



Daily dialect-speaking and wages among native Dutch speakers

Yuxin Yao¹ · Jan C. van Ours^{2,3,4,5,6}

© The Author(s) 2018

Abstract

Our paper studies the effects of daily dialect-speaking on hourly wages of native Dutch workers. The unconditional difference in median hourly wage between Standard Dutch speakers and dialect speakers is about 10% for males and 8% for females. Taking into account differences in personal characteristics, family characteristics and geographical differences, wage differences are reduced with about 6%-points.

Keywords Dialect-speaking · Wage penalty · Job characteristics

JEL Classification J24 · I2

1 Introduction

Language skills are an important determinant of labor market performance. Previous studies have focused on the effect of language proficiency on earnings of male immigrants. Recent examples are Bleakley and Chin (2004), Miranda and Zhu (2013a, b), Di Paolo and Raymond (2012) and Yao and van Ours (2015). However, it is not only language proficiency that affects labor market performance. Also, language speech patterns may be important, i.e. it may matter whether a worker speaks a standard language or a dialect. Though among linguists there is no common definition of

✉ Jan C. van Ours
vanours@ese.eur.nl

Yuxin Yao
yyx.eco@gmail.com

¹ Department of Economics, East China Normal University, Shanghai, China

² Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands

³ Tinbergen Institute, Amsterdam/Rotterdam, The Netherlands

⁴ Department of Economics, University of Melbourne, Parkville, Australia

⁵ CEPR, London, UK

⁶ IZA, Bonn, Germany

dialects, a dialect is usually referred to a variation of a language used by a particular group. A dialect may associate with social class. As for example is apparent from the “My Fair Lady” lyrics of the song “Why can’t the English?”: “Look at her, a prisoner of the gutter, condemned by every syllable she utters (...). An Englishman’s way of speaking absolutely classifies him. The moment he talks he makes some other Englishman despise him.”

A few studies investigated how speech patterns affect labor market outcomes. Gao and Smyth (2011) find a significant wage premium associated with fluency in standard Mandarin for dialect-speaking migrating workers in China. Carlson and McHenry (2006) presents the results of a small experiment on how speaking dialect affects employment probability. Bendick Jr. et al. (2010) using an experimental set-up studies the effects of a (mostly) French accent for white job applicants to New York City restaurants. These accents were considered as “charming” and they increased the probability of being hired as a waiter or waitress. To study the effects of speech patterns, Grogger (2011) uses NLSY data in combination with audio-information about how individuals speak. In the US labor market, black workers with a distinct black speech pattern earn less than white workers whereas black workers who do not sound distinct black earn the same as white workers. Grogger (2018) finds that lower wages of black workers are related to their speech pattern. These speech-related wage differences might be related to occupational sorting such that black workers whose speech is similar to white workers sort themselves into occupations that involve intensive interpersonal interactions. An alternative explanation mentioned by Grogger (2018) is employer discrimination, i.e. employers are prejudicial tasted against certain dialects. Grogger (2018) also finds that lower wages of US Southerners are not related to their speech pattern but are largely explained by family background and residential location. According to Das (2013) language and accents provide information about an individual’s social status. The spoken language may be a source of discrimination affecting earnings and promotion. In other words, a speech pattern may be a signal of unobserved productivity. It could also be that non-standard speech patterns reduce productivity at the workplace, for example because differences in language speech pattern between workers increase production costs (Lang 1986).

Language skills and speech patterns are acquired at early ages up to the teenage years. After this period it is hard to pick-up a different related speech. Rickford et al. (2015) present an analysis of audio recordings of participants in the Moving to Opportunity (MTO) project. In the MTO project, participants were randomly assigned to receive housing vouchers providing an opportunity to move to lower-poverty areas. Exploiting this randomization, Rickford et al. (2015) find that moving to lower-poverty neighborhoods at a young age changes speech patterns while moving at a higher age does not. Also using MTO data, Chetty et al. (2016) find that moving when young (before age 13) to lower-poverty neighborhoods is beneficial in terms of college attendance and earnings. Grogger (2018) suggests that the combination of the two findings is consistent with the notion that speech patterns affect earnings.

A dialect is a variation of the standard language, used in limited regions and different in mainly pronunciation, and sometimes vocabulary and grammar. Dialects

can be acquired without training and play a role in informal communication, while the standard language is the instruction medium at schools. Speaking with a local dialect accent may signal lower language ability, limited education and lack of experience communicating with people from other regions. Moreover, a similar speech pattern can signal cultural affinity. People are more likely to trust those who speak the same dialect and conduct trade (Falck et al. 2012). All this implies that speaking the major languages results in an advantage in economic activities. Although in dialect-speaking areas dialect can be viewed as a separate skill, the return to dialects is somewhat limited in other areas in the country. Therefore, it is of interest to explore how dialect speech patterns affect labor market performance and whether it is premium or penalty in the labor market.

Our paper studies the relationship between dialect-speaking and hourly wages. We study the Netherlands as an example of a country with a lot of commuting such that the spatial segregation is limited. This is not only because there are various dialects spoken, but also Dutch natives are more homogeneous in terms of culture, physical characteristics and economic wealth than natives from larger countries. Moreover, to compare native dialect speakers with Standard Dutch speakers, we obtain purer effects of speech pattern than comparing immigrants with natives. We focus on native Dutch individuals who indicated not having problems with reading or speaking Dutch. We perform simple regressions starting with daily-dialect speaking as the only right-hand side variable and gradually introducing personal characteristics, family characteristics and geographical differences. The difference in wages between dialect-speakers and non dialect-speakers becomes smaller as more explanatory variables are introduced. We find that the unconditional difference in hourly wages between Standard Dutch and dialect speakers is about 10% for males and 8% for females. If we take into account personal characteristics and province fixed effects male dialect speakers earn 4% less while there is no significant penalty on female dialect speakers. Our findings suggest a significant wage penalty of daily dialect-speaking behavior on wages of males. Moreover, in provinces with a higher share of dialect-speakers, the wage penalty is less severe. Nevertheless, although our results are robust to various sensitivity tests our analysis does not go beyond conditional correlations.¹

Our paper is set-up as follows. In Sect. 2 we provide a description of data and linguistic background of Dutch. Section 3 presents our parameter estimates. Section 4 discusses possible explanations for our main findings. Section 5 concludes.

¹ In Yao and van Ours (2016), we explore whether there is a potential causal relationship from daily dialect-speaking to wages. For this, we use an instrumental variable approach and propensity score matching. The results are very similar to the pooled OLS-estimates we present below.

2 Data and background

2.1 Linguistic background

The formal language of the Netherlands is Standard Dutch. It is spoken in its purest form in Haarlem, a city close to the capital Amsterdam. Standard Dutch is spoken throughout the Netherlands but there are also many regional languages and dialects. Frisian, mostly spoken in the province of Friesland, is recognized as a separate language and promoted by the local government. In Friesland, both Standard Dutch and Frisian are considered official languages and instruction media at school. More than 80% of the adult inhabitants understand verbal Frisian, but only a small minority can write the language (Gorter 2005). In our paper for simplicity we refer to Frisian as a dialect. Other official regional languages include Limburgish, spoken in the province of Limburg by about 75% of the inhabitants. Low Saxon dialects are spoken in the provinces of Groningen, Drenthe, Overijssel and Gelderland by approximately 60% of the inhabitants. Other provinces have their own dialects such as Brabantish, spoken in Noord-Brabant or Zeelandic in Zeeland [see an overview in Driessen (2005) and Cheshire et al. (1989)].

Distances between languages depend on characteristics such as vocabulary, pronunciation, syntax and grammar. To quantify distances between languages various methods are used. Levenshtein (1966) proposed an algorithm based on the minimum number of steps to change a particular word from one language to the same word in a different language. The overall distance between two languages is based on the average difference for a list of words for which often but not always the 100 words from Swadesh (1952) are used. Levenshtein's method can be based on written words but can also be based on phonetic similarities. This is especially helpful when comparing dialects as often these are spoken but not used in writing. Van Bezooijen and Heeringa (2006) use two samples of Dutch dialects and apply the Levenshtein distance measure to calculate the average linguistic distances between provincial dialects and Standard Dutch. In our paper, we use their distances, which are based on the New Dialect Sample. These distance measures are calculated from 100 words. The first column of Table 1 provides information about the linguistic distance of dialects to Standard Dutch averaged at the level of provinces. As shown, the linguistic distance to Standard Dutch of the dialect spoken in a particular province is the largest in Friesland and the smallest in Flevoland, Noord-Holland and Zuid-Holland.

2.2 Data

Our dataset is from the Longitudinal Internet Studies for the Social sciences (LISS) panel survey which is representative of the Dutch speaking population permanently residing in the Netherlands. In the LISS panel, background demographic variables are collected monthly while on specific topics data are collected annually (see for details: www.lissdata.nl). We use seven waves of panel data from 2008 to 2014. As is usual in studies on language effects, we rely on self-reported information about

Table 1 Dialect-speaking by province. *Source* linguistic distance: Van Bezooijen and Heeringa (2006)

Province	Linguistic distance	Geographical distance to Haarlem	Percentage of the sampled individuals speaking a dialect				
			Daily	Regularly	Sometimes	Never	Total
Drenthe	19	144	34	12	17	37	100
Flevoland	12	57	1	5	11	83	100
Friesland	37	116	48	9	13	30	100
Gelderland	28	97	14	9	20	57	100
Groningen	28	163	22	12	20	46	100
Limburg	32	170	68	7	10	15	100
Noord-Brabant	28	102	22	15	27	36	100
Noord-Holland	12	25	3	2	8	87	100
Overijssel	29	128	25	15	28	32	100
Utrecht	18	51	3	4	11	82	100
Zeeland	29	121	29	16	26	29	100
Zuid-Holland	12	47	3	2	10	85	100
Average	23	88	17	8	16	57	100

Distance to Haarlem in kilometers. Averaged over the individuals in our sample

speech patterns. Respondents indicate their intensity of dialect-speaking by answering the questions *Do you ever speak dialect?* with possible answers *Yes—daily*, *Yes—regularly*, *Yes—once a while* and *No—never*. The indicator for dialect-speaking we use in our analysis is whether or not an individual speaks a dialect daily.² To focus on the effects of dialect speaking pattern and exclude the effect of language deficiency, we remove from the sample immigrants and native individuals who indicate having problems in reading or speaking Dutch. We merged the dataset with a variable measuring the geographic distance from residential municipality to Haarlem.

The second column of Table 1 shows that the geographical distance to Haarlem of the individuals in our sample is on average the largest in Limburg and Groningen, and the smallest in Noord-Holland where Haarlem is located. Dialect-speaking is the most prevalent in Limburg where 68% of the individuals in our sample speak dialect daily, followed by Friesland with 48% and Drenthe with 34%. Flevoland, Noord-Holland, Utrecht and Zuid-Holland only have a tiny proportion of the sample speaking dialect daily. In addition, the so-called “Randstad” provinces have the highest share of individuals who never speak a dialect.³

² Since Frisian is an official language rather than a dialect, we refer to the survey question “*which language do you generally speak at home?*” for respondents from Friesland.

³ The Randstad is a megalopolis in the central-western Netherlands consisting primarily of the four largest Dutch cities (Amsterdam, Rotterdam, The Hague and Utrecht) and their surrounding areas. In our paper, the “Randstad” provinces are defined as Noord-Holland, Flevoland, Utrecht, and Zuid-Holland.

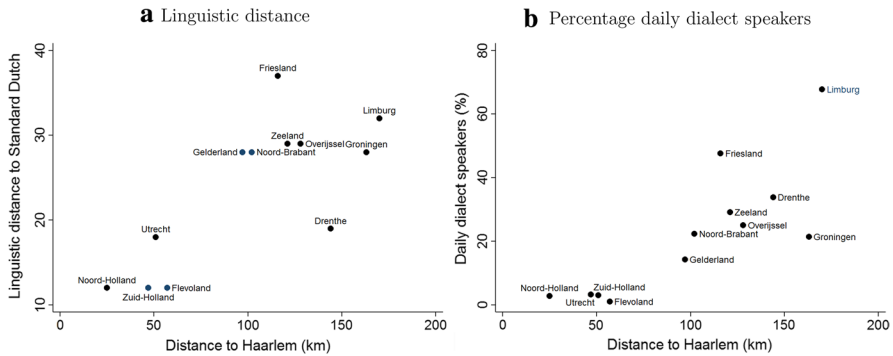


Fig. 1 Linguistic distance, percentage daily dialect speakers and geographical distance to Haarlem by province. *Source:* see Table 1

Figure 1 provides a graphical representation on the relationship between dialect characteristics and the geographical distance to Haarlem at the provincial level. Figure 1a shows that linguistic distance and geographical distance are highly correlated with Friesland and Drenthe as outliers. In Drenthe, the linguistic distance to Standard Dutch is smaller than in other provinces with the same geographical distance to Haarlem while in Friesland the linguistic distance to Standard Dutch is larger than it is in comparable provinces. Figure 1b shows that there is also a strong correlation between the share of (daily) dialect speakers and the distance to Haarlem.

In 2008, the minimum hourly wage in the Netherlands was about 7.8 Euro while in 2014 it was about 8.5 Euro. In the original data, some of the hourly wages are far below the minimum wage. To avoid a bias in the parameter estimates we removed from the sample all observations with an hourly wage below 7.5 Euro (2.6% of the observations). Our empirical analysis is based on 5721 observations of 1761 males and 5597 observations of 1756 females.⁴ The densities and cumulative distributions of hourly wages by dialect status are presented in Figure 2. The differences between males and females are not large but individuals who speak dialect daily on average have lower hourly wages.⁵ The median wage of Standard Dutch speaking males is 18.75 Euro while for Standard Dutch speaking females this is 15.86 Euro. Among dialect-speakers the median hourly wage is 16.76 Euro for males and 14.79 Euro for females. So, on average for males dialect speakers earn 10.6% less than non-dialect speakers. For females the difference is 6.7%. Of course, these differences need not be related to dialect-speaking itself but may be explained by other personal characteristics that correlate with dialect-speaking.

⁴ We have an unbalanced panel but find no indication for selective attrition, i.e. conditional on all control variables, drop out from the panel is uncorrelated to labor market outcomes.

⁵ According to the nonparametric Kolmogorov-Smirnov test, the distribution of hourly wages of Dutch speakers and dialect speakers are statistically different from each other. This holds for both males and females.

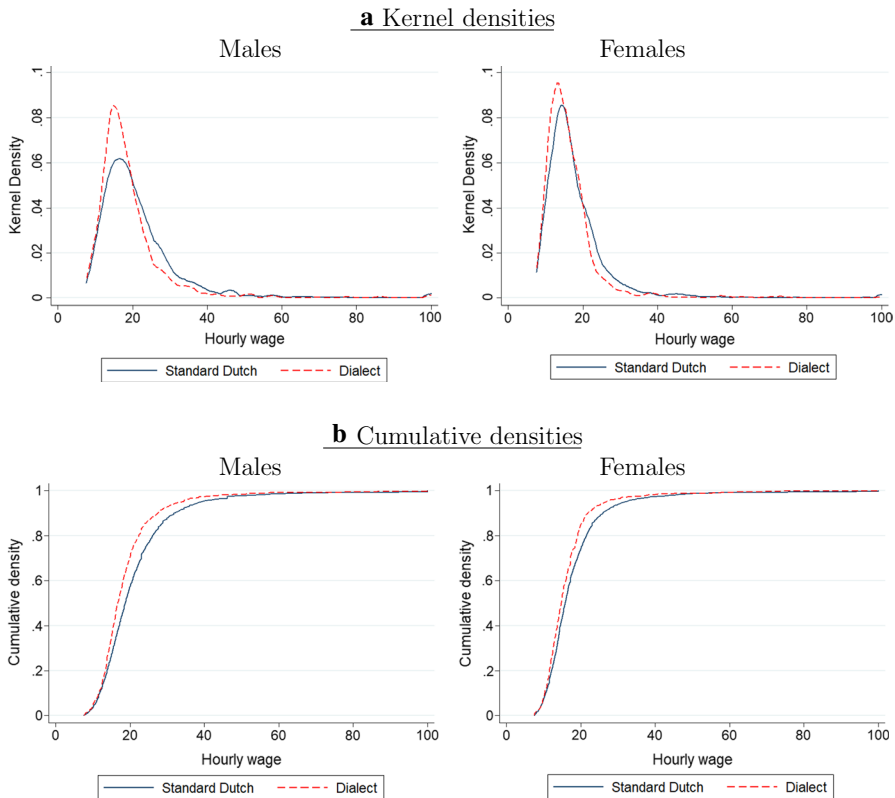


Fig. 2 Kernel densities and cumulative distribution of hourly wages; by gender and daily dialect-speaking

3 Wage effects of dialect-speaking

3.1 Baseline results

Although we have panel data, from 2009 to 2014, only 5–6% of individuals changed dialect-speaking status. So, dialect-speaking behavior hardly changes over time. Furthermore, since speech patterns are unlikely to change quickly a change in self-reported dialect-speaking status probably does not reflect a big change in speech pattern. Therefore, it is not possible to use a individual fixed-effects approach to account for unobserved differences between individuals. Instead, we use a pooled cross-section set-up relating daily dialect-speaking to log hourly wages:

$$Y_{it} = \alpha_t + \gamma D_{it} + \beta_1 X_{it} + u_{it} \quad (1)$$

where Y_i represents log hourly wages, the α_t indicate calendar year fixed effects, D_{it} is a dummy variable for daily dialect-speaking, X_i is a vector of individual, family and province characteristics, β_1 represents a vector of parameters and u_{it} is an error

Table 2 Baseline parameter estimates effect of daily dialect-speaking log hourly wages; pooled cross-section

	(1)	(2)	(3)	(4)	(5)
Males	− 0.100*** (0.022)	− 0.061*** (0.019)	− 0.054*** (0.019)	− 0.037* (0.019)	− 0.041** (0.020)
R ²	0.015	0.289	0.292	0.300	0.301
Females	− 0.080*** (0.021)	− 0.041** (0.019)	− 0.032 (0.020)	− 0.024 (0.021)	− 0.028 (0.023)
R ²	0.010	0.172	0.176	0.177	0.180
Individual characteristics	No	Yes	Yes	Yes	Yes
Family characteristics	No	No	Yes	Yes	Yes
Province characteristics	No	No	No	Yes	No
Province fixed effects	No	No	No	No	Yes

Based on 5721 observations of 1761 males and 5597 observations of 1756 females; all estimates contain year fixed effects. In parentheses, standard errors are clustered at the level of the individual

term. The parameter of main interest is γ . The personal characteristics are age, educational attainment, having a religion and having a partner. Family characteristics are number of children and urbanization level of the residence and province characteristics are log of per capita GDP, log of population, area of main roads (km²) and log of employment. The appendix provides details of all variables used in the analysis.

Table 2 presents OLS parameter estimates where the pooled cross-section nature of the data is taken into account by clustering the standard errors at the level of the individual. In the first column only calendar year fixed effects are taken into account. In the second column individual characteristics are included. The third column further adds family characteristics. Column (4) further includes province characteristics, while column (5) substitutes province characteristics for province fixed effects.

The first column of Table 2 shows that on average daily dialect speakers have lower hourly wages by 8% (females) to 10% (males). The second and third column show that this wage penalty is reduced once individual and family characteristics are introduced. The R² of the regression increases a lot once individual characteristics are introduced. Individual characteristics absorb about 4%-point of the wage penalty, family characteristics add another 1%-point. For females the wage penalty becomes insignificant. If province characteristics or province fixed effects are also included, the wage penalty further reduces to about 4% for males and about 3% for females, where for females the wage penalty is no longer different from zero at conventional levels of significance.⁶

Table 3 reports the OLS parameter estimates of the fourth column of Table 2 in detail. For both males and females, age is positively related to hourly wages.

⁶ We also investigated whether adding the linguistic difference of a provincial dialect changed our results finding that this is not the case.

Table 3 Daily dialect-speaking and log hourly wages; pooled OLS estimates

	Males	Females
Dialect daily	– 0.037 (0.019)*	– 0.024 (0.021)
Age 25–34	0.162 (0.054)***	0.162 (0.032)***
Age 35–44	0.379 (0.053)***	0.290 (0.033)***
Age 45–54	0.488 (0.054)***	0.312 (0.032)***
Age 55+	0.606 (0.055)***	0.393 (0.036)***
Lower secondary educ.	0.031 (0.047)	– 0.175 (0.072)**
Intermediate sec educ.	0.126 (0.045)***	– 0.078 (0.070)
Higher education	0.388 (0.046)***	0.151 (0.071)**
Number of children	0.002 (0.008)	– 0.003 (0.008)
Partner	0.054 (0.021)**	– 0.059 (0.020)***
Very urban	– 0.013 (0.028)	– 0.046 (0.029)
Moderately urban	0.042 (0.031)	– 0.047 (0.032)
Slightly urban	0.041 (0.033)	– 0.063 (0.033)**
Not urban	– 0.006 (0.034)	– 0.044 (0.034)
Religious	– 0.007 (0.018)	0.028 (0.018)
Log GDP per capita	– 0.042 (0.073)	0.030 (0.089)
Log Employment	– 0.091 (0.194)	0.021 (0.201)
Log Population	0.191 (0.199)	– 0.003 (0.201)
Area Main Roads/100	– 0.001 (0.000)	– 0.000 (0.000)

Based on 5721 observations of 1761 males and 5597 observations of 1756 females; all estimates contain calendar year fixed effects. In parentheses, standard errors clustered at the level of the individual

Workers age 55-plus earn substantially more than workers aged 45–54, who in their turn earn more than workers ages 35–44, et cetera. For males the age effect is stronger than for females. Educational attainment also has a positive effect on hourly wages, although again this effect is stronger for males than for females. The main difference between males and females is the effect of the partner. While males with a partner earn more than males without a partner, for females this is the other way around. Of course, this is not necessarily a causal effect, i.e. it could even be reverse causality with women with higher earnings being less likely to have a partner or men with higher earnings being more likely to have a partner. Religion, urbanization level of the municipality of residence and provincial characteristics do not have significant effects on hourly wages.

3.2 Additional parameter estimates

To investigate the sensitivity of our baseline estimates we performed a range of sensitivity analysis of which the results are presented in panels b to e in Table 4. For reasons of comparison, panel a reports the main baseline estimate of Table 2, column (4).

Table 4 Sensitivity analysis parameter estimates effects of dialect-speaking on log hourly wages

	Males	Females
a. Baseline estimates		
Daily dialect-speaking	– 0.037 (0.019)*	– 0.024 (0.021)
b. Province heterogeneity		
1. Randstad provinces	– 0.180 (0.058)***	– 0.020 (0.069)
2. Non-Randstad provinces	– 0.038 (0.021)*	– 0.032 (0.024)
3. Friesland	0.060 (0.062)	– 0.030 (0.072)
4. Other provinces	– 0.054 (0.020)***	– 0.024 (0.024)
c. Less intense dialect-speaking		
Ever dialect-speaking	– 0.032 (0.018)*	– 0.019 (0.019)
d. Intensity of dialect-speaking		
Daily dialect-speaking	– 0.053 (0.024)**	– 0.035 (0.025)
Regular dialect-speaking	– 0.033 (0.025)	– 0.029 (0.027)
Once in a while dialect-speaking	– 0.017 (0.020)	– 0.005 (0.023)
e. Educational heterogeneity		
Daily dialect-speaking	– 0.012 (0.022)	– 0.018 (0.025)
Daily dialect-speaking × high education	– 0.081 (0.041)**	– 0.035 (0.040)

Based on 5721 observations of 1761 males and 5597 observations of 1756 females except for panel b. All regressions include individual characteristics, family characteristics, calendar year fixed effects and province fixed effects. Standard errors are clustered at the level of the individual

In panel b, we investigate provincial heterogeneity of the wage penalty. We examine dialect-speaking separately for the “Randstad” area and the remaining eight provinces. In the Randstad provinces the share of daily dialect-speakers is only 1–3%, substantially below average. As shown in panel b1, the wage of male dialect speakers in the Randstad is about 16% lower than the wage of male non-dialect speakers. As shown in panel b2, in the non-Randstad provinces, for males the wage penalty of dialect-speaking is about 4%. It is striking that for males the wage penalty is lower in the provinces where the linguistic distance to Standard Dutch is larger. Apparently, for males the wage penalty of dialect-speaking is not so much related to linguistic distance but instead to the share of people speaking dialect. If more people speak a dialect the wage penalty of dialect-speaking is smaller.

As in Friesland Frisian is recognized as a co-official language, we examine whether the wage penalty still exists. As shown in panel b3, we do not find a significant wage penalty of dialect-speaking, and for males there is an insignificant wage premium of dialect-speaking. In panel b4 we find robust results by excluding individuals from Friesland. For females, dialect-speaking has no significant wage effect in either of the provincial subsamples.

In panel c we use as indicator of dialect-speaking, whether an individual ever spoke dialect during his or her lifetime. We find the parameter estimates are somewhat smaller (in absolute terms) but not very different from the baseline estimates. In panel d, we add dummies for each category of dialect-speaking, ranging from daily dialect-speaking to dialect-speaking once in a while. This is to measure

Table 5 Pooled cross-section parameter estimates of daily-dialect speaking; other labor market outcomes

	Males	N	Females	N
a. Employment	– 0.001 (0.016)	10,291	– 0.014 (0.023)	11,927
b. Log working hours	0.024 (0.014)	6362	– 0.030 (0.028)	6397
c. Non-manual job	– 0.075 (0.027)***	6862	– 0.098 (0.028)***	6763
d. Professional ranking	– 0.221 (0.116)*	7714	– 0.269 (0.122)**	7612

All regressions include individual characteristics, family characteristics, calendar year fixed effects and province fixed effects. In parentheses, standard errors clustered at the level of the individual

whether the wage penalty increases with the frequency of dialect usage. Not surprisingly, for males daily dialect-speaking has a significant negative effect of 5.3%, while speaking dialects regularly and once in a while have smaller and non-significant effects.

Panel e shows the heterogeneity of the wage penalty according to educational attainment. We include an interaction term of daily dialect-speaking with a dummy for having a higher vocational or university degree. We find that for both genders there is a larger wage penalty of dialect-speaking for the high-educated, while dialect-speaking is less of a concern for low-educated people. This may be because high-educated are targeted at occupations which require communication skills with people all over the country. Moreover, dialect-speakers with a higher education may be perceived by employers to have a lower unobserved ability whereas for low educated workers dialect-speaking is less of a problem.

In Table 5 we indicate how dialect-speaking affects other labor market outcomes i.e. employment, working hours and type of jobs. An individual is considered to be employed if he or she has any type of paid work, including family business and self-employment. Table 5 shows that neither for males nor for females daily dialect-speaking affects the probability to have job. Conditional on having a job, daily dialect-speaking does not have a significant effect on working hours either. However, conditional of having a job, daily dialect-speakers are less likely to have a non-manual job. Moreover, daily dialect-speakers are more likely to have a professionally lower ranked job.

4 How to explain our findings?

Our major finding is that dialect-speaking is related to a wage penalty as well as occupation sorting. Dialect-speakers are sorted to occupations with less income, or they are paid less than their counterparts in the same occupation. There are several potential mechanisms accounting for this disadvantage.

The first mechanism responsible for this is the inability to speak Standard Dutch. Yao et al. (2016) for example investigate the relationship between dialect speaking and academic performance of 5–6 year old children in the Netherlands. They find that dialect speaking has modest negative effects on language skills of young boys while young girls are not affected. However, it is not clear that this effect persists

over time till adulthood. Giesbers et al. (1988) reporting the results of an investigation at primary schools in two dialect areas in the southern part of the Netherlands conclude that dialect-speaking is not necessarily an educational disadvantage. However, they do find that dialect-speaking children are disadvantaged in terms of school performance and their choice of secondary education. One of the surprising findings is that school teachers give lower grades to dialect-speaking children under their supervision. However, essays of dialect-speaking children were graded similarly as essays by non-dialect-speaking children by external graders, unaware of the speech patterns of the children.

A second possible mechanism is perception of lower productivity. Even if workers do not have difficulty in speaking Standard Dutch at work, colleagues and employers may still recognize a dialect accent. Whether dialect-speakers actually have a lower productivity is hard to tell. From Table 4, the high educated workers face a larger wage penalty than their counterparts. We consider this fact as an indication that perceived productivity is not in line with actual productivity. The fact that a wage penalty of daily dialect-speaking is only present for males but not for females may also be related to differences in location of the job. Females are more likely to work closer to home than males. In our sample, the average distance for females between work and home was 10.5 km with a commuting time of 20.6 min. For males the average distance to work was 16.3 km and 27.5 min.

Our main finding of a wage penalty for dialect speaking is seemingly at odds with existing literature on the effect of being bilingual which suggests that there is a wage premium for being proficient in a second language (Christofides and Swidinsky 2010; Di Paolo and Raymond 2012; Chen et al. 2014; Cappellari and Di Paolo 2018). However, there is a clear distinction between speaking an additional language and speaking a dialect. Speaking a second language signals social identity, increases search efficiency and decreases transaction cost in the local labor market. Our context is different from bilingual systems in the sense that a dialect cannot be viewed as a separate language. Dialect speech pattern will affect how people speak the standard language and associate with social status. Moreover, although speaking a dialect increases search efficiency in the local market, it may not compensate for the disadvantage in national labor markets unless the local area is more developed than the country average, like for example, Catalonia in Spain and Shanghai in China. This is in line with our finding of a bigger wage penalty for dialect-speaking for individuals living in a Randstad province.

5 Conclusions

We investigate whether daily dialect-speaking affects earnings. Using data from the Netherlands we conclude that male workers who speak a dialect daily have significantly lower hourly wages. For females we also find negative effect but this is not significantly different from zero. High-educated male workers face a larger wage penalty of dialect speaking than their counterparts. Finally, we find province heterogeneous effects such that in areas where more people speak dialect, there is a

smaller wage penalty. Our data do not allow us to make a distinction between various mechanisms that lead to a wage penalty for male workers. We can only speculate that the wage penalty is partly related to occupational sorting and partly to conscious or unconscious discrimination by employers or co-workers.

Our findings are all the more surprising since there are clear limitations to our empirical analysis. Although our results are robust to various sensitivity tests our analysis does not go beyond conditional correlations. Furthermore, our indicator of dialect-speaking is self-reported and related to daily use. So, it is not clear which dialect is spoken nor is it clear how the speech pattern is perceived by listeners. Similarly, if an individual indicates speaking Standard Dutch it is not clear whether this is also the perception of listeners. It could be that a self-reported Standard Dutch speaker is perceived to speak according to a particular area in the Netherlands, i.e. speech pattern can reveal area of residence. Also, it is not clear whether an individual speaks dialect at work. It may be that a daily dialect-speaker speaks Standard Dutch at work with a flavor of dialect. Even if the pronunciation makes the meaning of words perfectly clear, it may be slightly different from area to area.

Acknowledgements We are grateful to CentER Data for making the LISS data available for this paper. We also thank participants at the meetings of the European Society for Population Economics 2016, the European Association of Labour Economists 2016 and seminars at Università Ca' Foscari (Venezia), University of Padova, Luxembourg Institute of Socio-Economic Research and the workshop “Language skills for social and economic inclusion” at Humboldt University (Berlin) for their comments on a previous version of the paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Details on the variables

Explanatory characteristics

Individual characteristics

- Daily dialect speaking: dummy variable, 1 if the answer to the question “Do you ever speak dialect?” is “Yes, daily”, 0 if “Yes, regularly”, “Yes, sometimes” or “No, never”.
- Age dummies: dummy variables for age categories 25–34, 35–44, 45–54 and 55+. Reference group is age below 25.
- Education dummies: dummy variables for educational categories primary school, intermediate and higher secondary school, intermediate and higher vocational school, university or higher. Reference group is primary school education.
- Religious: dummy variable for having a religion.
- Partner: dummy variable for having a partner living at the same address.

Family characteristics

- Number of children: number of dependent children in the household.
- Urbanization: dummy variables for degree of urbanization of the municipality of residence (inhabitants per square-kilometer); very urban (1500–2500), moderately urban (1000–1500), slightly urban (500–1000), not urban (less than 500). Reference group is extremely urban (2500 or more).

Province characteristics

- GDP per capita: GDP, value added at market prices of the total economy, per capita.
- Employment: number of residents having a paid work for at least one hour a week, including self-employees.
- Population: total number of people residing in a province by January 1.
- Area Main Roads: total area of main roads in km².

Dependent variables

- Employment: dummy variable for having any type of paid work, including family business and self-employment.
- Working hours: average hours of work per week.
- Hourly wage: calculated based on monthly earnings and working hours.
- Manual jobs: unskilled and trained manual work, semi-skilled manual work, agrarian profession and skilled and supervisory manual work. Non-manual jobs include all other jobs, i.e. academic, independent, supervisory, commercial or mental work professions at any level.
- Professional ranking: based on average monthly earnings. The ranking is as follows:

Profession	Monthly earnings (Euro)	Professional ranking
Unskilled and trained manual work	1623	1
Semi-skilled manual work	2447	2
Agrarian profession	2562	3
Skilled and supervisory manual work	2646	4
Other mental work	2865	5
Intermediate academic or independent	3164	6
Intermediate supervisory or commercial	3335	7
Higher academic or independent	4160	8
Higher supervisory profession	5230	9
Average	3398	5

Table 6 Sample characteristics by gender and daily dialect-speaking

	Males		Females	
	No	Yes	No	Yes
Dialect speakers				
Speaks dialect (%)				
Never	68	0	72	0
Once in a while	21	0	19	0
Regularly	11	0	9	0
Daily	0	100	0	100
Personal characteristics				
Age	44	46	43	45
Education (%)				
Primary education	7	6	6	6
Lower secondary education	16	28	20	32
Intermediate secondary education	37	40	37	40
Higher education	40	27	36	22
Number of children	1.1	1.0	1.1	1.1
Living with a Partner (%)	79	80	78	80
Has a religion(%)	20	18	21	15
Urbanized area (%)	45	17	43	19
Province characteristics				
Log(GDP in Euro per capita)	10.5	10.4	10.5	10.4
Log(Employment)	6.9	6.5	6.9	6.4
Log(Population (1000))	14.4	14.0	14.4	14.0
Area in use of main roads (km ²)	112	107	111	103
N	8154	2265	10,240	1922
Labor market and job characteristics				
Employment (% population)	76	78	70	64
Weekly working hours	39.8	40.0	28.5	26.9
Hourly wage (Euro)	20.9	18.6	17.5	15.5
Manual job (%)	19.9	38.1	9.2	22.2
Professional ranking	5.0	4.5	4.6	4.1

The level of education dummy variables are based on Statistics Netherlands categories, primary education, lower secondary education, intermediate secondary education and higher education. In an urbanized area population density is above 1500 inhabitants per squared kilometer. *N* is the number of observations

Table 6 shows summary statistics by gender and daily dialect-speaking. Of the non-daily dialect-speakers about 70% has never spoken a dialect, 20% does it once in a while and 10% speaks dialect regular but not daily. Compared to standard-Dutch speakers daily dialect-speakers have slightly lower education and are less likely to live in urbanized areas. There is not much difference in province characteristics between the two groups. In terms of labor market characteristics, employment rate and weekly working hours are similar but daily dialect-speakers on average have lower hourly

wages. Similarly, employed daily dialect-speakers has a much higher share of manual jobs and a lower professional ranking.

References

- Bendick M Jr, Rodriguez RE, Jayaraman S (2010) Employment discrimination in upscale restaurants: evidence from matched pair testing. *Soc Sci J* 47:802–818
- Bleakley H, Chin A (2004) Language skills and earnings: evidence from childhood immigrants. *Rev Econ Stat* 86(2):481–496
- Cappellari L, Di Paolo A (2018) Bilingual schooling and earnings: evidence from a language-in-education reform. *Econ Educ Rev* 64:90–101
- Carlson HK, McHenry MA (2006) Effect of accent and dialect on employability. *J Employ Couns* 43:70–83
- Chen Z, Lu M, Xu L (2014) Returns to dialect identity exposure through language in the Chinese labor market. *China Econ Rev* 30:27–43
- Cheshire J, Edwards V, Munstermann H, Weltens B (1989) Dialect and education: some European perspectives. Clevedon, Multilingual Matters Ltd
- Chetty R, Hendren N, Katz LF (2016) The effects of exposure to better neighborhoods on children: new evidence from the moving to opportunity experiment. *Am Econ Rev* 106(4):855–902
- Christofides LN, Swidinsky R (2010) The economic returns to the knowledge and use of a second official language: English in Quebec and French in the rest-of-Canada. *Can Pub Policy* 36:137–158
- Das MB (2013) Exclusion and discrimination in the labor market. World Bank, Washington Background paper for the World Development Report
- Di Paolo A, Raymond JL (2012) Language knowledge and earnings in Catalonia. *J Appl Econ* 15(1):89–118
- Driessen G (2005) In Dutch? Usage of Dutch regional languages and dialects. *Lang Cult Curric* 18:271–285
- Falck O, Heblich S, Lameli A, Sudekum J (2012) Dialects, cultural identity and economic exchange. *J Urban Econ* 72(2–3):225–239
- Gao W, Smyth R (2011) Economic returns to speaking 'standard Mandarin' among migrants in China's urban labour market. *Econ Educ Rev* 30:342–352
- Giesbers H, Kroon S, Liebrand R (1988) Bidialectalism and primary school achievement in a Dutch dialect area. *Lang Educ* 2(2):77–93
- Gorter D (2005) Three languages of instruction in Frysland. *Int J Sociol Lang* 171:57–73
- Grogger J (2011) Speech patterns and racial wage inequality. *Journal of Human Resources* 46:1–25
- Grogger J (2018) Speech and wages. *J Hum Resour* (forthcoming)
- Lang K (1986) A language theory of discrimination. *Q J Econ* 101:363–382
- Levenshtein V (1966) Binary codes capable of correcting deletions, insertions and reversals. *Cybern Control Theory* 10:707–710
- Miranda A, Zhu Y (2013a) The causal effect of deficiency at English on female immigrants' labor market outcomes in the UK. IZA Discussion Paper No. 7841
- Miranda A, Zhu Y (2013b) English deficiency and the native immigrant wage gap. *Econ Lett* 118(1):38–41
- Rickford JR, Duncan GJ, Gennetian LA, Gou RY, Greene R, Katz LF, Kessler RC, Kling JR, Sanbonmatsu L, Sanchez-Ordóñez AE, Sciandra M, Thomas E, Ludwig J (2015) Neighborhood effects on use of African-American Vernacular English. *Proc Natl Acad Sci USA* 112(38):11817–11822
- Swadesh S (1952) Lexico-statistic dating of prehistoric ethnic contacts. *Proc Am Philos Soc* 96:121–137
- Van Bezooijen R, Heeringa W (2006) Intuitions on linguistic distance: geographically or linguistically based? In: Koole T, Nortier J, Tahitu B (eds) *Vijfde sociolinguïstische conferentie*. Eburon Uitgeverij BV, Delft, pp 77–87
- Yao Y, Ohinata A, van Ours JC (2016) Educational consequences of language for young children. *Econ Educ Rev* 54:1–15
- Yao Y, van Ours JC (2015) Language skills and labor market performance of immigrants in the Netherlands. *Labour Econ* 34:76–85
- Yao Y, van Ours JC (2016) The wage penalty of dialect speaking. CEPR Discussion Paper (11610)