_{item}
	would quality	6125.35	effects	0.4710	0.6132
	Fixed effects	estimate	SE	z-value	р
LMI ^A	LMI	0.3754	0.0756	4.9662	<0.001
	Age cohort	0.7716	0.0766	10.0723	<0.001
	Sex	0.1998	0.0772	2.5886	0.0096
	Education	0.0921	0.0806	1.1432	0.2529
	LMI*Age cohort	-0.3805	0.1509	-2.5206	0.0117

359

360 Table 6 Spatial characteristics of the dialect change rate and the LMI prototypes, using the Kruskal-Wallis test and measuring the spatial autocorrelation using Moran's I

	Dependence of distribution (Kruskal-Wallis test)			Snatial a	autocorrelatio	n (Moran's	Δ
	χ^2	df	p	/	I _{expected}	SD	., р
Dialect change rate							
(10 variables)	266.82	124	< 0.001	-0.0223	-0.002	0.0012	< 0.001
Minimal (LMI ^A)	141.17	124	0.1388	-0.004	-0.002	0.0012	0.0946
Cumulative (LMI ^B)	155.14	124	0.0305	-0.005	-0.002	0.0012	0.0013
Comprehensive (LMI ^C)	152.06	124	0.0442	-0.0053	-0.002	0.0012	0.0068
Cohort-based (LMI ^D)	153.02	124	0.0394	-0.0055	-0.002	0.0012	0.0034

362

363

To contextualise linguistic mobility and dialect change with spatial processes, we provide our insights regarding the

364 spatial distribution of dialect change and LMI. Although a Kruskal-Wallis test (Table 6) shows significant differences across

our 125 localities with regard to the dialect change rate based on the ten lexical variables ($\chi^2 = 266.82$, df = 124, p < 0.001), these also stem from the different regional baselines regarding dialect change, i.e., there is also spatial variation in the lexical items in the SDS. Regarding LMI prototypes, we also find significant local differences using the Kruskal-Wallis test, except in the case of LMI^A.

369 Due to the aforementioned, potentially different, regional baselines of dialect change and the regional variation present 370 in Swiss German dialects, we will address the expected correspondence of spatial patterns in the dialect change trends and 371 linguistic mobility. In the polygons of the 125 SDATS survey sites, Figure 5A and 5C chart the dialect change rates, while 372 Figure 5B and 5D show the LMI^D values, for the two age cohorts. Urban areas are traditionally associated with more dialect 373 change [58]. Those SDATS localities that had more than 10,000 inhabitants in 2018, qualifying for the 'city' rank in 374 Switzerland, are shown with magenta edges. It is visible from the maps in Figure 5A and C, however, that urban areas, 375 especially in the cantons of Bern (BE) and Zurich (ZH), show the least dialect change in both age cohorts. Population of the 376 survey localities shows a significant negative linear correlation with language change ($R^2 = 0.032$, F(1,498) = 16.4824, p < 0.032) 377 0.001), meaning that the higher the population, the less language change can be expected. Dialect change in the non-urban 378 areas below 10,000 inhabitants ($\mu = 40.06\%$, SD = 19.12%) is almost significantly higher (t = 1.9069, df = 498, p = 0.057) than 379 in the urban areas ($\mu = 36.47\%$, SD = 20.78%). As the largest urban localities change remarkably less than rural areas do, and 380 urbanity is a fluid category, this definition of 'urban' seems to influence the result of the test. Also, contrary to naive 381 expectations, rural and mountainous areas, especially in the centre and south-east (e.g. cantons of Lucerne, Nidwalden, 382 Obwalden, Uri and Valais - LU, NW, OW, UR, VS, respectively), show the most dialect change in the younger cohort. Moran's 383 I analysis, however, shows significant negative spatial autocorrelation in dialect change rate (I = -0.0223, SD = 0.0012, p < 0.0012, 384 0.001). This means that there is a higher likelihood of finding different dialect change rates in nearby localities rather than 385 similar ones. In other words, dialect change does not cluster in space.

386 Geographic patterns of LMI^D also show a negative spatial autocorrelation (I = -0.0055, SD = 0.0012, p < 0.0034) with 387 little visual similarity across the age cohorts (Figure 5B and D) except in rural parts of BE and LU, and the generally suburban 388 region of ZH. Regarding the spatial autocorrelation of the other LMI prototypes, Moran's I values are also negative, i.e., mobile 389 or non-mobile speakers do not cluster in space. LMI, thus, cannot be safely interpreted as a function of geography in our sample. 390 Although trends of slight correspondence were seen in Figure 4, visual inspection of Figure 5 does not reveal spatial 391 correspondence between LMI^D and dialect change rates except for parts of BE and ZH being less mobile and showing less

- 392 change. A table containing all 125 SDATS localities with their average dialect change rates and LMI values can be found in
- 393 S2_Appendix.





400 **Discussion**

401 In this section we discuss LMI as a practical framework for researchers analysing language variation and change by 402 reflecting on its composition and evaluation in our study. Following this, we present the limitations of LMI and its 403 implementation, and make further suggestions about adapting the LMI framework for use in other studies.

404 LMI as a useful framework for language variation and change studies

With the growing (spatial) mobility of the modern world, any linguistic survey dealing with spatial variation must consider the mobility of its participants. To this end, many previous studies used an estimation for the 'localness' of a dialect by measuring the exposure of the participants to a reference locality, such as by applying the principles of the Regionality Index [1,15–17]. While RI shows the exposure to the reference dialect, LMI, in the form in which we implemented it, quantifies the exposure to variation outside the reference dialect. Without claiming its superiority in every scenario, if the availability of metadata allows, we suggest implementing LMI as a quantitative indicator of exposure for purposes similar to that of RI, for the following reasons.

412 Mobility has been shown to influence language change in several languages, including in theoretical work (e.g. 413 [3,9,10]) and in studies on different linguistic levels [1,2,5,6,8], less complex from the point of view of metadata. Thus, we 414 suggest that a more extensive quantitative explanatory factor aggregated from linguistic biography data would be successful 415 for the task of elucidating the effect of mobility on language change. Our contribution shows the wider applicability of LMI by 416 evaluating prototypes that simulate differences in metadata availability and theoretical considerations. In addition, our data 417 features different baselines of dialect change (due to the different age cohorts in the data). Moreover, the LMI framework, 418 consisting of setting up linguistic distances or judgements and weighting them by exposure and relational weights, is flexible 419 as these steps can be tailored to the needs of the researcher and according to the available data. A further advantage of LMI is 420 its cumulative nature, also enabling comparison across individuals when data is missing, which would otherwise affect 421 statistical (e.g. mixed-effects) modelling.

422 Discussion of the findings specific to the application to SDATS data

The evaluation was conducted through modelling dialect change in Swiss German using the LMI prototypes and the SDATS survey design variables (age, sex, educational background). Our evaluation tested whether the patterns of linguistic mobility characterising speakers that acquired their dialect under different circumstances (due to age, education, residency etc.) are suitable predictors of language change patterns that might not be the direct consequences of the speakers' dialect acquisition. This complex language change task makes the evaluation of the predictors difficult in terms of establishing causality, although the spatial diversity and covert prestige make Swiss German dialects an ideal environment for actual mobility-induced change to be present, as opposed to media- or prestige-induced change, for example. From the point of view of demonstrating the usefulness of LMI, the performance of the LMI prototypes beyond the fact that they are meaningful predictors is not of great importance.

432 In the Swiss German context, the results of the mixed-effects models can be interpreted in line with expectations: 433 speakers with higher linguistic mobility (LMI) show more dialect change (Table 4). This corresponds to previous observations 434 showing that being sedentary is a countereffect for dialect change in the Swiss German context (e.g. [19,39]). Each LMI 435 prototype proved to be a significant predictor when controlling for the survey design variables (Table 4). Beside the mixed-436 effects model, we also showed the role of age and sex on dialect change in bivariate tests (visually in Figure 3), and linear 437 regression models relating the LMI prototypes to language change (Figure 4) showed the cumulative LMI^B and cohort-based 438 LMI^D as better sole predictors over the others. Regarding LMI, differences across the distribution of its prototypes also reflect 439 the role of age (Figure 2A-D): no clear age-related pattern is present for LMI^A, but to a certain degree other prototypes show 440 higher mobility for the younger cohort. The fact that younger speakers show more dialect change, which we expect LMI to 441 predict, indicates that it is crucial to consider age when implementing LMI for SDATS.

442 Amongst the LMI prototypes, LMI^D holds the top position (based on AIC and z-values) as a predictor. We associate 443 this success with LMI^D addressing the differences between the metadata availability of the two age cohorts, a fact that makes 444 it more complex than LMI^C. Its top position is not so clear, however, which means that the inclusion of the design variables as 445 fixed effects regulates the explanatory power of the LMI prototypes. That is, the composition of the LMI prototypes affected 446 their explanatory power less when we controlled for the design variables in the model. As the cumulative prototype (LMI^B), 447 not involving exposure weights, performed similarly to LMI^D, we can conclude that in our study, it was not the fine-grained 448 details of the LMI implementation that were important but rather whether we successfully captured those mobility-related 449 factors that impact dialect change. In comparison to other predictors, LMI has a stronger effect on dialect change than sex and 450 educational background do (i.e. an increase of two standard deviations in the LMI values increases the odds of dialect change 451 more than binary switches in sex and educational background do), which is an additional argument in support of the usefulness 452 of LMI.

The high *SD*_{speaker} value compared to the slope estimates means that random effects capture more variance in dialect change than LMI, sex and educational background do. Therefore, between-speaker variation, not accounted for by the LMI prototypes, is still very important for dialect change, which means that parts of the speakers' linguistic biographies that could not be included in LMI (such as contacts not present in the metadata, personality, dialect attitudes etc.) make a difference and that local baselines of dialect change also involved in personal variation (random speaker effect) still contain crucial effects.

Corroborated by the strong age effect on dialect change shown by the bivariate statistics (Figure 3), age cohorts maintain the highest estimate in each model, again showing that the manner of including an age effect in LMI is crucial when explaining language change. The significance of the interaction of age and LMI^A (Table 5) means that the minimalistic LMI^A is a weaker predictor in terms of the age effect. The negative slope estimate of the interaction of age with LMI^A tells us that the difference in LMI's effect is smaller than expected within the younger cohort than in the older cohort.

Sex being a significant predictor in the case of each LMI prototype means that more change is expected in the dialect of men compared to that of women, when keeping age cohort, educational background and linguistic mobility constant. Slope estimates of sex are very similar across the four models, which means that their effect is not influenced by the composition of the LMI prototypes. Indeed, the LMI prototypes do not comprise any dialectal exposure in a manner directly differentiated by sex.

Educational background is not significant in any of the models, with a large *SE* diminishing its worth as a predictor. The explanation for the lack of importance of educational background may partly be due to the varied educational backgrounds being suboptimally categorised into two groups (also recall that the dataset includes many young speakers attending higher education who are not yet eligible for the 'with tertiary education' label), and the fact that social stratification does not extensively influence dialectal differences in Switzerland.

The spatial analysis, showing that LMI values and dialect change rate do not cluster or correspond to each other in space, reveals that the trends within and correspondence between LMI and dialect change are independent from their spatial variation. Low mobility may not necessarily be the cause of little dialect change. Similarly, change may not occur even in very mobile speakers. Because of the restricted linguistic sample and since SDATS speakers are more sedentary than average, we can only speculate on a cause. Perhaps the pattern is an artefact of the selected lexical variables or, due to the levelling of Swiss German variation [19,39], locally there is no longer a more prestigious variant for them to adopt.

479 Urban lifestyle is often associated with higher general mobility, but in our sample linguistic mobility does not 480 correspond to this preconception (Figure 5B and D, magenta-edged polygons). Contrary to popular opinion, the population of 481 SDATS survey localities shows a slightly negative correlation with dialect change rate across the ten items, which means that 482 there is less dialect change in localities with a larger population. Correspondingly, rural areas in Figure 5A and C, represented 483 by white polygon edges, show more change (white to redder hues) than cities (magenta edges), especially in the younger age 484 cohorts. Considering the ongoing levelling, the rural population seems to align with the local urban varieties [59], although 485 given that rural dialects often define their own identities by rejecting urban variants, this pattern might be an artefact of the 486 linguistic sample. Intuitively, however, the spatial pattern of dialect change corresponds to the increasing geographical 487 mobility, and thus is more difficult to unravel based on our restricted sample of linguistic items. In order to find true regional 488 and geographic effects in dialect change, an explicit dialectometric study would be necessary, investigating a larger set of 489 linguistic items.

490 **Limitations**

491 Beyond the default limitations and potential drawbacks associated with dialect surveys and their data elicitation 492 practices, such as (socio)linguistic interviews and (unsupervised) metadata collection, the following limitations may also hinder 493 other LMI implementations.

Owing to the limitations of the unsupervised metadata collection, inconsistent answers may always occur and some data may not be sufficient for inferences to be drawn on linguistic mobility. For example, geolocation problems arise due to the inconsistency or ambiguity of localities indicated by the speakers. The ongoing COVID-19 pandemic affected several SDATS questionnaire items directed at short-term mobility and social networks. Effects of short-term mobility were tested in two unpublished master theses in relation to dialect change [60,61], but no predictive effect was found. Additionally, questionnaire items regarding current linguistic connections do not necessarily characterise long-term contact.

500 Specific problems with metadata items which impact the composition of LMI include the availability of information. 501 Information is available about the speakers' parents, their current most significant peer (the partner), place of education (if 502 currently attending), and workplace. In addition, we can infer from different kinds of aggregate information on the reference 503 locality, which (indirectly) indicates childhood and adolescent peer effects. There is, however, no information about the 504 duration of contact between the speaker and the people elicited from their social networks, such as their flatmates and their 505 closest personal and work-related peers. Regarding residence outside the reference locality, information is available about 506 locations and the durations are recorded, but not the age at which the speaker lived there. Due to this, it was not possible to 507 evaluate LMI by testing its predictive performance against mobility predictors based on the number of years spent away from 508 the reference locality (cf. [6.8]). The discrepancy between the availability of information about the vounger and the older cohort 509 regarding their adolescence and other potential LMI components is also a limitation of our study. For instance, information is 510 missing on pensioners' last workplace, as is the last place of education of those not attending any education at the time of the 511 survey. Due to their age, the older cohort's baseline regarding dialect change is closer to the SDS, whereas for the younger 512 cohort there is a higher probability of dialect change being inherited. This makes the older cohort more optimal subjects for 513 assessing whether linguistic mobility impacts dialect change directly, as dialect change may have happened during their lives 514 that was actually caused by some of the factors included in LMI.

515 Given that older linguistic material was used as the basis for calculating linguistic distance, estimates of exposure may 516 be biased. In our study, the linguistic distance values based on 70-year-old data are somewhat conservative, as due to patterns 517 of dialect change [19,39], linguistic distance values on the Swiss Plateau (Ger.: *Schweizer Mittelland*) have probably decreased 518 more since the 1950s than they did in mountainous, more isolated areas.

519 **Recommendations for applying the LMI framework in other studies**

520 In this section, we will summarise a few issues to consider in order to facilitate the adoption of LMI in other studies, 521 organised according to the steps outlined for the implementation of the LMI framework. Following this, we provide a few 522 scenarios where we see the implementation of LMI is warranted.

523 Most importantly, a researcher must fine-tune the framework for their own goals. There is no single best LMI 524 implementation for any study, since any composition will have a certain degree of subjectivity and its predictive power will 525 depend on the speakers in the dataset, and on the linguistic scenario of the research. Prior to implementing LMI, crucial factors 526 that impact the outcome variable in the specific study should also be determined. Such factors may include, beside those tested 527 in the present study (age, sex, educational background, spatial variation), the role of the standard language, the strength of local 528 identity (e.g. [62]) and the power relationship between urban and rural dialects [59,63], boosting or countering the effects of 529 mobility. By the same token, an inclination or reluctance towards dialect change may also be associated with personality and 530 identity, for example, openness, extraversion or pride in one's dialect [41]. In terms of language change studies, age and the 531 time elapsed between points of comparison may be crucial as change accumulates over time, meaning that younger speakers 532 would typically show more dialect change. This is exemplified by the emergence of the age-cohort-based LMI^D as the best

predictor in our study. In order to determine the items of metadata to include in the composition of LMI, we advise testing metadata items as single predictors of the phenomenon investigated and also as aggregates, similarly to our relational weight model, to detect those that affect the outcome variable. Studies may also benefit from a similar spatial analysis to that presented here to determine whether the linguistic phenomenon investigated is governed by regional processes rather than the mobility of individuals.

538 Determining the linguistic distance between the speaker and the factor's locality is a possible way to estimate potential 539 linguistic effects despite the exact effects of the factor being unknown. If the calculation of linguistic distances is not possible, 540 other judgements could be made, including using proxies like geographical distance or travel time (e.g. [5,47]), since linguistic 541 variation is, to a great extent, spatially autocorrelated [64]. Regarding the mismatch between growing linguistic and geographic 542 distances, however, caution should be exercised.

Flat rates essentially address the general unfeasibility of quantifying the actual impact of people encountered. By utilising such average values, a certain dialect variety or place can be assigned to factors, and available data about the relationship to places (through specific people encountered) can be quantitatively used. Linguistic distances between pairs of localities are flat rates in our implementation, but other studies may choose other estimations of linguistic differences; however, any composition should be kept universal enough to be a valid, comparable measure for all speakers in a certain sample.

548 Exposure weights, also flat rates in our study, will be specific to available metadata, research questions, language and 549 culture, beside practical and theoretical considerations, while relational weights can be estimated through the researchers' own 550 assessment, beyond our models that are based on influential factors. Since significant effects are confirmed with our LMI 551 prototypes featuring different components, the global meaningfulness of LMI may not depend heavily on the minor details of 552 its construction but whether it successfully captures long-term exposure to dialectal variation from (often noisy) data. Using a 553 double-weighting system similar to our study is, however, not the only way to operationalise individuals' relations and exposure 554 to influencing factors. The determination of relational weights, beyond flat rates, is especially challenging given the vast 555 between-speaker variation regarding contact with the factors, which could be hardly quantified without very specific metadata 556 elicitation.

557 LMI can also be implemented with little metadata collected. Naturally, the implementation of LMI has better 558 possibilities when a larger abundance of factors is collected, such as information on peers, especially before and around 559 adolescence [21], or a comprehensive collection of metadata that allows researchers to assess the duration, intensity and 560 temporal sequence of linguistic effects on an individual. Nevertheless, the significant performance of LMIA, despite its 561 minimalistic composition, means that implementing LMI should be possible for most linguistic survey data, since at least some 562 information is usually available regarding survey design variables. LMIA's different effects in the two age cohorts, however, 563 raise caution for implementations using little metadata. Through the success of the cumulative strategy not using exposure 564 weights (LMI^B), an LMI adding up influential factors in the metadata may also be a possibility for other studies. In certain 565 cases, especially for studies with a smaller number of speakers, setting LMI or components of it manually may also be 566 reasonable, e.g. based on qualitative or anecdotal information available, but comparability should be maintained in 567 correspondence studies.

568 Beyond similar language change studies to the one presented in this paper, the suitability of LMI seems also to be 569 affirmed for lifespan studies (e.g. [65]), investigating real-time effects of mobility and exposure to variation, rather than 570 apparent-time ones in points recorded in different surveys. Similarly to studies finding that women [27], people with a higher 571 level of education [28], people displaying extraversion [49], or social centrality [34] show more proneness to language change, 572 LMI may also be useful for contributing to research on sociodemographic characteristics that make people more likely to adopt 573 innovations. Further studies, thus, may find that the linguistically mobile (i.e. those with a higher exposure to varied dialects 574 or languages, and in the case of most languages, exposure to the standard) are more likely to be linguistic innovators, while the 575 non-mobile are the laggards resistant to change.

576 Dialectological research mostly references their participants to a single survey locality which they represent. The LMI 577 framework could be a step towards the goal of addressing the phenomenon of mobility causing dialect change, achieved by 578 measuring separately the degree to which an individual is exposed to different localities and varieties, in a multifaceted 579 localisation. Beyond calculating one universal value, linguistic mobility could be implemented as a spatially directed predictor 580 based on the general mobility network of the population, in correspondence with Trudgill's theory of linguistic gravity [9], 581 rather than being focused on individuals. For example, based on (historic) commuting and relocation patterns, it could be 582 investigated whether the linguistic mobility patterns or language change patterns of an individual conform to the strongest 583 spatial streams. In addition, the term 'mobility' may represent the linguistic effects of exposure not only to other regions but 584 other, social factors (e.g. social classes, other social groups or networks) as well.

585

586 Acknowledgements

- 587 We are grateful to Yves Scherrer for allowing us to use the corpus of SDS data digitised by himself and his colleagues,
- 588 based on which linguistic distances, the foundation of the LMI values, were calculated. We would like to thank Karen Beaman
- 589 for her very helpful and supportive review and the editors for their work. We also thank Eugen Unterberger, Christian Zimmer

590 and others for the helpful comments on the previous versions of the research at conferences and gatherings in 2022, and Katie

- 591 Finnegan for her extremely helpful copy-editing work. Importantly, we would like to thank all colleagues and students at the
- 592 University of Bern for their help with the SDATS data collection and in the transcription of the data.
- 593

594 **References**

- 595 1. Chambers JK. Region and language variation. English World-Wide. 2000;21: 169–199.
- 5962.Hernández-Campoy JM. Exposure to contact and the geographical adoption of standard features: Two
complementary approaches. Lang Soc. 2003;32: 227–255. doi:10.1017/S0047404503322043
- Britain D. Sedentarism and nomadism in the sociolinguistics of dialect. In: Coupland N, editor. Sociolinguistics:
 Theoretical Debates. Cambridge: Cambridge University Press; 2016. pp. 217–241.
- 6004.Auer P. The Geography of Language: Steps toward a New Approach. Freiburg Arbeitspapiere zur Ger Lingustik.6012013; 1–39. Available: https://portal.uni-freiburg.de/sdd/fragl/2013.16
- 5. Beaman K V. Identity and mobility in linguistic change across the lifespan: the case of Swabian German. In: Ziegler
 A, Edler S, Kleczkowski N, Oberdorfer G, editors. Urban Matters Current Approaches of International
 Sociolinguistic Research (Studies in Language Variation). Amsterdam: John Benjamins; 2021. pp. 27–60.
 doi:https://doi.org/10.1075/silv.27.02bea
- 606 6. Bowie D. The Effect of Geographic Mobility on the Retention of a Local Dialect. University of Pennsylvania. 2000.
- 607 7. Britain D. Space, Diffusion and Mobility. 2nd ed. In: Chambers JK, Schilling N, editors. The Handbook of Language
 608 Variation and Change. 2nd ed. Chichester, West Sussex, England: Wiley; 2013. pp. 469–500.
 609 doi:10.1002/9781118335598.ch22
- 6108.Regan B. Dialectology meets sociophonetics: The social evaluation of ceceo and distinción in Lepe, Spain. In:611Chappell W, editor. Recent Advances in the Study of Spanish Sociophonetic Perception. John Benjamins; 2019. pp.61285–121.
- 6139.Trudgill P. Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. Lang614Soc. 1974;2: 215–246.
- 61510.Milroy L. Introduction: Mobility, contact and language change Working with contemporary speech communities. J616Socioling. 2002;6: 3–15. doi:10.1111/1467-9481.00174
- 617
 11.
 Vertovec S. Super-diversity and its implications. Ethn Racial Stud. 2007;30: 1024–1054.
 618
 doi:10.1080/01419870701599465
- Dodsworth R. Migration and Dialect Contact. Annu Rev Linguist. 2017;3: 331–346. doi:10.1146/annurev-linguistics-011516-034108
- 621 13. Bailey C-JN. Variation and linguistic theory. Washington, DC: Center for Applied Linguistics; 1973.
- 622 14. Chambers JK, Trudgill P. Dialectology. 2nd ed. Cambridge: Cambridge University Press; 1998.
- Chartier N. Old Stereotypes "Live Free or Die": Addressing the Evaluation Problem of Non-rhoticity in New
 Hampshire. University of Washington. 2020. Available: https://www.proquest.com/docview/2397710959?pq origsite=gscholar&fromopenview=true
- 626 16. Jones BG. The Spatiality of Perceptual Dialectology. University of Washington. 2021.
- Moroz AK. Exploring Applications of Rootedness in Sociolinguistic Research in Southern Oregon. University of
 Washington. 2021.
- Inderwood GN. American English dialectology: Alternatives for the Southwest. Linguistics. 1974;12: 19–40.
 doi:10.1515/ling.1974.12.128.19
- 19. Juska-Bacher B. Wortgeografischer Wandel im Schweizerdeutschen. Sommersprossen, Küchenzwiebel und

- 632 Schmetterling 70 Jahre nach dem SDS. Linguist online. 2010;42.
- 633
63420.Takemura A. Dialect acquisition in the view of parental origins: The case of Kansai dialect of Japanese. Kobe Pap
Linguist. 2010;7: 78–90.
- Baxter GJ, Croft W. Modeling language change across the lifespan: Individual trajectories in community change.
 Lang Var Change. 2016;28: 129–173. doi:10.1017/S0954394516000077
- 637 22. Labov W. Sociolinguistic patterns. Philadelphia: University of Pennsylvania Press; 1991.
- Cheshire J, Milroy J. Syntactic variation in non-standard dialects: background issues. In: Milroy J, Milroy L, editors.
 Real English: The Grammar of English Dialects in the British Isles. London: Routledge; 1993. pp. 3–33.
- Stanford JN, Pan Y. The sociolinguistics of exogamy: Dialect acquisition in a Zhuang village. J Socioling. 2013;17:
 573–607. doi:10.1111/josl.12052
- Lansing JS, Abundo C, Jacobs GS, Guillot EG, Thurner S, Downey SS, et al. Kinship structures create persistent channels for language transmission. Proc Natl Acad Sci. 2017;114: 12910–12915. doi:10.1073/pnas.1706416114
- 644 26. Persyn D. On dialects, networks, and labour mobility. Leuven; 2017. Report No.: 58.
- Leemann A, Jeszenszky P, Steiner C, Studerus M, Messerli J. SDATS Corpus Swiss German dialects across time and space. Bern; 2020. Available: osf.io/s9z4q
- Labov W. The social stratification of English in New York city. Washington, D.C.: Center for Applied Linguistics;
 1966.
- Tamminga M. Leaders of language change: Macro and micro perspectives. In: van der Velde H, Haug Hilton N,
 Knooihuizen R, editors. Language Variation European Perspectives VIII: Selected papers from the Tenth
 International Conference on Language Variation in Europe (ICLaVE 10), Leeuwarden, June 2019. Leuwarden: John
 Benjamins; 2021. doi:10.1075/silv.25.12tam
- 65330.Labov W, Rosenfelder I, Fruehwald J. One hundred years of sound change in Philadelphia: Linear incrementation,654reversal and reanalysis. Language (Baltim). 2013;89: 30–65. Available: https://www.jstor.org/stable/23357721
- 655 31. Milroy J, Milroy L. Linguistic Change, Social Network and Speaker Innovation. J Linguist. 1985;21: 339–384.
- Eckert P. Jocks and burnouts: Social categories and identity in the high school. New York: Teachers College,
 Columbia University; 1989.
- Lippi-Green RL. Social network integration and language change in progress in a rural alpine village. Lang Soc.
 1989;18: 213–234. doi:10.1017/S0047404500013476
- 66034.Fagyal Z, Swarup S, Escobar AM, Gasser L, Lakkaraju K. Centers and peripheries: Network roles in language
change. Lingua. 2010;120: 2061–2079. doi:10.1016/j.lingua.2010.02.001
- Blythe RA, Croft W. How individuals change language. PLoS One. 2021;16: 1–23.
 doi:10.1371/journal.pone.0252582
- 66436.Hotzenköcherle R, Schläpfer R, Trüb R, Zinsli P, editors. Sprachatlas der deutschen Schweiz (1962-2003). 8 vols.665Bern (I–VI)/ Basel(VII–VIII): Francke;
- Berthele R. Vor lauter Linguisten die Sprache nicht mehr sehen Diglossie und Ideologie in der deutschsprachigen
 Schweiz. In: Christen H, editor. Dialekt, Regiolekt und Standardsprache im sozialen und zeitlichen Raum Beiträge
 zum 1 Kongress der Internationalen Gesellschaft für Dialektologie des Deutschen. Wien: Edition Praesens; 2004. pp.
 111–136.
- 67038.Studler R. Cognitive cultural models at work: The case of German-speaking Switzerland. Yearb Ger Cogn Linguist671Assoc. 2017;5: 93–106. doi:10.1515/gcla-2017-0007
- 672 39. Christen H. Convergence and divergence in the Swiss German dialects. Folia Linguist. 1998;32: 53–68.
 673 doi:10.1515/flin.1998.32.1-2.53
- 67440.Jeszenszky P, Steiner C, Leemann A. Reduction of Survey Sites in Dialectology: A New Methodology Based on
Clustering. Front Artif Intell. 2021;4: 1–23. doi:10.3389/frai.2021.642505
- 41. Steiner C, Jeszenszky P, Leemann A. Variation and Change in Swiss German Agreement Morphology: Spatial,
 Social, and Attitudinal Effects. J Linguist Geogr. 2022;11: 1–42. doi:10.1017/jlg.2022.12
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2021. Available: https://www.r-project.org/
- 680 43. Cambon J, Hernangómez D, Belanger C, Possenriede D. tidygeocoder: An R package for geocoding. J Open Source
 681 Softw. 2021;6: 3544. doi:10.21105/joss.03544
- 44. Pebesma E. Simple features for R: Standardized support for spatial vector data. R J. 2018;10: 439–446.
 doi:10.32614/rj-2018-009
- 684 45. Goebl H. "Stammbaum" und "Welle." Zeitschrift für Sprachwiss. 1983;2: 3–44.
- 685
 686
 686
 687
 46. Scherrer Y. dialektkarten.ch Interactive dialect maps for German-speaking Switzerland and other European dialect areas. In: Krefeld T, Lücke S, Mutter C, editors. Berichte aus der digitalen Geolinguistik (II): Akten der zweiten Arbeitstagung des DFG-Langfristvorhabens VerbaAlpina und seiner Kooperationspartner am 18062019. Munich:

- 688 Korpus im Text, University of Munich; 2021.
- 47. Jeszenszky P, Hikosaka Y, Imamura S, Yano K. Japanese Lexical Variation Explained by Spatial Contact Patterns.
 690 ISPRS Int J Geo-Information. 2019;8: 1–30. doi:10.3390/ijgi8090400
- 48. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. J Stat Softw.
 2011;45: 1–67. doi:10.18637/jss.v045.i03
- 49. Labov W, Ash S, Boberg C. The Atlas of North American English. 2006. doi:10.1515/9783110206838
- 69450.Steiner C, Jeszenszky P, Stebler V, Leemann A. Extraverted innovators and conscientious laggards? Investigating
personality effects on language change. Lang Var Change.
- 69651.Trudgill P. Sociolinguistic typology and the speed of linguistic change. J Hist Socioling. 2020;6. doi:10.1515/jhsl-6972019-0015
- 698
69952.Michel JB, Kui Shen Y, Presser Aiden A, Veres A, Gray MK, Pickett JP, et al. Quantitative analysis of culture using
millions of digitized books. Science (80-). 2011;331: 176–182. doi:10.1126/science.1199644
- 53. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67.
 doi:10.18637/jss.v067.i01
- 54. Gelman A, Hill J. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press; 2006.
- 70455.Bybee JL. Diachronic Linguistics. In: Geeraerts D, Cuyckens H, editors. The Oxford Handbook of Cognitive705Linguistics. Oxford: Oxford University Press; 2012. pp. 1–55. doi:10.1093/oxfordhb/9780199738632.013.0036
- 706 56. Fox J, Weisberg S. An {R} Companion to Applied Regression. Third. Thousand Oaks, California: SAGE; 2019.
- 707 57. Barton K. MuMIn: Multi-Model Inference. R package version 1.43.17. R package. CRAN; 2020.
- 58. Britain D. "Rural" and "Urban" in Dialectology. In: Busse B, Warnke IH, editors. Handbuch Sprache im urbanen
 Raum Handbook of Language in Urban Space. Berlin, Boston: De Gruyter; 2022. pp. 52–73.
 doi:https://doi.org/10.1515/9783110296334-003
- Taeldeman J. The influence of urban centres on the spatial diffusion of dialect phenomena. In: Auer P, Hinkens F,
 Kerswill P, editors. Dialect Change: Convergence and Divergence in European Languages. Cambridge: Cambridge University Press; 2005. pp. 263–284. doi:10.1017/CBO9780511486623.012
- 60. Grunder LP. Das heisst 'der Hung fouget' u niid 'der Hund folget': Eine Studie zum Einfluss des Mittelbernischen auf das Oberländische. Universität Bern. 2021.
- 61. Wagner JL. "Hesch mir es Taschetuech us de Schublade? Ich ha mir bim Schlittschue Laufe de Schnupfe gholt.
 "Eine Studie zum Sprachwandel und Einfluss des Standarddeutschen auf die Lexik der Dialekte in der Deutschschweiz. University of Bern. 2021.
- Beaman K V., Tomaschek F. Loss of Historical Phonetic Contrast Across the Lifespan: Articulatory, Lexical, and
 Social Effects on Sound Change in Swabian. In: Beaman K V., Buchstaller I, editors. Language Variation and
 Language Change Across the Lifespan. New York: Routledge; 2021. pp. 209–234.
- 63. Siebenhaar B. Sprachvariation, Sprachwandel und Einstellung : der Dialekt der Stadt Aarau in der Labilitätszone
 723 zwischen Zürcher und Berner Mundart. University of Zurich. 2000. Available:
 724 https://www.researchgate.net/publication/316673055 Sprachvariation Sprachwandel und Einstellung Der Dialekt
- https://www.researchgate.net/publication/316673055_Sprachvariation_Sprachwandel_und_Einstellung_Der_Dialekt
 der_Stadt_Aarau_in_der_Labilitatszone_zwischen_Zurcher_und_Berner_Mundartraum
- 72664.Nerbonne J, Kleiweg P. Toward a dialectological yardstick. J Quant Linguist. 2007;14: 148–166.727doi:10.1080/09296170701379260
- Baayen RH, Beaman K V., Ramscar M. Deconfounding the Effects of Competition and Attrition on Dialect across the Lifespan. Language Variation and Language Change Across the Lifespan. 2021. pp. 235–264.
 doi:10.4324/9780429030314-10
- 731

732 Supporting information

733

S1 Appendix LMI setup - R Markdown report in html format. The report describes with code the construction of

The The Table 734 LMI components from the SDATS survey metadata.

- 735 S2 Appendix Mixed models R Markdown report in html format. The report presents the composition of the four
- 736 LMI prototypes evaluated in this article, conducting the statistical tests and modelling reported here including their summary
- results. The reproducible code for both reports can be accessed at <u>https://osf.io/hfbpk/</u>