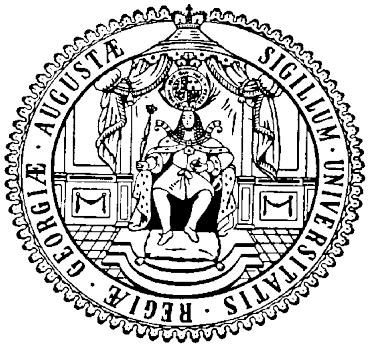
Linguistic Relativity and Economic Outcomes

How Language Influences Thought and Behaviour of Individuals

A thesis submitted in fulfillment of the requirements
for the degree of Doctor Rerum Politicarum
from the Faculty of Economic Sciences
University of Göttingen



by

Florian Rottner

born in Nürnberg

Göttingen, 2022

First Supervisor:	Prof. Dr. Robert Schwager
Second Supervisor:	Prof. Dr. Thushyanthan Baskaran
Third Supervisor:	Prof. Dr. Michael Wolff

"Now, the beauty of linguistic diversity is that it reveals to us just how ingenious and how flexible the human mind is. Human minds have invented not one cognitive universe, but 7,000... there are 7,000 languages spoken around the world."

Lera Boroditsky

Abstract

Research by linguists, psychologists and cognitive scientists has shown that the grammatical and lexical structures of languages affect human cognition. This hypothesis of linguistic relativity has been taken up by economists who began to study the impact of language structures on economic outcomes and people's beliefs. In this thesis, I empirically investigate how differences in language structure influence economic outcomes and people's beliefs, thereby contributing to the growing body of economic literature about linguistic relativity and its effects.

In the three main chapters of this thesis, I study the effects of three different language structures; i) how the differences in the grammatical marking of future events in languages influence the future-orientation of companies (**Chapter 2**), ii) the difference in the distinction of politeness groups in second person pronouns of languages and their effect on the attitude of individuals towards foreigners (**Chapter 3**) and iii) how the gender system of a language influences norms about the role of women in family and work life in a society and labour market outcomes for women (**Chapter 4**).

Overall, this thesis provides further evidence that differences in the structure of languages not only affect human perception, but also have an impact on people's beliefs and economic outcomes. The grammatical marking of future events in the language of board members influences how future-oriented a company acts (**Chapter 2**). The distinction of at least two politeness groups in second person pronouns reduces the trust individuals put into foreigners (**Chapter 3**). A gender system based on biological sex reinforces the belief in the traditional role of women in family and working life and also has a negative impact on women's labour market outcomes (**Chapter 4**).

Zusammenfassung

Forschung von Linguisten, Psychologen und Kognitionswissenschaftlern hat gezeigt, dass die grammatikalischen und lexikalischen Strukturen von Sprachen die Wahrnehmung des Menschen beeinflussen. Diese Hypothese der linguistischen Relativität wurde in den letzten Jahren von Ökonomen aufgegriffen und es wurde begonnen den Einfluss von Sprachstrukturen auf ökonomische Ergebnisse und die Überzeugungen von Menschen zu untersuchen. Ich untersuche in dieser Dissertation empirisch, wie sich die strukturellen Unterschiede zwischen Sprachen auf die Überzeugungen von Menschen und ihre ökonomischen Ergebnisse auswirken. Damit leiste ich einen Beitrag zur wachsenden ökonomischen Literatur, die sich mit der linguistischen Relativität und ihrer Effekte befasst.

In den drei zentralen Kapiteln dieser Arbeit untersuche ich die Effekte von drei Sprachstrukturen: i) wie sich die Unterschiede in der grammatischen Markierung von zukünftigen Ereignissen in Sprachen auf die Zukunftsorientierung von Firmen auswirken (**Kapitel 2**), ii) die Unterschiede in der Anzahl von Höflichkeitsgruppen, die in den Pronomina der zweiten Person einer Sprache unterschieden werden, und ihre Effekte auf die Einstellung, die Menschen gegenüber Fremden haben (**Kapitel 3**) und iii) wie das Geschlechtersystem einer Sprache die Normen einer Gesellschaft über die Rolle der Frau im Familien- und Berufsleben und die Arbeitsmarktergebnisse von Frauen beeinflussen (**Kapitel 4**).

Insgesamt liefert diese Dissertation weitere Belege dafür, dass die Unterschiede in den Strukturen von Sprachen nicht nur die menschliche Wahrnehmung beeinflussen, sondern sich auch auf die Überzeugungen von Menschen und ihr Verhalten auswirken. Die Art und Weise, wie zukünftige Ereignisse in den Sprachen von Vorstandsmitgliedern einer Firma grammatisch markiert werden, beeinflusst wie zukunftsorientiert sich diese Firmen verhalten (**Kapitel 2**). Werden in den Pronomina der zweiten Person mindestens zwei Höflichkeitsgruppen unterschieden, reduziert dies das Vertrauen, das eine Personen Fremden entgegenbringt (**Kapitel 3**). Basiert das Geschlechtersystem einer Sprache auf den biologischen Geschlechtern, verstärkt das den Glauben von Personen an traditionelle Geschlechterrollen in Familien- und Berufsleben und es hat zusätzlich auch einen negativen Einfluss auf die Arbeitsmarktergebnisse von Frauen (**Kapitel 4**).

Acknowledgements

Zuallererst möchte ich mich bei meinem Doktorvater Robert Schwager bedanken. Er hat mir die Möglichkeit gegeben, meinen eigenen Interessen und Forschungsfragen nachzugehen. Dabei stand er mir auf meinem Weg zu jeder Zeit mit Rat und hilfreichen Kommentaren zu meiner Forschung und im Arbeitsalltag zur Seite. Des Weiteren gilt mein Dank auch Thushyanthan Baskaran und Michael Wolff, die sich bereit erklärt haben Mitglieder meiner Prüfungskommission zu sein.

Ich möchte mich auch bei meinen Kollegen Raphael Brade, Kamila Danilowicz-Gösele, Monika Jackmann, Katharina Lerche, Ansgar Quint und Ann-Marie Sommerfeld für ihre hilfreiche Kommentare, den interessanten Gesprächen während den Mittagessen oder beim Kaffee und für die angenehme Zusammenarbeit am Lehrstuhl bedanken. Mein Dank geht auch an all die anderen Menschen in Göttingen, die mich auf meinem Weg begleitet haben, insbesondere allen Mitspielern bei der wöchentlichen Fußballrunden, die für den nötigen sportlichen Ausgleich gesorgt haben.

Weiterhin möchte ich all meinen Freunden in Nürnberg und anderswo danken, die mich teilweise schon seit der Grundschule begleiten. Ich danke euch für all die schönen Tage und Abende, mit unvergesslichen Erlebnissen und spannenden Gesprächen. Ihr habt mich immer daran erinnert, dass es noch eine Welt jenseits der Universität und der Forschung gibt.

Ganz besonders möchte ich mich bei meiner Familie und insbesondere meinen Eltern, Beate und Peter, und meiner Schwester Steffi bedanken. Ich wusste euch immer an meiner Seite und ihr habt mich bei all meinen Entscheidungen auf meinem Weg unterstützt. Ohne euch wäre ich heute nicht die Person, die ich bin. Dafür werde ich euch immer dankbar sein.

Mein größter Danke gilt meiner Verlobten Antje. Diese Promotion ist genauso ihre Leistung wie sie meine ist. Ich danke dir für deine Liebe und deine Unterstützung, deren ich mir immer sicher war, trotz der vielen Kilometer zwischen uns. Sie waren meine Quelle der Kraft und Inspiration. Du warst immer und bist mein Vorbild und meine Motivation mehr zu erreichen und meinen Träumen zu folgen. Ohne dich wäre diese Promotion nicht möglich gewesen.

Contents

1 General Introduction **1**

1.1 Linguistic Relativity 1

1.2 Thesis Outline 7

2 Language and the Future: Board Members and the Investment in the Future **9**

2.1 Introduction 10

2.2 Future-Time Reference in Languages 13

2.3 Hypothesis 15

2.4 Data 16

2.4.1 Dependent Variable 16

2.4.2 Independent Variable 18

2.4.3 Controls 20

2.5 Empirical Approach 23

2.5.1 Model 23

2.5.2 Unbalanced Panel Data 23

2.6 Main Results 25

2.6.1 Intangible Asset Growth Rate 25

2.6.2 R&D Expenditures 28

2.7 Discussion 31

2.8 Conclusion 32

3 Language and Xenophobia: The Effect of Politeness Distinction in Pronouns on your Attitude towards Foreigners **34**

3.1 Introduction 35

3.2 Politness Distinction 39

3.3 Hypothesis 41

3.4 Data . . . 42
3.4.1 Dependent Variable . . . 43
3.4.2 Independent Variable . . . 43
3.4.3 Control Variables . . . 45
3.5 Empirical Approach . . . 47
3.6 Main Results . . . 48
3.7 Robustness Tests . . . 50
3.7.1 Immigrants . . . 50
3.7.2 General Preference Survey . . . 53
3.7.3 General Trust and Language . . . 58
3.8 Conclusion . . . 59

4 Language and Gender: How Linguistic Differences Influence a Woman's Labour Market Outcomes 61

4.1 Introduction . . . 62
4.2 Linguistic relativity . . . 67
4.3 Data . . . 68
4.3.1 Socio-Economic and Demographic Data . . . 68
4.3.2 Language Features . . . 71
4.4 Empirical Approach . . . 74
4.5 Main Results . . . 75
4.5.1 Gender Intensity Index . . . 75
4.5.2 Individual Language Features . . . 77
4.6 Robustness Tests . . . 83
4.6.1 Immigrants . . . 83
4.6.2 Wider Sample . . . 86
4.7 Conclusion . . . 88

A Language and the Future 90

A.1 Company descriptives . . . 90
A.2 Additional Regressions . . . 94
A.2.1 Intangible Asset Growth Rate . . . 94
A.2.2 R&D Expenditures . . . 100

B Language and Xenophobia **106**

B.1 Regressions with SE Clustered at Country Level 106

B.2 Marginal Effects 108

B.2.1 Immigrants 108

B.2.2 Official Language 110

B.2.3 GPS 112

B.3 Summary Statistics 116

C Language and Gender **119**

C.1 Data 119

C.2 Non-Linear Models 123

C.3 Alternative GII 124

C.4 Sex-based Only Dataset 125

Bibliography **127**

Declaration for Admission to the Doctoral Examination **138**

Author Contributions **139**

List of Tables

2.1 Distribution of Dependent Variables 17
2.2 Company Descriptives 18
2.3 Descriptive Statistics Strong-FTR: Mean 18
2.4 Descriptive Statistics Strong-FTR: Median 19
2.5 Descriptive Statistics Inflectional Marking: Mean 20
2.6 Descriptive Statistics Inflectional Marking: Median 21
2.7 S-Test for Sample Selection Bias 24
2.8 Effect on Intangible Asset Growth Rate: Mean 26
2.9 Effect on Intangible Asset Growth Rate: Median 27
2.10 Effect on R&D Expenditures: Mean 29
2.11 Effect on R&D Expenditures: Median 30

3.1 Trust in Foreigners 43
3.2 Politeness Distinction Across Individuals 44
3.3 Politeness Distinction by Country 45
3.4 Language Effect on Trust 48
3.5 Language Effect on Trust: Cultural Dimensions 49
3.6 Politeness Distinction Across Languages 51
3.7 Language Effect on Trust: Immigrants 52
3.8 Language Effect on Trust: Not Speaking Official Language 53
3.9 Politeness Distinction Across Languages (GPS) 54
3.10 Language Effect on Trust (GPS) 55
3.11 Language Effect on Trust: Cultural Preferences (GPS) 56
3.12 Language Effect on Trust: Immigrants (GPS) 57
3.13 Language Effect on Trust: Not Speaking Official Language (GPS) 58

3.14 Language Effect on General Trust 59

4.1 Summary statistics: Dependent Variables 70
4.2 Summary Statistics: Descriptives 71
4.3 Gender Expression in Pronouns 73
4.4 Summary statistics: Language Features 74
4.5 Effect of Language on Gender Norms and Labour Market Outcomes . . 76
4.6 Effect of Individual Language Features: Society's Beliefs 78
4.7 Effect of Individual Language Features: Female Labour Market Outcomes I 79
4.8 Effect of Individual Language Features: Female Labour Market Outcomes II 80
4.9 Effect of Individual Language Features: Female Labour Market Outcomes III 82
4.10 Effect of Individual Language Features: Immigrant Subsample I 84
4.11 Effect of Individual Language Features: Immigrant Subsample II 85
4.12 Effect of Individual Language Features: Female Labour Market Outcomes ESS Round 5-9 87

A.1 Observations per Country (R&D expenditures) 90
A.2 Observations per Sector (R&D Expenditures) 91
A.3 Observations per Country (Intangible Assets Growth Rate) 92
A.4 Observations per Sector (Intangible Assets Growth Rate) 93
A.5 Effect on Intangible Asset Growth Rate $<$ 200%: Mean 94
A.6 Effect on Intangible Asset Growth Rate $<$ 150%: Mean 95
A.7 Effect on Intangible Asset Growth Rate $<$ 75%: Mean 96
A.8 Effect on Intangible Asset Growth Rate $<$ 200%: Median 97
A.9 Effect on Intangible Asset Growth Rate $<$ 150%: Median 98
A.10 Effect on Intangible Asset Growth Rate $<$ 75%: Median 99
A.11 Effect on R&D Expenditures Proportions $<$ 1: Mean 100
A.12 Effect on R&D Expenditures Proportions $<$ 0.75: Mean 101
A.13 Effect on R&D Expenditures Proportions $<$ 0.25: Mean 102
A.14 Effect on R&D Expenditures Median Proportions $<$ 1: Median 103

A.15 Effect on R&D Expenditures Median Proportions < 0.75: Median 104
A.16 Effect on R&D Expenditures Median Proportions < 0.25: Median 105

B.1 Language Effect on Trust: SE clustered at country level 106
B.2 Language Effect on Trust: Cultural Dimensions and SE clustered at country level 107
B.3 Language Effect on Trust: Immigrants (Marginal Effects) 108
B.4 Language Effect on Trust: Immigrants and Cultural Dimensions (Marginal Effects) 109
B.5 Language Effect on Trust: Not Speaking Official Language (Marginal Effects) 110
B.6 Language Effect on Trust: Not Speaking Official Language and Cultural Dimensions (Marginal Effects) 111
B.7 Language Effect on Trust: Immigrants (GPS) (Marginal Effects) 112
B.8 Language Effect on Trust: Immigrants and Cultural Preferences (GPS) (Marginal Effects) 113
B.9 Language Effect on Trust: Not Speaking Official Language (GPS) (Marginal Effects) 114
B.10 Language Effect on Trust: Not Speaking Official Language and Cultural Preferences (GPS) (Marginal Effects) 115
B.11 Politeness Distinction by Country (GPS) 116
B.12 List of Languages 117
B.13 List of Languages (GPS) 118

C.1 Summary statistics: Descriptives (Female Subsample) 119
C.2 List of Feminine Occupations 120
C.3 List of Languages and its Features 121
C.4 Summary Statistics: GII per Country 122
C.5 Non-Linear Models 123
C.6 Effect of Language and Gender Norms and Labour Market Outcomes (Alternative GII) 124
C.7 Effect of Language and Gender Norms and Labour Market Outcomes: Only Sex-based 125

C.8 Effect of Language and Gender Norms and Labour Market Outcomes: Only Sex-based ESS Round 5-9 126

List of Abbreviations

EBIT	**E**arnings **B**efore **I**nterest and **T**axes
EBITDA	**E**arnings **B**efore **I**nterest, **T**axes, **D**epreciation and **A**mortization
EPO	**E**uropean **P**atent **O**ffice
ERP	**E**vent **R**elated **P**otential
ESS	**E**uropean **S**ocial **S**urvey
fMRI	**f**unctional **M**agnetic **R**esonance **I**maging
FTR	**F**uture-**T**ime **R**eference
GII	**G**ender **I**ntensity **I**ndex
GPS	**G**lobal **P**reference **S**urvey
ISSP	**I**nternational **S**ocial **S**urvey **P**rogram
MCAR	**M**issing **C**ompletely **A**t **R**andom
WALS	**W**orld **A**tlas of **L**anguage **S**tructure
WVS	**W**orld **V**alue **S**urvey

Chapter 1

General Introduction

1.1 Linguistic Relativity

The idea that language influences the way an individual thinks and perceives the reality around them has existed in linguistics, philosophy, anthropology, and psychology for centuries. Different versions of this concept can be found in the writings of Roger Bacon in medieval times and all-around Europe in the late 17th and early 18th century, e.g., in the ideas of Locke, Diderot, and Herder. The work of these scholars were driven by concerns regarding the translation of religious and scientific knowledge between different languages and interest about the origin of languages during the development of the human cognition and culture. Language also played an important role in strengthening the perceived superiority of one's own culture during the time of expansion and colonialism by European nations. In the 19th century, this tradition was taken up by Humboldt, who saw a different *Weltanschauung* (world view) in each language, and it is also implicit in Saussure's work on structuralism in languages (Gumperz et al., 1996; Lucy, 1997; Malmkjaer, 2009).

Today, the idea of linguistic relativity is most commonly associated with the two American linguists Edward Sapir and Benjamin L. Whorf, who were both influenced by Franz Boas.[1] Their work with Native American languages in the early 20th century led Sapir and Whorf to formulate their hypothesis that the various categories and distinctions in languages cause their speakers to perceive the reality around them differently and, because of the differences in perceived reality, to also act differently (Boroditsky, 2001;

[1] Hence, the common use of "the Sapir-Whorf hypothesis" as synonym for the hypothesis of linguistic relativity.

Gumperz et al., 1996; Lucy, 1997; Malmkjaer, 2009). Whorf formulated this idea that language determines the thoughts and actions of individuals in the following way.

> *From this fact proceeds what I have called the "linguistic relativity principle," which means, in informal terms, that users of markedly different grammars are pointed by their grammars towards different types of observations and different evaluations of externally similar acts of observation, and hence are not equivalent as observers, arrive at somewhat different views of the world.* (Whorf, 1956, p. 221)

> *We are thus introduced to a new principle of relativity, which holds that all observers are not led by the same physical evidence to the same picture of the universe, unless their linguistic backgrounds are similar, or can in some way be calibrated.* (Whorf, 1956, p. 214)

The formulation of this hypothesis was followed by a series of empirical studies by anthropologists and linguists testing the hypothesis as well as by a few psychologists investigating the relationship between language and memory (Lucy, 1992). Work on the perception of colour showed that Dani people[2] were easily able to learn English colour categories despite having only two colour categories in their own language. Thereby, these findings undermined the strong Whorfian view and the hypothesis of linguistic determinism that thought and action are entirely determined by language (Heider, 1972; Rosch, 1975; Rosch, 1978). Therefore, the view of linguistic determinism has since been abandoned by researchers in the field.

The strong form of Whorf's hypothesis, linguistic determinism, was not tenable, but researchers began to pursue weaker forms of linguistic relativity. One of those hypothesis for example is Slobin's *thinking for speaking* (Slobin, 1987). Speakers of different languages might be biased to attend to and encode different aspects of their experience because their language makes those aspects grammatically obligatory and therefore more salient. In English, for example, the speaker has to include tense to say that *the elephant ate the peanuts*. In Mandarin on the other hand the inclusion of the time when an event occurred is optional. In Russian indicating tense is not enough, the speaker also has to indicate the gender of the peanut-eater and if all of the peanuts were eaten

[2] A tribe in New Guinea.

or only some (Boroditsky, 2006). Slobin draws inspiration for his idea of *thinking for speaking* in the early work of Sapir, who wrote:

> *[The forms of each language] establish a definite relational feeling or attitude towards all possible contents of expression and, through them, towards all possible contents of experience, in so far, of course, as experience is capable of expression in linguistic terms.* (Sapir, 1949, p. 153)

Since then, researchers have explored many different domains of languages that are likely to reveal linguistic influences on the thought and behaviour of individuals. One of those domains is how languages describe spatial dimensions (e.g., Bowerman (1996), Levinson (1996), Li and Gleitman (2002) and McDonough, Choi, and Mandler (2003)). The Dutch language for example relies heavily on relative spatial terms to describe the locations of objects, e.g., the remote is to the left of the TV. Tzeltal[3] on the other hand uses a system of absolute reference, similar to the north/south distinction in Dutch. To test if this difference in the use of spatial terms between the two languages has cognitive consequences, speakers of Dutch or Tzeltal were seated at a table with an arrow laying in front of them, either pointing right/north or left/south. Following that, they were rotated 180 degrees to a second table with two arrows, one pointing right/south and the other left/north. The participants were then asked to point at the arrow which is like the one before. When Dutch speakers saw an arrow pointing right on the first table, they also chose the arrow pointing right on the second table. They approached the question following relative spatial terms. Tzeltal speakers did exactly the opposite. If the first arrow pointed north, they chose on the second table also the arrow pointing north, though it pointed to the left now (Levinson, 1996).

Another dimension in which languages differs is the description of time. English speakers use horizontal spatial terms to talk about time, e.g., meetings are pushed back, or the good times are laying ahead of us. Mandarin speakers on the other hand also use vertical metaphors to talk about time, earlier events are up and later ones are down, in addition to horizontal terms (Scott, 1989). This leads Mandarin speakers to confirm faster that March comes before April if they see a vertical array of objects compared to a horizontal array and the reverse was true for English speakers (Boroditsky, 2001).

[3] A Mayan language.

An interesting combination of both domains can be found among the Australian Aboriginal Community of Pormpuraaw. The Pormpuraawan languages use absolute cardinal directions (north, south, east, west) to describe locations, which leads Pormpuraawans to organize time from east to west instead of from left to right like English speakers. Therefore, when asked to order photos of humans by age, Pormpuraawans order them from east to west irrespective of their own orientation at that point, whereas English speakers order them from left to right relative to their own line of sight (Boroditsky and Gaby, 2010).

The grammatical categorisation of objects is another domain in which languages differ. Many languages have grammatical gender system whereby all nouns are assigned a gender, while other languages are genderless. Moreover, the languages that assign a gender to nouns also differ in the number of genders they have. Some only distinguish between masculine and feminine whereas some also assign neuter or other genders (Boroditsky, L. A. Schmidt, and Philipps, 2003). The grammatical gender assigned to an object has an influence on the mental representation of this object by individuals. When Spanish and German speakers were asked to rate similarities between pictures of females or males and pictures of objects, which had opposite genders in Spanish and German, they both rated grammatically feminine objects to be more similar to females and grammatically masculine objects to be more similar to males (Boroditsky and Philipps, 2003). For German speakers *die Brücke* (bridge) is more similar to females, because the German language assigns it to the feminine gender. For Spanish speakers on the other hand, it is more similar to a males, since *el puente* (bridge) is masculine in Spanish.

However, research is not only being conducted about the question if a language influences thought and behaviour but there is also a nuanced range of proposals on how language may influence them. These proposals range from Whorf's linguistic determinism to Slobin's thinking for speaking to more recent proposals such as language as a spotlight, which sees certain constructs or characteristics of language as a spotlight that make certain aspects of reality more salient to the speaker.[4,5]

[4]For a detailed overview of the different proposals in the literature see P. Wolff and Holmes (2011).

[5]For further readings on the history on linguistic relativity and recent advances in the field see Boroditsky (2006), Casasanto (2008), Lucy (2016).

In addition to the work by linguists and anthropologists, linguistic relativity has also been studied in the field of neuroscience and cognitive perception. Researchers in these fields use a variety of methodological techniques to find tangible physiological evidence for the effect of language on cognitive processing in the human brain. Thereby, they add biologically grounded evidence to the question how language influences thought and behaviour of individuals.

Stutterheim et al. (2012) used eye-tracking to show that language affects the focus of vision of a person. Speakers of languages that typically include endpoints in their descriptions, take more looks at the endpoint region of a visual stimulus compared to speakers of languages that do not typically include endpoints.[6] Additionally, their duration of the fixation on the area of interest is also longer.

Another method to investigate the effect of language on cognitive perception assesses event-related potentials (ERP), a measure for electrophysiological responses of the brain to a sensory, cognitive or motor stimulus.[7] Boutonnet, Athanasopoulos, and Thierry (2012) used ERP to look at effects of grammatical constructs (in their case grammatical gender) on categorical perception. When Spanish-English bilinguals were asked to perform a semantic categorization of three pictures, they showed a greater modulation of an ERP marker of morphosyntactic processing than English monolinguals. This effect was found despite the fact that the study was conducted in English. Hence, Spanish grammatical gender categories weren't needed to solve the task. These findings demonstrate that speakers of languages with a grammatical gender unconsciously retrieve these information even when they are not needed to categorise objects.

Additional methods to measure the effect of language on cognitive perception are verbal interference tasks, lateralisation studies and studies using functional magnetic resonance imaging (fMRI). A verbal interference task is a dual task methodology, whereby participants have to perform a verbal task, e.g., remembering and repeating digits, at the same time as a nonverbal cognitive task. This limits the reliance on verbal resources

[6] The inclusion of endpoints is a difference between [+aspect] and [-aspect] languages (Athanasopoulos and Bylund, 2013; Bylund, Athanasopoulos, and Oostendorp, 2013; Stutterheim and Nüse, 2003).

[7] Nancy Kanwisher gives an easy to understand introduction into ERPs in her undergraduate course The Human Brain, which can be viewed on the MITCBMM YouTube channel: 2.9 - Event-Related Potentials (ERPs).

for the purpose of performing categorization tasks. Therefore, these studies demonstrated that humans use linguistic resources in decision-making even when they are faced with non-verbal stimuli. In most individuals, language processing involves the left hemisphere of the brain to a greater degree. Lateralisation studies use this knowledge about language processing and show visual stimuli to participants either in their right field of view, which is also processed in the left hemisphere of the brain, or in the left field of view, which is processed in the right hemisphere of the brain. They show that visual information in the right field of view, which is first processed by the language-dominant left hemisphere, is affected by language to a greater degree. fMRI provides measurements for activity in the brain that are based on the oxygen level of the blood. Active neurons, similar to muscle cells, need more oxygen and this rise in oxygenation of the blood can be detected by the scanner. Similar to ERP, fMRI can therefore detect neuron clusters in the brain, which are activated by stimuli.[8,9]

In recent years, cross-linguistic differences have also drawn the attention of economists as possible determinants of economic outcomes. Contrary to the work by linguist, psychologists and cognitive scientist studies by economist normally include a much larger sample size and they investigate the effect of language structures on economic outcomes and people's beliefs instead of cognitive effects. So far, existing research has mainly focused on four language features: tense (M. K. Chen, 2013; S. Chen et al., 2017; Chi et al., 2020), pronouns (Alesina and Giuliano, 2010; Davis and Abdurazokzoda, 2016), gender (Santacreu-Vasut, Shoham, and Gay, 2013; Hicks, Santacreu-Vasut, and Shoham, 2015; Gay et al., 2018) and mood (Kovacic and Orso, 2016; Bernhofer, Costantini, and Kovacic, 2021).[10]

[8]Nancy Kanwisher gives an easy to understand introduction into fMRIs in her undergraduate course The Human Brain, which can be viewed on the MITCBMM YouTube channel: 2.11 - fMRI.

[9]See Athanasopoulos and Casaponsa (2020) for an detailed overview of the literature about neuroscientific approaches to linguistic relativity.

[10]See Mavisakalyan and Weber (2018) for a more comprehensive overview of the literature on linguistic structures and economic outcomes.

1.2 Thesis Outline

This thesis contributes to the growing strand of literature investigating the influence of language characteristics on economic outcomes. I use cross-linguistic differences and survey data in my thesis to investigate the effects of certain language characteristics on the behaviour of individuals and their economic outcomes.

In the German language, it is possible to talk about future events in the present tense, whereas English requires its speakers to use the future tense to talk about future events. This difference in future time reference is found in many languages around the globe. Recent research has shown that the grammatical representation of the future in languages influences how future-oriented an individual acts in their private life. In **Chapter 2**, I combine data from the World Atlas of Language Structure (WALS), BoardEX and Amadeus to examine whether this effect of a language on a person's behaviour is also evident in the business world. Do companies with board members whose native language grammatically separates the present from the future invest less into the future of their business model and therefore act less future-oriented? I find evidence that inflectional marking of the future in languages has an effect on the future orientation of companies, which confirms that the effect of language on future orientation can also be found in the business world.

Individuals and societies differ in their attitudes toward foreigners. Some societies are considered very immigration-friendly while others are known for their more hostile attitude towards immigration. Previous research concerned with attitudes towards foreigners and immigration focused primarily on economic, socio-economic, and cultural variables to explain these differences. In **Chapter 3**, I use cross-linguistic differences in how languages distinguish different politeness groups in their second person pronouns to add language differences as another explanatory variable to this body of literature. Using data from the World Value Survey (WVS) and the WALS, I find that individuals who speak a language without politeness distinctions have a higher probability to respond that they trust foreigners, which is used as a measurement of an individual's attitude towards foreigners.

Another domain in which languages differ, is how they categorize objects. Many languages assign a gender to objects and in many cases this grammatical gender system is based on biological sex whereas other languages have a non-sex-based gender system or assign no gender at all. In **Chapter 4** I use data from the WALS about the grammatical gender system of languages and combine it with data from the European Social Survey (ESS) to investigate the effect of language on gender norms regarding the role of women in the labour market and labour market outcomes for women. The WALS contains four different features of languages concerned with gender and gender assignment in languages. I find that a gender system based on biological sex in a language reinforces traditional gender norms regarding women's roles in family and work life. Furthermore, it is also associated with worse labour market outcomes for women. Women who speak such language reduce their labour supply, either entirely or by reducing their working hours, and work in occupations traditionally considered more feminine.

Chapter 2

Language and the Future: Board Members and the Investment in the Future

In this chapter, I investigate differences in how languages grammatically express the future affect the future-orientation of companies. For my analysis I use two different language features related to the future; future-time reference and the inflectional marking of the future. I combine these with data from BoardEx about board members of companies and company information from Amadeus. A higher number of board members speaking a language with an inflectional marking of the future results in a lower intangible asset growth rate. On the other hand, however, it also leads to higher expenditure on R&D expenditures. The different effects can possibly be attributed to poor R&D data availability in Amadeus. A higher number of board members speaking a language with a strong-future time reference is associated with a higher intangible asset growth rate and more R&D expenditures. These results are not in line with previous research associating strong-future time reference with less future-orientation by individuals and companies. Linguists' critic that the future-time reference classification of languages is problematic is supported to some extent by these results. Nevertheless, I find evidence for the hypothesis that the future-orientation of a company is influenced by the language spoken by its board members.

2.1 Introduction

There are about 7000 languages worldwide. These languages differ in how they grammatically treat the future. In German it is perfectly fine to say *Es regnet morgen* (It rains tomorrow). In English on the other hand it sounds odd to talk in the present tense about future events. It is more natural to use *will* or *is going to* to mark the future (It will/is going to rain tomorrow.). In German, the present tense can thus be used to talk about the present and the future (weak future-time reference), while English speakers are required to make a distinction between the two (strong future-time reference). In this chapter I find that such an difference in how languages grammatically handle the future has an effect on the behaviour of companies. Companies with a majority of board members who speak a language which has an obligatory grammatical distinction between future and present act less future-oriented. These results extend the existing literature on factors that determine the level of R&D expenditure in companies and thus also help to explain the different long-term trends in productivity in countries.

Recent research by M. K. Chen (2013) and Roberts, Winters, and M. K. Chen (2015) showed that this is not only of interest for linguists but also affects the decision making of individuals. Chen demonstrates in his paper that individuals with a native language with no required distinction between present and future act more future oriented: They save more, retire with more wealth, smoke less, practice safer sex, and are less obese. His findings hold up both across countries and within countries. In this article, I examine whether the native language of an individual and its grammatical identification of the future form also influence the business world and the decisions in it. More precisely, do companies with a management board that consists of members who speak a language related to more future orientation, also act more future-oriented?

A possible measurement for the future-orientation of a company is its investment into R&D. Does a company invest more into R&D when its board is composed of for example more German speakers than English speakers? This is not only an important question for companies and their owners, who look for long-term success but it is also of interest for the economy as a whole. Economic theory points to R&D as a main source of long term productivity growth (Solow, 1956; Romer, 1990). Therefore, the amount of R&D expenditures are important for the long term development of economies. But

companies might not undertake research on their own but outsource it to suppliers. Other companies might not undertake research at all but buy market ready products or even whole companies to develop their business model for the future. To capture this kind of future-oriented behaviour of a company I also look at its intangible assets growth rate.

To address my hypothesis, I use data from BoardEx about board members of European companies from 2005 to 2017. The native language of the board members is approximated by their nationality. This information is then combined with data about R&D expenditures of the companies from the Amadeus database of the Bureau van Djik. The information about the future-time reference (FTR) of a language, which tells you how and when a language requires the speaker to mark the future, come from the WALS (Dahl and Velupillai, 2013) and M. K. Chen (2013). This provides me with a comprehensive panel data set containing observations from 1895 companies over a 13-year period to examine the relationship between FTR characteristics of board members' native languages and companies' R&D expenditures.

I find a negative effect of inflectional marking on the growth rate of intangible assets, so companies with more board members who speak a language that grammatically distinguishes the future from the present by modifying the verb do act less future-oriented. When I use R&D expenditures as benchmark for the future-oriented behaviour of a company the effect turns positive. However, this change in direction of the effect most likely doesn't reflect the actual effect, but is due to the poor availability of data on R&D expenditure of enterprises in the EU in my data sources. The effect of the Strong- and Weak-FTR classification is positive throughout all of my regressions and therefore not in line with my hypothesis and the findings from the literature. But due to the criticism of various linguist (Section 2.7) I would argue that this classification is connected with many problems and is not suitable to differentiate languages. Therefore, I do not put much emphasis on these results and focus on the inflectional marking.

Closest to my own research is work done by Chi et al. (2020). They find in their work that countries with a weak FTR language and companies from such countries invest more into R&D. My research adds to the existing literature by not assigning companies the characteristic of the official language of the country they are based in but to focus

directly on board members and their native language. Far-reaching decisions for the future of a company, like R&D investments, are made by its board. So to evaluated a possible effect of a language characteristic on the future-orientation of a company, you have to look at the language of its board members, because they are the individuals who make these decisions and their thinking and ultimately their decision making is influence by their native language. A second important point is, that the current literature on the effect of FTR on economic decision making uses the classification in weak and strong FTR first introduced into economic research by M. K. Chen (2013). As mentioned before the linguists Pullum (2012), and Dahl in his comment to a guest post by M. K. Chen (2012) criticise that an accurate classification of languages into strong or weak FTR is not so easily possible and that languages differ in their FTR marking on many parameters for which information is often lacking in grammar. The WALS Dahl and Velupillai (2013) focus on inflectional marking as one characteristic in which languages differ in their marking of future events. I use this characteristic as a second variable to investigate the effect of FTR in languages on the economic behaviour and future-orientation.

In a follow up study, Roberts, Winters, and M. K. Chen (2015) investigate if the findings of M. K. Chen (2013) for savings and grammatical marking of the future are robust when controlled for geographic and historical relatedness of languages. In general, the statistical correlation between the two variables is weaker when controlled for relatedness but the correlation remained reasonably robust. Fuchs-Schündeln, Masella, and Paule-Paludkiewics (2020) and Guin (2016) also find that a weak-FTR language leads to a higher savings of individuals. S. Chen et al. (2017) find a similar savings behaviour for companies. The relationship between FTR and earnings management is investigated by Fasan et al. (2016) and Jaehyeon Kim, Y. Kim, and Zhou (2017). They both find that companies from weak-FTR countries are less likely to engage in earnings management. Pérez and Tavits (2017) find that weak-FTR languages are linked to a higher support for future-orientated policies. They randomly assign the language to bilingual persons in a survey and find that individuals are more likely to support such policies if the survey is conducted in the weak-FTR language. Galor, Özak, and Sarid (2016) find that individuals speaking a language with an obligatory inflectional marking of

the future are 4 percentage points less likely to attend college.

In addition to my primary research question, this research also expands the already existing literature about factors determining the level of R&D expenditures within a country. Some of these factors are tax incentives, location factors, democratic institutions and compensation schemes for board members. A large strand of literature finds a positive effect of tax incentives on R&D activities (Hall and Van Reenen, 2000; Bloom, Griffith, and Van Reenen, 2002; Ernst and Spengel, 2011). However, tax incentives do not only affect the quantity of R&D but also the quality. Ernst, Richter, and Riedel (2014) use patent applications to the European Patent Office (EPO) as a proxy for R&D activities of companies. They show that a low tax rate on patent earnings raises the average profitability and innovation level of projects. On the other hand, R&D tax credits and allowances exert a negative impact on project quality. Important location factors for R&D activities are high-quality infrastructure and the supply of R&D staff (Cantwell and Piscitello, 2005). For public R&D expenditures, democratic institutions play a role (Jungbu Kim, 2011). On an individual level, Rapp, Schaller, and M. Wolff (2012) find that a share based compensation of board members yield higher investments into R&D.

This chapter is structured as follows. First, in section 2.2 I explain in more detail what FTR means and how it differs between languages. Section 2.3 details my hypothesis and how your native language might affect your decision making. Section 2.4 explains my variables of interest in more detail and which controls I will use for my analysis. Section 2.5 presents my empirical model and the results of my regressions are presented in section 2.6. Section 2.7 discusses issues surrounding the interpretation of my results and known criticism from the literature. In Section 2.8, I conclude my findings.

2.2 Future-Time Reference in Languages

Languages differ widely in how and when they require their users to mark the future. As mentioned above, an English speaker mostly uses some form of *will* or *going to* when they are speaking about the future. For example, if I want to tell a friend what I'm going to do tomorrow, I can't say *I go to the theatre*. In the English language it is obligatory

to say *I'm going to the theatre tomorrow*. In German it is perfectly fine to use the present tense to talk about future plans (**1**).

(1)	Morgen	regnet	es	
	Tomorrow	rains	it	
	It	will rain	tomorrow	

Just because there is no obligation to use the future tense in German doesn't mean there is no possibility to grammatically mark the future (**2**). It would also be totally fine to say:

(2)	Morgen	wird	es	regnen
	Tomorrow	will	it	rain
	It	will rain	tomorrow	

which is not so commonly used but also correct.

Even within Europe, these differences are surprisingly widespread. It ranges from Finnish (**3**) with almost no distinguishing between present and future time to French (**4**), which has separate and obligatory forms of verbs to use in the future tense.

(3)	Tänään	on	klymää	
	Today	is	cold	
	It	is	cold	today
	Huomenna	on	klymää	
	Tomorrow	is	cold	
	It	will be	cold	tomorrow

It also is equally correct to use the present tense for the present and the future in Finnish like it is in German as mentioned before. In English, it again requires this auxiliary construct with *will* to grammatically mark the future. Another way to grammatically mark the future which is commonly used in languages is inflectional marking. Inflectional in

general means the modification of a word to express different grammatical categories, i.e., tense, case or gender. Thus in French present and future differ in the form of the verb.

(4)	Il	fait	froid	aujourd'hui
	It	do	cold	today
	It	is	cold	today
	Il	fera	froid	demain
	It	do	cold	tomorrow
	It	will be	cold	tomorrow

I will use both differences in marking the future I just described in my analysis. The first characteristic is the obligatory marking of future events, which is the central characteristic of the Weak- and Strong-FTR classification. This is the criteria that Chen uses in his analysis. The second characteristic I want to exploit is inflectional marking of the future like in French. This means not only an obligatory marking of the future but it is also done by the modification of the verb and not like in English with an auxiliary construction. This is the feature which is used in the WALS in the chapter about the future tense (Dahl and Velupillai, 2013) and also by Galor, Özak, and Sarid (2016). I consider the inflecting marker to be the better distinction, as the distinction made by Chen is criticized by some linguists[1]. I have nevertheless included the Strong-FTR classification for completeness, as it is used in the literature.

2.3 Hypothesis

In this chapter, I test the hypothesis that being required to speak in a certain form about the future affects the decision making of an individual, i.e., they act less future-oriented. If a language requires the speaker to distinguish between present and future grammatically, the future will be conceptual more distant. This distance leads to a less future-oriented behaviour. In the business environment, investments in the future lead to costs today, but the possible rewards from it are sometime in the future. For a speaker

[1] For more information see Section 2.7

of a language with a certain grammatical marking of the future these possible rewards are even further in the future due to the grammatical distinction. Therefore, language speakers prefer to spend money on projects that yield a reward today. For speakers of Weak-FTR languages on the contrary, it might be easier to invest. By equating present and future grammatically, the future seems closer. Hence, the possible reward of an investment is mentally also closer, which makes it easier to bear the costs today.

For the Strong- and Weak-FTR distinction that M. K. Chen (2013) makes in his paper this hypothesis is straight forward. A person whose native language has a Strong-FTR time reference acts less future-oriented than a speaker of a Weak-FTR language. Therefore, a company with board members who speak more Strong FTR languages should also act less future-oriented if this effect is transferred to the business world.

For the second characteristic the hypothesis changes slightly and gets more specific. I argue that it does not only matter if a language requires a speaker to distinguish between future and present but also how it requires them to differentiate. It might make a difference if a speaker is required to use an auxiliary term to talk about the future like in English or they have to use a *special* form of the verb, like in French. To really alter the word of what you are doing, in this case spending money/bearing costs to get a reward in the future, might make mentally a huge difference. In doing so, it becomes really clear that the earning is in the future because the verb *earn* is in its future form. When an auxiliary construct is used the verb *earn* remains in its present form and thereby might not be perceived to be in the future. At least not as distant in the future compared to a language in which the speaker is required to use a future form of the verb itself.

2.4 Data

2.4.1 Dependent Variable

As a measurement for the future orientation of a company, I use tow different variables. The market and the demand of customers change over time, new technologies emerge and old ones disappear. Companies must constantly adapt and evolve their business model to stay relevant and competitive in a changing world. One way to measure

how future orientated a company acts, is their expenditures into R&D to develop their business model and to be prepared for the market of tomorrow. The data about R&D expenditures of companies come from the Amadeus database of the Bureau van Djik which has financial data for thousands of European firms. The data is available per company per year and denominated in local currency. To make it comparable between companies, I relate R&D expenditures to the total assets of the company. This results in a proportion, that shows how much of the company's current capital is invested into the future. Observations with negative R&D expenditures are excluded, because they can't be reasonably explained. Furthermore, observations with a proportion of R&D expenditures to total assets over 50% are excluded, because they seem unreliably high and are most likely reporting errors. Investment into R&D and development of new products for yourself is not the only way to prepare your company for the market of the future. Another possibility is to buy the research and developments of others. To capture this, I will also look at the growth rate of the intangible assets of the companies. As you can see in table 2.1, there are growth rates that seem to be unreliably high. Because of this observations above the 93rd percentile are excluded from the regressions. The 93rd percentile with a growth rate just over 100% is still pretty high, but the lowest value I can use without running into sample selection problems. I will talk more on sample selection problems in section 2.7.

TABLE 2.1: Distribution of Dependent Variables

R&D expenditures		Growth Rate of int. assets	
Minimum	-0.267	Minimum	-1
25th Percentile	0	25th Percentile	-0.068
50th Percentile	0.0142	50th Percentile	0.005
75th Percentile	0.063	75th Percentile	0.148
93rd Percentile	0.245	81st Percentile	0.248
97th Percentile	0.496	87th Percentile	0.460
98th Percentile	0.687	93rd Percentile	1.027
Maximum	140.080	Maximum	1.62×10^8
Mean	0.149	Mean	11631.830

For the R&D regressions I'm left with 4830 observations from 868 companies that range from 2005 to 2017. These companies are based in 10 different European countries and

come from 34 different sectors.[2] On average there are 5.57 observations per company. The data set for regressions with the intangible assets growth rate as dependent variable has 9968 observations from 1895 companies that range from the years 2007 to 2017. These companies are based in 24 different European countries and come from 37 different sectors.[3] On average there are 5.26 observations per company.

TABLE 2.2: Company Descriptives

	R&D Expenditures	Intangible Asset Growth Rate
Female Share	0.14	0.13
Avg. Age of Board Members	62.37	62.11
Country Share	0.85	0.83
Avg. R&D	0.05	
Avg. Growth Rate		0.01
Avg. Nb. Observation	5.56	5.26

2.4.2 Independent Variable

TABLE 2.3: Descriptive Statistics Strong-FTR: Mean

R&D Expenditures

		Mean	Variance	Min	Max	Observations
SFTR	overall	0.797	0.356	0	1	N = 4830
	between		0.365	0	1	n = 868
	within		0.050	0.462	1.464	T-bar = 5.565

Intangible Asset Growth Rate

		Mean	Variance	Min	Max	Observations
SFTR	overall	0.759	0.389	0	1	N = 9968
	between		0.398	0	1	n = 1895
	within		0.048	0.230	1.359	T-bar = 5.260

Note: T-bar is the average number of years observed for all companies

The data about Strong-FTR and Weak-FTR languages is adopted from M. K. Chen (2013). He bases his data mostly on the research of Dahl (2000) and Thieroff (2000)

[2]For a complete list of countries and sectors see table A.1 and table A.2 in the appendix.
[3]For a complete list of countries and sectors see table A.3 and table A.4 in the appendix.

about the characteristics of European languages and extends it with other sources for non-European languages. The information about the board members of a company in a certain year come from BoardEx. The data also ranges from 2005 to 2018. Both databases are connected by the nationality of the board member. The board member gets assigned the language characteristics of the official language of their nationality. The variable is 0 if the language is a Weak-FTR language, like German, and 1 if it is a Strong-FTR language, like English. Afterwards, the data for all board members per company and year are collapsed to form a continuous variable that ranges from 0 to 1. The observation takes the average value of all the language variables of the board members. That means if a company has 10 board members in 2010, of whom four are German speakers and six are English speakers, the value of the language variable would be 0.4 for this year. The descriptive statistics are shown in Table 2.3.

TABLE 2.4: Descriptive Statistics Strong-FTR: Median

R&D Expenditures

		Mean	Variance	Min	Max	Observations
SFTR	overall	0.808	0.390	0	1	N = 4830
	between		0.400	0	1	n = 868
	within		0.057	0.031	1.642	T-bar = 5.565

Intangible Asset Growth Rate

		Mean	Variance	Min	Max	Observations
SFTR	overall	0.764	0.421	0	1	N = 9968
	between		0.432	0	1	n = 1895
	within		0.057	-0.111	1.576	T-bar = 5.260

Note: T-bar is the average number of years observed for all companies

Another way to condense the language variable of board members per company is to use the median instead of the mean. You can possibly argue that decisions in a board are not made by a compromise that represents the preferences of all board members according to their share, but by a majority winner takes it all kind of vote and therefore the median language characteristic is of more interest than the mean value. The descriptive statistics for the median value of the language variable are presented in table 2.4. As one can see from the data, there is variance of the language characteristic between and

within companies. However, the variation within companies is rather small.

For my second approach, I will use data from the WALS (Dahl and Velupillai, 2013) about whether a language has an inflectional marking of the future. The data for the board members is the same as before and they are again linked by the nationality of the board members. The variable is 0 if the language has no inflectional marking and 1 if there is an inflectional marking of the future. The observations are again collapsed to get one observation per company and year which is the mean of all board members and a continuous variable between 0 and 1. Table 2.5 shows the descriptive statistics for the mean of the inflectional marking variable.

TABLE 2.5: Descriptive Statistics Inflectional Marking: Mean

R&D Expenditures

		Mean	Variance	Min	Max	Observations
Inflectionl	overall	0.403	0.447	0	1	N = 4830
Marking	between		0.442	0	1	n = 868
	within		0.049	-0.397	0.903	T-bar = 5.565

Intangible Asset Growth Rate

		Mean	Variance	Min	Max	Observations
Inflectionl	overall	0.275	0.411	0	1	N = 9968
Marking	between		0.407	0	1	n = 1895
	within		0.044	-0.141	0.859	T-bar = 5.260

Note: T-bar is the average number of years observed for all companies

Additionally, I will also use the median to condense the language variables of all board members to one value per company per year. There is again variance between the companies but very little within the companies over time. Descriptive statistics are shown in table 2.6.

2.4.3 Controls

As mentioned before, the quantity and quality of R&D activities of companies greatly depends on country specific factors, like taxation, infrastructure and institution. To control for these factors I will use country fixed effects for the general situation in a country, like the underlying institutions, that don't usually change on a yearly basis. To

TABLE 2.6: Descriptive Statistics Inflectional Marking: Median

R&D Expenditures

		Mean	Variance	Min	Max	Observations
Inflectionl	overall	0.421	0.490	0	1	N = 4830
Marking	between		0.480	0	1	n = 868
	within		0.062	-0.436	1.310	T-bar = 5.565

Intangible Asset Growth Rate

		Mean	Variance	Min	Max	Observations
Inflectionl	overall	0.278	0.445	0	1	N = 9968
Marking	between		0.439	0	1	n = 1895
	within		0.057	-0.558	1.153	T-bar = 5.260

Note: T-bar is the average number of years observed for all companies

capture possible tax incentives for R&D expenditures and the general economic state of a country, which can change more often, I will use an interaction term between the country and the year fixed effect.

A second effect I want to control for is the fact that different amounts of R&D activities between companies from different countries can just be rooted in different cultural preferences of its board members. It might be that some cultures are just more future oriented than others and therefore board members from these cultures do invest more into R&D than board members with a different cultural background. To test for a possible cultural effect, I will use two cultural dimensions from the Global Preferences Survey (GPS) (Falk et al., 2016; Falk et al., 2018). Investments into R&D are normally risky at least to some extent and have a potential return somewhere in the future. Therefore, I will include the GPS's measurements for *patience* and *risk preference*.

Different business sectors can greatly differ in their R&D activity. In pharmacy, R&D is a very important part of the business model. Companies have to constantly develop new drugs, which is a very costly process. In retail, on the other hand R&D plays nearly no significant role. I will use sector fixed effects to control for these fundamental differences between different sectors as I am not interested in the effect of the sector on the volume of the R&D expenditures.

To make investments a company first needs the financial capacities to make them, so the general financial situation could also play an important role for the R&D activities of a company. To address this issue I will also control for that by using the EBIT margin (ratio earnings before interest and taxes to operational revenue) and the EBITDA margin (ratio earnings before interest, taxes, depreciation and amortization to operational revenue) of the companies.

Languages can't be assumed to be independent from each other because they have common ancestors. As Roberts, Winters, and M. K. Chen (2015) argue in their paper, this can lead to an overestimation of the correlation (Galton's problem). Therefore, I will add dummies for the language families and language genera to control for possible effects of historical relatedness between languages. These dummies are again specific to an individual and then collapsed on the company level to form a continuous variable between 0 and 1 as the other language characteristics. For a company with a board consisting of five members of whom three are English speaking (Germanic language) and two are French speaking (Romance language) for example the variable for Germanic languages would take the value 0.6.

Another effect that should be taken into account is the effect of someone entering or leaving the board. New board members try to implement their ideas for the future of the company and therefore the R&D expenditures rises or they invest in new intangible assets. A change of board members might also be a sign for an overall poor financial situation of the company which leads to cuts in the R&D budget or the sale of intangible assets. Board members usually knew beforehand when their tenure ends and it could be that they do not want to make big decisions about the future of the company in their last year and leave it to their successors instead. I will control both, for new member entering the board and an old member leaving it because these two events do not have to occur at the same time. There might be some delays in finding a successor for a leaving member or there is always the possibility for an increase or decrease of the overall board size.

Country Share gives the share of board members with a nationality equal to the country the companies is based in. This share is 83% and 85% for the samples of companies used in the regressions of the intangible asset growth rate and the proportion of the

R&D expenditures, respectively.

2.5 Empirical Approach

2.5.1 Model

I will use the following equation for all my regressions:

$$y_{i,t} = \beta_0 + \beta_1 \ell_{i,t} + \beta_2 X_{i,t} + \beta_3 X_i + \epsilon \tag{2.1}$$

The dependent variable $y_{i,t}$ is either the R&D expenditures in relation to the amount of total assets or the intangible assets growth rate of company i in year t. The variable $\ell_{i,t}$ is the language variable I am interested in. In the first approach, this is the share of Strong-FTR language speaking board members in company i in year t. In the second approach, it is the share of board members speaking a language with an inflectional distinction between future and present in company i in year t. $X_{i,t}$ is a vector of company and time variant control variables and X_i is a vector of only company variant control variables.

I estimate a random effects model with standard errors clustered at the company level. As my dataset is unbalanced I will use the Swamy-Arora estimator of the variance components (Swamy and Arora, 1972; Baltagi, 2013).

2.5.2 Unbalanced Panel Data

Unfortunately I don't have observations for all companies for every year, so my dataset is unbalanced. This wouldn't be a problem if you could argue that the data is missing completely at random (MCAR), this means that the missing data are a totally random set of the data. If this were be the case, it would be possible to just use the same empirical methods as in the case of a balanced panel dataset. But in the case of company you can easily argue that the data of some companies is missing because they were founded later, went bankrupt, merged with another company or were just too small to report any data. They are not missing completely at random but their missing is conditional on other variables, i.e., the economic situation in the market they are operating in or their size. A sample selection problem arises if this selection is related

to the idiosyncratic errors, even when controlled for the conditional explanatory variables. Wooldridge (2010) suggests a simple test to check if the selection is related to the idiosyncratic errors. A lead of the selection indicator, $s_{i,t+1}$ is added to the regression with all other explanatory variables. For observations that are in the sample every time period, $s_{i,t+1}$ is always zero. But for attriters, $s_{i,t+1}$ switches to one in the period just before attrition. Selection in the succeeding time period should not be significant in the equation at time t, when the idiosyncratic errors are uncorrelated to the selection.

TABLE 2.7: S-Test for Sample Selection Bias

	(1)	(2)	(3)	(4)	(5)	(6)
Strong-FTR	0.028		0.028	0.012		0.012
	(0.037)		(0.037)	(0.013)		(0.013)
Inflectional-FTR		-0.139	-0.139		0.298***	0.290***
		(0.156)	(0.157)		(0.106)	(0.105)
Attrition Control						
$s_{i,t+1}$	-0.025	-0.025	-0.025	0.000	-0.000	0.000
	(0.017)	(0.017)	(0.017)	(0.005)	(0.005)	(0.005)
Country	√	√	√	√	√	√
Country × *Year*	√	√	√	√	√	√
Language Family	√	√	√	√	√	√
Language Genus	√	√	√	√	√	√
Sector Fixed Effects	√	√	√	√	√	√
Sector × *Year*	√	√	√	√	√	√
Observations	9968	9968	9968	4830	4830	4830

Note: All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Growth rates in the regressions are restricted to be smaller than the 93rd percentile. Negative proportions and proportions over 0.5 are excluded from the regression for R&D Expenditures. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Table 2.7 show the results for these tests. In all four regressions the selection indicator is statistically not significant and therefore there is no sample selection bias. Consequently, the same methods can be used as in the case of MCAR data.

2.6 Main Results

2.6.1 Intangible Asset Growth Rate

Tables 2.8 and 2.9 show the results for the regressions with the intangible assets growth rate as dependent variable and Strong-FTR and Inflectional-FTR as variables of interest respectively. Table 2.8 uses the mean value for the language characteristics and table 2.9 the median. The control variables are introduced in the same order in both regressions. Regression one controls only for country fixed effects and the interaction term between country fixed effects and year fixed effects. In the second regression, I additionally add variables for language family and genus to control for possible effects of relatedness between languages. Regression three adds the GPS's values for *patience* and *risk preference*. In regressions four to six, I add controls for the financial situation, sector fixed effects, interaction terms between sector fixed effects and year fixed effects, the share of board members being nationals of the country the company is located in, the average age of the board, the share of female board members, the share of new board members, the share of board members who left the board that year and the amount of total assets a company has. In regression seven I add additionally the growth rate of the next period as a control variable. I drop all observations above the 93rd percentile of the growth rate. Therefore, my sample selection is conditional on the intangible asset growth rate. To prevent a possible sample selection bias this conditionality must be taken into account.

The coefficient of the Strong-FTR characteristic is not significant in any of the seven regressions, but it is also positive in all regressions in contrast to my predictions. Inflectional distinction between future and present as variable of interest is also not significant in any of the regressions. Its coefficient is negative in all regressions as expected. The coefficient for the share of members leaving the board is significant at 10% level and yields a negative effect of about 0.05. The size of the company is significant at the 1% level. But the coefficient is very small.

In table 2.9 I use the median value of the Strong-FTR and the Inflection-FTR characteristic as independent variable. The coefficient for Strong-FTR is significant and positive from regression two onwards. So a switch from a board with a majority of Weak-FTR

TABLE 2.8: Effect on Intangible Asset Growth Rate: Mean

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	-0.002	0.034	0.038	0.038	0.038	0.040	0.041	0.028
	(0.030)	(0.032)	(0.032)	(0.032)	(0.032)	(0.032)	(0.034)	(0.037)
Inflectional-FTR	-0.047	-0.058	-0.129	-0.133	-0.130	-0.127	-0.126	-0.143
	(0.032)	(0.165)	(0.154)	(0.155)	(0.154)	(0.153)	(0.154)	(0.155)
Financials								
EBIT Margin			0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share			-0.015	-0.015	-0.016	-0.014	-0.014	-0.012
			(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)
Mean Age				-0.000	-0.001	-0.001	-0.001	-0.001
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female Share				0.023	0.025	0.021	0.021	0.021
				(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Enter					-0.030	-0.030	-0.030	-0.030
					(0.021)	(0.020)	(0.020)	(0.020)
Exit					-0.048*	-0.047*	-0.047*	-0.047*
					(0.025)	(0.025)	(0.025)	(0.025)
Total Assets						0.000***	0.000***	0.000***
						(0.000)	(0.000)	(0.000)
Culture								
Patience							0.002	-0.040
							(0.051)	(0.069)
Risk Taking								0.122
								(0.159)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	9968	9968	9968	9968	9968	9968	9968	9968
Clusters	1895	1895	1895	1895	1895	1895	1895	1895

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Growth rates in the regressions are restricted to be smaller than the 93rd percentile.

language speaker to a Strong-FTR language speaking majority in regression eight is accompanied by a 5.5 percentage points higher intangible asset growth rate. This is again in contrast to what I expected and to Chen's findings. In contrast the coefficient for the Inflection-FTR is significant in all regressions an is always negative. A change from a board with a majority of speakers of a language with no inflectional distinction to a board with a majority of speakers of a language with inflectional distinction yields 10

TABLE 2.9: Effect on Intangible Asset Growth Rate: Median

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	0.028	0.058**	0.059**	0.059*	0.059**	0.061**	0.061**	0.055*
	(0.027)	(0.029)	(0.030)	(0.030)	(0.030)	(0.030)	(0.031)	(0.032)
Inflectional-FTR	-0.077**	-0.102**	-0.107**	-0.105**	-0.103**	-0.102**	-0.103**	-0.102**
	(0.030)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
Financials								
EBIT Margin			0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share			-0.012	-0.012	-0.012	-0.011	-0.010	-0.008
			(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)
Mean Age				-0.000	-0.001	-0.001	-0.001	-0.001
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female Share				0.022	0.025	0.021	0.021	0.020
				(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Enter					-0.030	-0.030	-0.030	-0.030
					(0.020)	(0.020)	(0.020)	(0.020)
Exit					-0.048*	-0.047*	-0.047*	-0.047*
					(0.025)	(0.025)	(0.025)	(0.025)
Total Assets						0.000***	0.000***	0.000***
						(0.000)	(0.000)	(0.000)
Culture								
Patience							0.005	-0.029
							(0.050)	(0.063)
Risk Taking								0.108
								(0.149)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	9968	9968	9968	9968	9968	9968	9968	9968
Clusters	1895	1895	1895	1895	1895	1895	1895	1895

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Growth rates in the regressions are restricted to be smaller than the 93rd percentile.

percentage points lower intangible asset growth rate. This is in line with my hypothesis. As in the regressions before the coefficient for the share of members leaving the board yield a significant negative effect and the coefficients for company's size is highly significant but very close zero.

Overall there are only significant results when I use the median of the language characteristic of the board members instead of the mean. Here the Strong-FTR characteristic has a positive effect, e.g. companies with a majority of board members speaking a Strong-FTR language act more future orientated. This effects contradicts my hypothesis and is also not in line with Chen's findings of the effect on an individuals behaviour. The effect for the Inflectional-FTR variable is as expected by my hypothesis. A company with a board with a majority of speakers of a language with inflectional distinction has a lower intangible asset growth rate, i.e., acts less future orientated.

2.6.2 R&D Expenditures

In my second approach, I use the proportion of R&D expenditures to total assets as dependent variable. The control variables are added in the same order throughout regressions one to seven as before. I drop all observations with a negative proportion or a proportion higher than 0.5. Therefore, my sample selection is conditional on the proportion of R&D expenditures to total assets. To prevent a possible sample selection bias this conditionality must be taken into account. As in the regressions for the growth rate I add the proportion of the next period to control for this conditionality.

The effect of the Strong-FTR characteristic isn't significant in all regressions, but is positive throughout all of them. The coefficient for the Inflectional-FTR characteristic is significant in all regressions. Its effect is negative and turns positive from regression two onwards. In regression eight it yields a positive effect of 0.288 that is significant at the 1% level. If a company's board switches from only speakers of a language with no inflectional distinction to a board with only speaker of a language with an inflectional distinction this will result in a 28 percentage points higher proportion of R&D expenditures to total assets. In regression eight only the controls for the financial situation of the company and the size of the company are significant. Both coefficients for the company's financials and the size are very close to zero.

In table 2.11 I use again the median value of the language characteristic instead of its mean. The coefficient for the inflectional marking of the future is only significant in the first regression. In all other regressions neither of the both language characteristics yields a significant effect. The cultural trait risk taking yields a positive effect of 0.17,

TABLE 2.10: Effect on R&D Expenditures: Mean

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	0.025**	0.028**	0.020*	0.019	0.019	0.018	0.024*	0.012
	(0.011)	(0.012)	(0.012)	(0.012)	(0.013)	(0.012)	(0.012)	(0.013)
Inflectional-FTR	-0.031***	0.134***	0.288***	0.292***	0.294***	0.301***	0.295***	0.288***
	(0.010)	(0.032)	(0.101)	(0.107)	(0.106)	(0.105)	(0.104)	(0.104)
Financials								
EBIT Margin			0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share				-0.002	-0.002	-0.002	-0.000	0.002
				(0.008)	(0.008)	(0.008)	(0.008)	(0.007)
Mean Age				-0.000	-0.000	-0.000	-0.000	-0.000
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female Share				-0.011	-0.012	-0.011	-0.011	-0.011
				(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Enter					0.002	0.002	0.002	0.002
					(0.006)	(0.006)	(0.006)	(0.006)
Exit					-0.009	-0.009	-0.009	-0.009
					(0.008)	(0.008)	(0.008)	(0.008)
Total Assets						-0.000**	-0.000**	-0.000**
						(0.000)	(0.000)	(0.000)
Culture								
Patience							0.063**	0.024
							(0.030)	(0.045)
Risk Taking								0.138
								(0.098)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	4830	4830	4830	4830	4830	4830	4830	4830
Clusters	868	868	868	868	868	868	868	868

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Negative proportions and proportions over 0.5 are excluded from the regression.

which is significant at the 10% level. From the other controls again the company's financials and size are highly significant but their effect size is very close zero.

Contrary to my findings from the regressions with the intangible assets growth rate the effect of the Inflectional-FTR characteristic of a language has a positive effect if I use the proportion of R&D expenditures to total assets as indicator of the future orientation

TABLE 2.11: Effect on R&D Expenditures: Median

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	0.017	0.018	0.013	0.012	0.012	0.012	0.014	0.004
	(0.011)	(0.011)	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)
Inflectional-FTR	-0.019**	-0.006	-0.005	-0.003	-0.003	-0.003	-0.003	0.000
	(0.009)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Financials								
EBIT Margin			0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share				-0.001	-0.001	-0.002	0.000	0.002
				(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Mean Age				-0.000	-0.000	-0.000	-0.000	-0.000
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female Share				-0.012	-0.012	-0.011	-0.011	-0.012
				(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Enter					0.002	0.002	0.002	0.002
					(0.006)	(0.006)	(0.006)	(0.006)
Exit					-0.009	-0.009	-0.009	-0.009
					(0.008)	(0.008)	(0.008)	(0.008)
Total Assets						-0.000**	-0.000**	-0.000**
						(0.000)	(0.000)	(0.000)
Culture								
Patience							0.056*	0.012
							(0.030)	(0.042)
Risk Taking								0.171*
								(0.093)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	4830	4830	4830	4830	4830	4830	4830	4830
Clusters	868	868	868	868	868	868	868	868

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Negative proportions and proportions over 0.5 are excluded from the regression.

of a company. A possible explanation for this change in the direction of the effect may be the different data that is available for the regressions with the R&D expenditures. The sample is not only much smaller, but also less diverse when it comes to countries where companies are located with a majority of board members who speak a language with the Inflectional FTR. Almost every one of them (98%) is located in France. This is of interest to my findings because, according to the World Bank's World Development

Indicators (The World Bank, n.d.), France has much higher R&D spending than, for example, Spain or Italy, where a significant proportion of companies are located in the data set that I use in the regressions with the growth rate of intangible assets. I would argue that, on average, companies from Spain or Italy also have a lower proportion of R&D spending than companies from France and I'm only left with companies with a high proportion. It is this enormous over-representation of French companies in my data that most likely drives my results and makes the effect positive compared to the results of previous regressions

2.7 Discussion

In his comment to M. K. Chen (2013), the British linguist Prof. Geoffrey K. Pullum criticizes Chen's coding of Strong- vs. Weak-FTR. He gives some simple examples that it is very well possible to speak about the future in present tense in English in certain circumstances, i.e. *My flight takes off at 8:30*. Therefore, he has no confidence in accurately describing English as Strong-FTR. Furthermore, he makes the point that if the facts are shaky for a so well studied language as English, how likely are they to be precise for less studied languages (Pullum, 2012)? A point that Dahl supports in his comment to Chen's answer to Pullum's critic (M. K. Chen, 2012). He says that in the EUROTYPE volumes (Dahl, 2000), one of the phenomena he is looking at, is the so-called *futureless area* in Northern Europe in which languages lack inflectional futures and future-time reference that is less systematically marked grammatically. He does therefore not introduce a binary coding like textitStrong- vs. Weak-FTR and focuses more on predictive statements and not obligatory marking in general. The FTR marking differs across languages on many parameters for which information is often lacking in grammar. Because of this, the chapter on future tense in the WALS (Dahl and Velupillai, 2013) focuses on inflectional marking, the second criteria I use in my analysis.

A second critic Pullum (2012) made, is the fact that a priori it is not clear if the correlation should be positive or negative, a point that Roberts, Winters, and M. K. Chen (2015) also briefly address in their follow-up study. You can easily argue that the grammatical distinction between future and present does not lead to thinking less about the future but instead to think more about it. If an individual has to use a specific grammatical

construct or a specific form of a word to speak about the future, the speaker has to pay more attention to the future and therefore might act more future orientated. This could be an explanation for my results on the inflectional distinction of future and present, which go against my hypothesis and are contrary to Chen's findings.[4]

With regard to the the interpretation of the results, it is important to mention that I can't completely rule out that language is reflecting deeper differences between individuals which drive the different behaviour instead of causing it. I try to rule out this possibility by including my control variables, especially the cultural preferences, to find a causal relationship. The introduction of the cultural variables has almost no effect on especially the significant coefficients of the language characteristics. If they both were markers for the same causal factor you would expect these two to interact more.

2.8 Conclusion

Overall, my findings on the influence of inflectional marking on the growth rate of intangible assets are consistent with my hypothesis and the literature. I find a negative effect of considerable size for both the mean and median aggregation of the language characteristic. These results are tantamount to a less future-oriented behaviour of companies with more board members who speak a language that grammatically distinguishes the future from the present by modifying the verb. The coefficient for the median regression is statistically significant at the 5% level.

When I use the proportion of R&D expenditures to total assets as a benchmark for the future-oriented behaviour of a company, the direction of the effect changes too positive. But this most likely does not reflect the actual effect but is due to the poor availability of data on R&D expenditure of enterprises in the EU in my data sources. Almost all companies with a majority of board members who speak a language that has an inflectional marking for the future are based in France. France spends more of its GDP on research and development than, for example, Spain or Italy, two other large European countries with a language that bears the inflection marking (The World Bank, n.d.). It is therefore highly likely that French companies also have higher R&D expenditure on average than Spanish or Italian companies. Therefore, I would argue that not only do

[4] Similar criticism is brought forward by Dahl (2009)

I lose half of the observations by moving from the growth rate of intangible assets to R&D expenditure, but I also lose disproportionately many observations with low R&D expenditure, which fully drives the effect. The effect of the Strong- and Weak-FTR classification is positive throughout all my regressions and therefore not in line with my hypothesis and the findings from the literature. My findings thus support the critical view expressed by some linguists and discussed in the 2.7 section. As Pullum (2012) argues the direction of the language effect is a priori not that clear and the distinction into the strong vs weak-FTR classification is in general not so clear cut. Therefore, I would not put too much emphasis on these results and concentrated more on the inflectional marking of future events.

In summary, the native language of person not only seems to play a role for decision making in her private life, but this effect also translate to the business world. The native language of its board members effects how future orientated a company acts.

Chapter 3

Language and Xenophobia: The Effect of Politeness Distinction in Pronouns on your Attitude towards Foreigners

In this chapter, I investigate cross-linguistic differences in politeness distinction in second person pronouns and their effect on an individual's attitude towards foreigners. I combine data from the World Value Survey, which is asking its participants about how much trust they put into foreigners with data from the World Atlas of Language Structure about the distinction of politeness groups in second person pronouns of languages. A politeness distinction in second person pronouns of a language results in its speakers trusting foreigners less compared to speakers of a language with no politeness distinction. In further regressions, I only include either immigrants or individuals not speaking the official language of their country of residence to increase the variation in my language variable. The effect of the language variable remains. Furthermore, the effect also remains after the inclusion of two different sets of cultural controls, indicating that it is a genuine language effect and not the representation of cross-cultural differences. Therefore, my results show that language structures should be included into research about people's attitude towards foreigners in addition to the economic and non-economic factors that have played a central role in the research so far.

3.1 Introduction

What factors determine people's attitude towards foreigners. Why are some countries in the world considered immigration countries and others not. This is one of the most exciting questions of our time. In the United States, the 2016 presidential election was won by a candidate who aggressively campaigned for a wall on the border with Mexico to limit immigration. In the same year, one of the slogans of the Leave campaign during the vote on Brexit was *take back control of our borders* to stop immigration from Eastern Europe. On the other hand, during the refugee crisis in 2015, Germany voluntarily opened its borders to people from Syria, thus coining the term *welcome culture*. Previous research has focused on non-economic factors, such as cultural values or political views, and economic factors, such as competition in the labour market, as explanations for the differences in people's attitudes towards foreigners. In my research I want to provide another approach to explain why people differ in their attitudes towards foreigners, namely their language. My research shows that the language someone speaks has an influence on their attitude towards foreigners.

About 7000 languages are spoken worldwide, which differ in many features. Germans for example divide people linguistically into two different groups. For family and close friends they use the personal *du* as a pronoun. But there is also the polite *Sie* for strangers or people of higher rank. In the German language, second person pronouns are thus divided into two groups of politeness. English does not know such a distinction, *you* is always used as a form of address completely independent of the social relationship between the speaker and the person addressed. Some languages even distinguish more than two groups of politeness. In Marathi[1], for example, there is a separate polite pronoun for priests. Other languages avoid pronouns completely as a sign of politeness towards the person addressed and use titles or kinship terms instead. My hypothesis is based on this distinction between languages that have a politeness distinction in their second person pronouns and those that have none. When strangers for one person linguistically belong to another group than family and close friends, i.e., people you trust, this has an effect on their behaviour. Every time you address these people, the language reminds you by using a different pronoun that they are strangers

[1] A language predominantly spoken in the state of Maharashtra, India.

who have not yet made it into the inner social circle. This leads to a more negative attitude towards foreigners than a speaker of a language that does not make a distinction between politeness groups. For these persons, foreigners are linguistically closer, since they are addressed in the same way as family members and close friends. This leads to a more friendly and welcoming attitude towards foreigners, as both groups are linguistically equal.

However, this linguistic effect should not only reflect cultural preferences of a society. It is an effect that exists in addition to culture and can theoretically go in a different direction than the effect of culture. This hypothesis is based on linguistic relativism (also Sapir-Whorf hypothesis) (Sapir, 1921; Slobin, 1996; Whorf, 1956), which in its weak form states that linguistic categories and usage influence thought and decisions. The linguistic differentiation of people into at least two groups, one that is close to us and one to which we are more distant, thus influences our behaviour and leads to a more adverse attitude towards foreigners.

To test my hypothesis I use a ordered probit model to empirically analyze an individuals attitude towards foreigners across languages, using data from the 6^{th} wave of the WVS, carried out from 2010 to 2014 (Inglehart et al., 2014). I'm interested in question V107, which asks participants directly how much trust they place into people of other nationalities. They can choose their answer from four categories with descending trust level from *Trust completley* to *Do not trust at all*. This is an advantage over indirect measurements of personal preferences such as voting or lobbying, as these are also influenced by preferences in other policy areas. A second advantage of the WVS data for my analysis is that it contains information about the language the participants speak at home. This gives me some national variations of the language feature due to immigrants and their descendants still speaking their mother tongue or the mother tongue of their parents at home, and for countries where several languages are spoken in everyday life. The information on the politeness distinction in second person pronouns of the different languages comes from the WALS (Helmbrecht, 2013). The WALS divides languages into four different groups based on whether they have a politeness distinction and if so, how many politeness groups they have. I reduce this to two groups, because for my research question I am only interested in whether there is a distinction

of politeness or not.

I find a negative effect of the politeness distinction on the degree to which respondents trust foreigners: The marginal effects show that people who speak a language with a politeness distinction are more likely to answer the question with *Do not at trust at all trust* or *Do not trust very much* than people whose language does not have a politeness distinction. In contrast, for the categories *Trust somewhat* and *Trust fully*, the marginal effects are positive. The size of the effects ranges from just under 2 to just under 8 percentage points. These effects are robust to a variety of control variables. When the country fixed effects are omitted to account for the effect of Hofstede's cultural dimensions (Hofstede, 2001), which have previously been found to affect attitudes toward foreigners (Leong and Ward, 2006), the results hold. Furthermore, I use two different subsamples to overcome the limitations of my original data set. The results of these additional regressions are consistent with previous results and support my original findings. In addition, I still repeat my regressions with a data set that uses cultural preferences for the GPS (Falk et al., 2016; Falk et al., 2018) instead of Hofstede's cultural dimensions as controls for cultural idiosyncrasies. These regressions also confirm my results.

There is a growing literature on the effect of language structures on the behaviour of individuals. Closest to my on work is research by Kovacic and Orso (2016) about the effect of the number of grammatical categories concerned with the expression of uncertainty on the attitude towards immigration. People who speak a language in which these specific grammatical forms are used more intensively have a higher intolerance towards immigration. M. K. Chen (2013) and Roberts, Winters, and M. K. Chen (2015) have shown that people who speak a language without necessary distinction between present and future act more future-oriented. They save more, retire with more assets, smoke less, practice safer sex and are less obese. Similar savings behaviour is also found for corporations (S. Chen et al., 2017). Jaehyeon Kim, Y. Kim, and Zhou (2017) find that in countries where langauges do not require speakers to grammatically mark future events managers are less likely to engage in earnings management as future consequences of it are perceived more imminent. Chi et al. (2020) demonstrate that languages with a more ambiguously encoding of future timing lead to higher R&D investments

on the country- and firm-level. Galor, Özak, and Sarid (2020) investigate the effect of language characteristics on educational attainment. The presence of a periphrastic future tense has an positive impact on educational attainment whereas the the presence of sex-based grammatical gender system has a negative effect on female educational attainment. Hicks, Santacreu-Vasut, and Shoham (2015) and Santacreu-Vasut, Shoham, and Gay (2013) also show that gender specific linguistic characteristics are associated with worse outcomes for women regarding the allocation of household tasks and the implementation of gender political quotas, respectively.

There is an extensive literature which theoretical and empirically investigates factors, like age, political views, education, employment status, skill composition of the labour market and cultural values, influencing personal attitudes towards foreigners. Leong and Ward (2006) examine the influence of cultural characteristics of societies and their impact on attitudes toward immigrants and multiculturalism in Europe. They conclude that certain cultural traits are associated with lower support for policies that promote social coexistence and lead to more pessimistic attitudes toward multiculturalism. Hjerm (1998) uses data from the International Social Survey Program (ISSP) 1995 to examine the effect of national attachment on xenophobia in four European countries. His findings conclude that civic national identity and national pride lead to lower levels of xenophobia whereas on the other hand ethnic national identity and national pride lead to higher levels of xenophobia. Gang, Rivera-Batiz, and Yun (2013) show in their paper that the change in the attitude of European citizens towards foreigners between 1988 and 2008, which was found by Eurobarometer surveys, can be explained by racial prejudice, economic conditions and educational attainment. Racial prejudice and economic strain leads to more negative attitudes while on the other hand educational attainment act as a powerful antidote against anti-foreigner attitudes. Ostapczuk, Musch, and Moshagen (2009) test the hypothesis that the positive effect of a respondent's education on their attitudes towards foreigners is not because highly educated people are actually less xenophobic, but because they are simply more likely to give socially desirable answers. They do indeed find a strong bias in self-reported attitudes towards foreigners, but even after controlling for social desirability, an effect of education on attitudes towards foreigners can be found. In another paper, it is shown that the skill

composition of natives relative to immigrants has an effect on attitudes toward immigrants (Mayda, 2006). Skilled individuals are more in favour of immigration in countries where natives are more skilled than immigrants and opposed otherwise. Facchini, Mayda, and P. Mishra (2011) come to the same conclusion, that skilled natives are less likely to favour skilled migration due to the perceived competition threat on the labour market. This effect leads to lower number of policies aiming to increase the intake of skilled immigrants despite the benefits this kind of immigration can have on the destination country. Scheve and Slaughter (2001) on the other hand show for the US, that less-skilled workers have a higher preference for policy which is limiting inflow of immigrants into the US. Individuals believe that the US economy is absorbing the influx of immigrants, at least in part, through changing wages. Facchini and Mayda (2012) find that interest groups play a statistically significant role in shaping migration policy for different sectors in the US. Sectors where trade unions are more important tend to have higher barriers to migration, while sectors with stronger business interest groups have lower barriers. In a comparative study on the public views regarding the equality of rights foreigners deserve between Germany and Israel Raijman, Semyonov, and P. Schmidt (2003) identify the perceived level of threat as the main determinant of support for foreigners' rights.

The outline of this chapter is as follows. Section 3.2 describes the politeness distinction in second person pronouns and how it differs across languages. Section 3.3 explains my hypothesis on how your language might influence your attitude towards foreigners. Section 3.4 describes my underlying data. Section 3.5 focuses on my model. The results of my regressions are presented in Section 3.6. In section 3.7, the results of several robustness tests are reported. In Section 3.8, I conclude my findings.

3.2 Politness Distinction

The language characteristic of interest for my research is the politeness distinctions in personal pronouns, and to be more precises in second person pronouns. Before I get into this characteristic of a language in more detail I would like to start with a small exmaple from my native language German. German has a binary politeness distinction. There are *du* (you.sg.familiar) and *ihr* (you.pl.familiar) as intimate or familiar pronouns

to address someone and *Sie* (you.honorific) as a formal pronoun of address, which does not distinguish numbers. The formal pronoun is normally used between adults, who are not in a close social relationship like family or friends. Normally the usage of pronouns is symmetrical, so if you are addressed with the familiar *du* you will answer with it. One common exception is between adults and children. Adults usually address all children with *du*, but receive a *Sie* if they are not in a close social relationship with the child. This politeness distinction and its symmetrical use leads in Germany to the custom of offering someone the *du*, when the relationship has grown closer. In the most cases the offer is initiated by the older person or the one with the higher status, for example in an work environment.

The WALS (Helmbrecht, 2013) distinguishes four different forms of politeness distinction in second person pronouns. The first one is very simple, there is no politeness distinction, so these languages have no personal pronouns which express different degrees of respect or intimacy toward the addressee. One well known example is the English language, which only uses *you* as second person pronoun to address someone.

The next group are languages with a binary politeness distinction. Languages in this group have a clear contrast between a pronoun that is a polite form of address and a familiar pronoun. This binary distinction may well be expressed by several distinct pronouns as long as these pronouns do not indicate more than one politeness distinction. One example is Polish, which uses two different pronouns to indicate the same degree of respect in different dialects. *Wy* is used in rural areas and *Pan/Pani* is used in urban areas. The language Taba, which is spoken in Indonesia, has the pronoun *meu* (2.SG.HON), which is a free pronoun fulfilling all grammatical functions, and the pronoun *h*= (2.SG.HON), which is a obligatory clitic only in subject form. Clitics have the form of affixes, but play a syntactic role at the phrase level. A common example for a clitic is the contracted forms of the auxiliary verb in *I'm*. The binarity does not refer to the number of pronouns, but describes the fact that linguistically two groups are distinguished. A close group, which is addressed with the familiar pronoun, and a more distant group, which is addressed with the polite pronoun. Also the pragmatic rules when to use which pronoun can differ between languages with a binary politeness

distinction. The mother-in-law will be addressed with the familiar *du* by the daughter-in-law in German, whereas she continues to be addresse with the polite *vous* in French.

The third group includes all languages that have two or more degrees of politeness within a pronominal paradigm. These systems are rare. One example is Marathi, which distinguishes between *tu*, used for familiy and intimate persons, *te* and *he* (2.SG.HON), used for people with higher social status, and *āpan* (2.SG.HON), used for priests and teacher in a very formal context.

The last group *pronoun avoidance* is rather different from the first three. Polite forms of address in these languages do not belong to the class of pronouns. Instead, status and kinship terms, titles and other complex nominal expressions are used. If there are second person pronouns they are usually used to address social equals or inferiors. This strategy can be found in languages of East and Southeast Asia such as Japanese, Burmese or Thai.

Brown and Gilman (1960) explain the usage of familiar and polite pronouns by two parameters, which are not fully independent. The first one is power. In this case the polite pronouns are used to express a difference in social rank between the interlocutors. The asymmetrical use of *Sie* and *du* between adults and children in German for example reflects this difference in social power. The other parameter is solidarity. This reflects the social distance between the interlocutors. If your conversation partner is a stranger, your social distance to him is greater and polite pronouns are used. On the other hand, the social distance to your family members or friends is very small and the familiar pronoun is used. The use of solidarity pronouns is always symmetrical.

3.3 Hypothesis

Persons speaking a language with a politeness distinction divided people linguistically into at least two groups. One that is close to them and one that is further away (The solidarity parameter Brown and Gilman (1960) mention) . I argue that this linguistic separation influences the way people think and behave towards foreigners. One possible explanation for how friendly and hospitable a society and its members are towards foreigners is therefore the language they speak and whether or not there is a difference

in politeness. For example, an English-speaking person addresses everyone with a *you*, whether it is a family member, a close friend or a complete stranger. This brings all these different people closer together linguistically and thus influences the person's behaviour. If a person is addressed in the same way as a family member, they are treated more like a family member, which leads to a friendlier and more welcoming attitude towards foreigners. A German-speaking person, on the other hand, makes a clear linguistic distinction between close persons and foreigners and is therefore reminded each time they are addressed whether they are a close or a more distant person. If they are reminded each time through their language, this distinction will also be reflected in their behaviour towards the foreign person.

This language effect should not only reflect the cultural differences between societies, but should also be a pure language effect on the behaviour of the individual. One explanation for such a language effect is the theory of linguistic relativity. In its weak form this theory states that linguistic categories and language use influence the thinking and decisions of individuals (Sapir, 1921; Slobin, 1996; Whorf, 1956). Therefore, a possible distinction in second person pronouns for different groups of people can influence the behaviour of an individual towards these groups beyond his personal and cultural preferences. This effect is reflected in people who speak a language with a difference in politeness, in a different attitude and behaviour towards foreigners.

3.4 Data

For my analysis I use three main data sources. WVS Wave 6 (2010-2014) (Inglehart et al., 2014) for information about attitude, socio-econmic status, world view ect, the WALS (Helmbrecht, 2013), which contains a multitude of grammatical and lexical characteristics of thousands of languages, among others politeness distinction in second person pronouns and Hofstede's cultrual dimensions (Hofstede, 2001). The survey data and the language data is combined via question V247 *What language do you normally speak at home?* in the WVS. So, individuals get attributed the value for politeness distinction of the language that is used in their household. They also get the cultural dimension of the country they are currently living in, as it is not possible to track where they were born or raised. After combining my data I delete all observations with missing data

in the politeness distinction variable and the cultural dimensions. Afterward I impute the data of all missing answers to survey questions using chained imputation with 35 iterations. I don't impute values for politeness distinction of missing languages and missing cultural dimensions as I'm not really convinced myself that one could retrieve plausible values for those by looking at answers individuals have given to survey questions about their lives or by looking at other languages or the culture of other nations. After the imputation I have 41,152 unique observations from 32 countries.[2]

3.4.1 Dependent Variable

As proxy for xenophobia and an individuals attitude towards foreigners I use question V107 from the WVS Wave 6 (2010-2014). In Question V107 participants are asked: *I'd like to ask you how much you trust people from various groups. People of another nationality.*. Respondents are given four possible answers with decreasing trust levels: *Trust completely*, *Trust somewhat*, *Do not trust at all* and *Do not trust very much*. You would expect that people with a more adverse attitude towards foreigners will show lower trust levels towards people with a different nationality. As you can see in table 3.1, the majority of responses are in the two middle categories, with just over 35% each. Approximately 24% answered *trust completely* and only about 4% have no trust at all in people of another nationality.

TABLE 3.1: Trust in Foreigners

Trust completely	Trust somewhat	Do not trust at all	Do not trust very much	Missing	Total
8,545	13,059	12,632	1,546	5,370	41,152

Note: Question V107 WVS Wave 6: I'd like to ask you how much you trust people from various groups. People of another nationality.

3.4.2 Independent Variable

The information about the politeness distinction in second person pronouns in languages is taken from chapter 45A of the WALS (Dryer and Haspelmath, 2013). The authors divide languages into four categories depending on how many different groups can be addressed based on politeness reasons. The four different groups are explained

[2]For a complete list of countries see appendix table 3.3.

in more detail in section 3.2 (Helmbrecht, 2013). For my analysis I encoded these four categories into a binary variable taking the value 0 if a language as no politeness distinction in second person pronouns at all and 1 if a language has any kind of distinction.[3] As I have no hypothesis why it should matter for an individual's attitude towards foreigners if their language differentiates only two politeness groups or more or expresses politeness by pronoun avoidance, *binary distinction*, *2 or more* and *pronoun avoidance* are combined in the value distinction. Slightly over 80% of the individuals in my sample speak a language with some sort of politeness distinction as shown in table 3.2.

TABLE 3.2: Politeness Distinction Across Individuals

No distinction	Distinction	Total
7,391	33,761	41,152

Note: Distribution of politeness distinction across all observed individuals. The politeness distinction variable always refers to the language that the respondents reported as the language they speak at home.

As you can see from table 3.3 the variation of the politeness variable within the countries is rather low, most of the means are very close to 0 or 1. This is not surprising, since in most countries one language is dominant and spoken by the majority of the inhabitants. Furthermore, even in countries with many regional languages like India these languages are highly related because of close geographical proximity and common ancestors and therefore, share the same rules or very similar rules for politeness distinction in second person pronouns. The only notable exception is Singapore with a mean of 0.5. The reasons for that is that almost all respondents from Singapore reported either English or Mandarin as their language spoken at home. Mandarin features a politeness distinction whereas English on the other hand has none.

From the WALS I also retrieve the family and the genus of each language. In general languages can not be assumed to be independent from each other, so I use their family and genus to control for possible correlations between languages, which are related.

[3] A list of all reported languages and their value for the politeness distinction variable can be found in the appendix table B.12

TABLE 3.3: Politeness Distinction by Country

Country	Mean	Std. Dev.	Freq.	Country	Mean	Std. Dev.	Freq.
Argentina	1	0	1021	Netherlands	0.995	0.067	1768
Australia	0.036	0.187	1434	New Zealand	0.013	0.113	776
Brazil	1	0	1486	Pakistan	0.961	0.194	847
Chile	1	0	1000	Peru	0.993	0.081	1210
China	1	0	2300	Philippines	1	0	512
Taiwan	1	0	1183	Poland	1	0	963
Colombia	0.995	0.068	1505	Romania	1	0	1498
Ecuador	0.981	0.137	1202	Russia	1	0	2343
Estonia	1	0	496	Singapore	0.520	0.500	1730
Germany	0.990	0.099	2027	Slovenia	1	0	9
Hong Kong	0.878	0.331	49	Sweden	0.989	0.106	1142
India	1	0	1871	Thailand	1	0	1152
Japan	1	0	2443	Trinidad	0.001	0.032	996
Malaysia	0.900	0.300	391	Turkey	1	0	1498
Mexico	0.974	0.159	1936	United States	0.072	0.259	2173
Morocco	0.006	0.076	1199	Uruguay	0.997	0.055	992

Overall

	Mean	Std. Dev.	Freq.
Total	0.820	0.384	41152

Note: Distribution of the politeness distinction variable of the language individuals speak at home grouped by their country of living. Frequency is the total number of observations from one country.

3.4.3 Control Variables

I use Hofstede's cultural dimensions (Hofstede, 2001) as a control for cultural characteristics and differences of the societies I look at. I'm especially interested in masculinity as it is a measurement for competitiveness within a society and power distance, a measurement of hierarchy and its acceptance in society, as these two have been identified to influence attitude towards foreigners before (Leong and Ward, 2006). As a second source for cultural characteristics and differences of societies I use data from the GPS (Falk et al., 2016; Falk et al., 2018) for additional regressions. Data from the GPS is not yet widely used in research about attitudes of individuals towards foreigners, but it gives me a second data set with a different composition of countries. This gives me the

opportunity to test my hypothesis for a wider range of countries and languages. Unfortunately, the GPS does not contain direct correspondences to Hofstede's dimensions of masculinity and power distance. The five preferences reported by the GPS that I use for my research are *Patience*, *Risk Preference*, *Positive Reciprocity*, *Negative Reciprocity* and *Altruism*.

I use a question from the WVS to control for the effect of the general trust level of a person. Question V24 of the WVS asks: *Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?* and gives two possible answers: i) *Most people can be trusted.* and ii) *Need to be very careful.*. The GPS also contains data on trust, but it is again aggregated on the country level. Therefore, I do not use it as it contains less information than the data from the WVS.

Furthermore, I use a wide range of information about the respondents, that is found in the WVS. Starting with information about the respondent's age (V241), gender (V240), religion (V144), education (V248), martial status (V57) and number of children (V58). A second big block of questions is about their financial and employment situation. Are they employed (V229)? Are they the chief wage earner in their family (V235)? In which income class would they sort themselves (V238) and in which income class they actually are based on their income (V239). For the self reported class they can choose between upper class, upper middle class, lower middle class, working class and lower class. Question V239 has 10 different ascending income groups and the respondent is asked to state in which of this 10 groups they fall with their income. I encode these 10 groups into five to match the possible answers from question V238. And finally are they in fear of losing their job or worried to not find one (V181). The scale of possible answers has five answers, *Very much*, *A good deal*, *Not much*, *Not at all* and *Don't know/No answer*. If you are unemployed or in fear of losing your job, strangers can be seen as competition in the job market. This might increase your reluctance towards them. This effect can be increased if you are the chief wage earner of your family or in lower income classes where the financial situation is tougher. I also control if the respondents (V246) or their parents (V243, V244) are immigrants themselves. Own experience with immigration and being a foreigner in a new country might increase your own openness towards foreigners because you can put yourself in their situation. Another question that is somewhat

linked to the ones before, is the question about the general happiness in life (V10). The respondents are asked to state if they are *Very happy*, *Rather happy*, *Not very happy* or *Not at all happy*. It might be the case that unhappy persons are more unfriendly to others in general or that they blame their misfortune onto others. Foreigners are a common target for such blame. Therefore, the happiness might affect an individual's attitude towards foreigners. The last question I use is about the political views of the respondents. They have to sort themselves into a left-right scale concerning their political position (V95). The scale rages from 1 (left) to 10 (right). I code this 10 scale into five different groups, *Left*, *Center Left*, *Center*, *Center Right* and *Right*. I would expect that people to the right of the political spectrum are more reserved towards foreigners.

3.5 Empirical Approach

I examine the effect of an individuals attitude towards foreigners using the following ordered probit model:

$$Pr(trust_j = i) = Pr(\kappa_{i-1} < \beta_1 + \beta_2 pd_j + \beta_3 X_j + \beta_4 X_\ell + \beta_5 X_C + \epsilon_j < \kappa_i.) \tag{3.1}$$

The dependent variable $trust_j$ is an individuals answer to the question if they trust people of another nationality. It takes on one of four possible outcomes (1 = *Do not trust at all*, 2 = *Do not trust very much*, 3 = *Trust somewhat* and 4 = *Trust completely*). The main independent variable of interest is pd_j. It takes on the value 1 if the language a person speaks at home has a politeness distinction, otherwise it is 0. X_j are characteristics of individual j, e.g., age, gender, religion, job status ect., and their answers to other questions of interest in the WVS, for example the political views or general happiness in life . X_ℓ are language specific characteristics, such as gender and family. X_C are either country fixed effects or Hofstede's cultural dimensions. I can only assign to the participants the masculinity and power distance values of their country of residence, as the WVS unfortunately does not contain any information about the participants' country of birth. So these variables have no variation within a country. In order to still control for their effect, I have to drop the country fixed effects. This is done in further regressions. κ denotes the cut off points between the different categories of

trust. Standard Errors are clustered at the language level.[4]

3.6 Main Results

Table 3.4 presents my main empirical findings. The coefficient for the politeness distinction in a language is negative as expected, i.e., people speaking a language with a distinction are more likely to trust people with another nationality less. If you look at the marginal effects for the four possible answers in column IV to VII, individuals have a higher chance to answer *do not trust at all* or *do not trust very much* and a lower probability to answer with *trust somewhat* and *trust completely*. The marginal effects are of considerable size, ranging from just under 2% to almost 8%. They are all statistically significant at the 1% level.

TABLE 3.4: Language Effect on Trust

	I	II	III	Do not trust at all	Do not trust very much	Trust somewhat	Trust completely
Politeness Distinction	-0.442***	-0.441***	-0.293***	0.079**	0.018**	-0.074***	-0.024***
	(0.147)	(0.144)	(0.112)	(0.035)	(0.009)	(0.028)	(0.009)
Trust			0.435***	-0.117***	-0.027*	0.109***	0.035***
			(0.032)	(0.011)	(0.016)	(0.009)	(0.006)
Immigrant			0.131***	-0.035***	-0.008	0.033***	0.011***
			(0.034)	(0.009)	(0.006)	(0.009)	(0.003)
Immigration Status Parents							
One Immigrant			0.127***	-0.033***	-0.010*	0.032***	0.011***
			(0.030)	(0.007)	(0.006)	(0.007)	(0.004)
Both Immigrants			0.028	-0.008	-0.002	0.007	0.002
			(0.048)	(0.013)	(0.003)	(0.012)	(0.004)
Country Fixed Effects	√	√	√	√	√	√	√
Language Family	√	√	√	√	√	√	√
Language Genus	√	√	√	√	√	√	√
Personal Characteristics	×	√	√	√	√	√	√
Employment Status	×	×	√	√	√	√	√
Financial Situation	×	×	√	√	√	√	√
Family Situation	×	×	√	√	√	√	√
Political Views	×	×	√	√	√	√	√
Observations	41152	41152	41152	41152	41152	41152	41152

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. Column I to III report coefficients for ordered probit regressions with different sets of control variables. Column IV to VII report the marginal effects for each of the four different possible answers given for question V107 of the WVS. The same set of control variables is used as in column III.

Individuals who trust people more in general also trust foreigners more, as one would expect. But the important point for my research is that the inclusion of the trust variable

[4]Regressions with standard errors clustered at the country level can be found in section B.1 in the appendix

does not change the direction and significance of the effect of the language variable. This suggests that the language variable does not just capture a general effect on trust that translates into higher trust in foreigners. The language variable has an additional effect on a person's attitude toward foreigners, in addition to the effect it might have on a person's trust in other people.

TABLE 3.5: Language Effect on Trust: Cultural Dimensions

	I	Do not trust at all	Do not trust very much	Trust somewhat	Trust completely
Politeness Distinction	-0.333***	0.092***	0.021	-0.085***	-0.027***
	(0.115)	(0.034)	(0.014)	(0.031)	(0.009)
Trust	0.416***	-0.115***	-0.026*	0.107***	0.034***
	(0.037)	(0.014)	(0.015)	(0.009)	(0.006)
Culture					
Masculinity	-0.003	0.001	0.000	-0.001	-0.000
	(0.003)	(0.001)	(0.000)	(0.001)	(0.000)
Power Distance	-0.012***	0.003***	0.001*	-0.003***	-0.001***
	(0.002)	(0.001)	(0.000)	(0.001)	(0.000)
Immigrant	0.122***	-0.034***	-0.008	0.031***	0.010***
	(0.039)	(0.010)	(0.006)	(0.011)	(0.003)
Immigration Status Parents					
One Immigrant	0.121***	-0.032***	-0.009*	0.031***	0.011**
	(0.040)	(0.012)	(0.005)	(0.010)	(0.004)
Both Immigrants	0.023	-0.006	-0.001	0.006	0.002
	(0.054)	(0.015)	(0.003)	(0.014)	(0.005)
Country Fixed Effects	×	×	×	×	×
Language Family	√	√	√	√	√
Language Genus	√	√	√	√	√
Personal Characteristics	√	√	√	√	√
Employment Status	√	√	√	√	√
Financial Situation	√	√	√	√	√
Family Situation	√	√	√	√	√
Political Views	√	√	√	√	√
Observations	41152	41152	41152	41152	41152

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. Column I reports coefficients for ordered probit regressions. Column II to V report the marginal effects for each of the four different possible answers given for question V107 of the WVS. The same set of control variables is used as in column III. Hofstede's cultural dimensions (Hofstede, 2001) are used to control for for cultural characteristics and differences of societies.

A person's own experience with immigration, either because they are immigrants themselves or indirectly through their parents' immigration history, has a positive effect on a person's trust towards foreigners. The experience of being a foreigner in a place where

new neighbours may have little trust leads people to be more open to others themselves. For the effect, it seems not to be important whether one has had the experience oneself or whether they only know it from the stories of their parents.

WVS does not ask for the country of origin of the participants. Therefore, it is only possible to assign to individuals the value of the cultural dimension of their country of residence. This leads to the fact that the cultural variables have no variation within a country. Therefore, it is only possible to measure the effect of the culture dimensions if the country fixed effects are omitted to allow for between countries variation. The results of the regression without country fixed effects are shown in table 3.5. The language variable still has a negative effect on the trust an individual has towards foreigners. Power Distance as a measure of hierarchy in a society leads to a higher level of xenophobia as previous research Leong and Ward, 2006 has also found. Masculinity has no statistically significant effect. The effects of the other variables remain unchanged when compared to column 3.

3.7 Robustness Tests

3.7.1 Immigrants

The vast majority of individuals in my data speak a language without politeness distinction (Table 3.3). The variation in this dimension is therefore very small. Moreover, the variation in languages within the group of languages that do not have a politeness distinction is also small. As can be seen from tables 3.3 and 3.6, these observations are predominantly from individuals who speak English and live in an English-speaking country. As an additional robustness check, I try to overcome these limitations by using different subsamples.

The first subsample contains only individuals who have experience with immigration, either directly or indirectly through their parents. First- or second-generation immigrants may still speak the language of their origin at home, leading to higher variation in the variables of interest. Table 3.6 shows that the proportion of observations of individuals speaking a language without politeness distinction increases, but the proportion speaking English also increases. This effect is driven primarily by immigrants to

TABLE 3.6: Politeness Distinction Across Languages

Full Dataset			Only Immigrants			Official language		
Politeness Distinctions in Pronouns	Freq.	Perc.	Politeness Distinctions in Pronouns	Freq.	Perc.	Politeness Distinctions in Pronouns	Freq.	Perc.
No distinction	7391	17.96%	No distinction	1539	31.05%	No distinction	253	14.99%
Distinction	33761	82.04%	Distinction	3418	68.95%	Distinction	1435	85.01%
Total	41152		Total	4957		Total	1688	

Languages with no politeness distinction

Language at home	Freq.	Perc.	Language at home	Freq.	Perc.	Language at home	Freq.	Percent
Albanian	5	0.07%	Albanian	5	0.32%	Albanian	5	1.98%
Arabic	1150	15.56%	Arabic	27	1.75%	Arabic	23	9.09%
Aymara	7	0.09%	Berber	1	0.06%	Aymara	7	2.77%
Berber	66	0.89%	English	1506	97.86%	Berber	66	26.09%
Brahui	33	0.45%				English	150	59.29%
English	6128	82.91%				Maori	2	0.79%
Maori	2	0.03%						
Total	**7391**		Total	**1539**		Total	**253**	

Note: Distribution of politeness distinction across all observed individuals. The politeness distinction variable always refers to the language that the respondents reported as the language they speak at home. The distribution is presented for the full data set and the two subsamples used for regressions, the results of which are presented in table 3.4, table 3.5, table B.3, table B.4, table B.5 and table B.6. The second part of the table shows the language distribution of those who, when asked what language they speak at home, reported a language that does not have a politeness distinction in its second person pronouns.

typical immigration countries such as the United States, Australia, and New Zealand, all of which are English-speaking countries and English being a language with no politeness distinction.

The results of the regressions with this subsample are reported in table 3.7. The politeness distinction variable continues to have a negative effect on the level of trust a person places in foreigners, both in the regressions with country fixed effects and in the regressions that omit them to test for the effect of cultural dimensions. The results for the other variables are also unchanged compared to the results for the full dataset reported in table 3.4.

The second subsample looks only at people who speak a language that is not the official language of the country in which they live. The official language of a country is taken from the CIA World Factbook (Central Intelligence Agency, 2020). This leads to a larger variation among languages that do not have a politeness distinction. However, the

TABLE 3.7: Language Effect on Trust: Immigrants

	Country Fixed Effects			Cultural Dimensions
	I	II	III	I
Politeness Distinction	-0.456**	-0.489**	-0.364**	-0.242**
	(0.187)	(0.191)	(0.150)	(0.103)
Trust			0.490***	0.503***
			(0.040)	(0.037)
Culture				
Masculinity				-0.000
				(0.004)
Power Distance				-0.010***
				(0.002)
Country Fixed Effects	√	√	√	×
Cultural Dimensions	×	×	×	√
Language Family	√	√	√	√
Language Genus	√	√	√	√
Personal Characteristics	×	√	√	√
Immigration History	×	×	√	√
Employment Status	×	×	√	√
Financial Situation	×	×	√	√
Family Situation	×	×	√	√
Political Views	×	×	√	√
Observations	4957	4957	4957	4957

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. For all regressions, only people who report being immigrants themselves or report at least one parent as an immigrant are included (Questions V243-245 in WVS). Column I to III report coefficients for ordered probit regressions with different sets of control variables. Hofstede's cultural dimensions (Hofstede, 2001) are used to control for for cultural characteristics and differences of societies instead of country fixed effects in column IV. Marginal Effects are reported in section B.2.1 in the appendix.

proportion of observations that do not have a politeness distinction remains the same compared to the full dataset.

In column three of table 3.8, when all control variables are added, the coefficient of the politeness discrimination variable loses its statistical significance but the direction of the effect is still negative. The marginal effects also still have the expected direction. I would argue that the lack of significance is most likely due to the small sample size. Because of this and because the effects are still in the same direction I would say that the results support my original results with the full data set. The results with the cultural dimensions instead of the country fixed effects are shown in column IV. Here the language variable remains significant at the 10% level. The results overall fall in line with the previous results of the full dataset and the first subsample.

TABLE 3.8: Language Effect on Trust: Not Speaking Official Language

	Country Fixed Effects			Cultural Dimensions
	I	II	III	I
Politeness Distinction	-0.644*	-0.613*	-0.565	-0.327*
	(0.354)	(0.336)	(0.351)	(0.186)
Trust			0.547***	0.574***
			(0.092)	(0.089)
Culture				
Masculinity				-0.003
				(0.003)
Power Distance				-0.016***
				(0.004)
Country Fixed Effects	√	√	√	×
Cultural Dimensions	×	×	×	√
Language Family	√	√	√	√
Language Genus	√	√	√	√
Personal Characteristics	×	√	√	√
Immigration History	×	×	√	√
Employment Status	×	×	√	√
Financial Situation	×	×	√	√
Family Situation	×	×	√	√
Political Views	×	×	√	√
Observations	1688	1688	1688	1688

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. Individuals who reported to speak at home the official language of the country they live in have been dropped all regressions. The official language of a country was taken from the CIA World Factbook (Central Intelligence Agency, 2020). Column I to III report coefficients for ordered probit regressions with different sets of control variables. Hofstede's cultural dimensions (Hofstede, 2001) are used to control for for cultural characteristics and differences of societies instead of country fixed effects in column IV. Marginal Effects are reported in section B.2.2 in the appendix.

3.7.2 General Preference Survey

In further regressions, I use the cultural preferences of the GPS instead of the cultural dimensions of Hofstede. Participants are again assigned the preferences of their country of residence, as I have no information about their country of birth. The procedure for imputation is the same as before, observations missing the politeness distinction or the preferences are deleted and then the missing answers to survey questions are imputed using 35 rounds of chained imputation. Missing values for the politeness distinction and the preferences are not imputed for the same reasons discussed in section 3.4.

The GPS covers other countries than Hofstede with its cultural dimensions. This gives

me the opportunity to test my hypothesis for a different and wider group of countries, even though the GPS preferences do not map exactly the same cultural characteristics of societies as Hofstede. Compared to before, mainly countries from North Africa and Sub-Saharan Africa are added.

TABLE 3.9: Politeness Distinction Across Languages (GPS)

Full Dataset			Only Immigrants			Official language		
Politeness Distinctions in Pronouns	Freq.	Perc.	Politeness Distinctions in Pronouns	Freq.	Perc.	Politeness Distinctions in Pronouns	Freq.	Perc.
No distinction	12414	27.82%	No distinction	1505	32.56%	No distinction	411	10.49%
Distinction	32203	72.18%	Distinction	3117	67.44%	Distinction	3506	89.51%
Total	44617		Total	4622		Total	3917	

Languages with no politeness distinction

Language at home	Freq.	Perc.	Language at home	Freq.	Perc.	Language at home	Freq.	Percent
Albanian	5	0.04%	Albanian	5	0.33%	Albanian	5	1.22%
Arabic	5823	46.91%	Arabic	616	40.93%	Arabic	23	5.60%
Aymara	7	0.06%	Berber	16	1.06%	Aymara	7	1.70%
Berber	246	1.98%	English	808	53.69%	Berber	66	16.06%
Brahui	33	0.27%	Ewe	11	0.73%	Brahui	33	8.03%
English	4326	34.85%	Hausa	25	1.66%	English	81	19.71%
Ewe	167	1.35%	Igbo	3	0.20%	Ewe	167	40.63%
Hausa	615	4.95%	Swahili	3	0.20%	Hausa	28	6.81%
Igbo	338	2.72%	Zulu	18	1.20%	Zulu	1	0.24%
Swahili	12	0.10%						
Zulu	842	6.78%						
Total	**12414**		Total	**1505**		Total	**411**	

Note: Distribution of politeness distinction across all observed individuals. The politeness distinction variable always refers to the language that the respondents reported as the language they speak at home. The distribution is presented for the full data set and the two subsamples used for regressions, the results of which are presented in table 3.10, table B.7 and table B.9. The second part of the table shows the language distribution of those who, when asked what language they speak at home, reported a language that does not have a politeness distinction in its second person pronouns. Data for the dataset with the GPS preference measures instead of Hofstede's cultural dimensions.

The proportion of languages without politeness distinction is slightly higher compared to the first data set. The group of languages without politeness distinction is no longer dominated by English, but by Arabic and English. This is mainly due to the countries of North Africa that have been added. Furthermore, some African languages have been added (Table 3.9).[5] The same two subsamples are used as before.

[5] A list of all reported languages and their value for the politeness distinction variable can be found in the appendix table B.13

The language variable again has a negative effect that is statistically significant. The marginal effects go in the same direction as before and their size is also comparable. Individuals who speak a language with politeness distinction are more likely to answer *Do not trust at all* and *Do not trust very much* and less likely to answer *Trust somewhat* and *Trust completely*. Furthermore, people who generally have a higher level of trust towards other people or people who have their own migration history are more likely to trust foreigners.

TABLE 3.10: Language Effect on Trust (GPS)

	I	II	III	Do not trust at all	Do not trust very much	Trust somewhat	Trust completely
Politeness Distinction	-0.459***	-0.453***	-0.307***	0.087***	0.014	-0.076***	-0.025***
	(0.134)	(0.133)	(0.103)	(0.033)	(0.009)	(0.025)	(0.008)
Trust			0.409***	-0.117***	-0.018	0.102***	0.033***
			(0.029)	(0.008)	(0.014)	(0.008)	(0.004)
Country Fixed Effects	√	√	√	√	√	√	√
Language Family	√	√	√	√	√	√	√
Language Genus	√	√	√	√	√	√	√
Personal Characteristics	×	√	√	√	√	√	√
Immigration History	×	×	√	√	√	√	√
Employment Status	×	×	√	√	√	√	√
Financial Situation	×	×	√	√	√	√	√
Family Situation	×	×	√	√	√	√	√
Political Views	×	×	√	√	√	√	√
Observations	44617	44617	44617	44617	44617	44617	44617

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. Column I to III report coefficients for ordered probit regressions with different sets of control variables. Column IV to VII report the marginal effects for each of the four different possible answers given for question V107 of the WVS. The same set of control variables is used as in column III.

To control for the cultural difference of societies and individuals, I must again omit the country fixed effects to allow for variation between countries. The language variable still has a negative effect and the margin effects also keep their direction. Of the preferences, *Risk Preference*, *Patience* and *Positive Reciprocity* have a positive effect and *Altruism* and *Negative Reciprocity* have a negative effect. The other variables have the same effect as before.

For the first sub-sample, I again look only at those individuals who stated that they were immigrants themselves or whose parents were immigrants. The results are shown in table 3.12. The effect of the language variable remains unchanged both in the regression with country fixed effects and in the regression with cultural preferences.

TABLE 3.11: Language Effect on Trust: Cultural Preferences (GPS)

	I	Do not trust at all	Do not trust very much	Trust somewhat	Trust completely
Politeness Distinction	-0.349***	0.101***	0.016	-0.088***	-0.029***
	(0.102)	(0.031)	(0.012)	(0.027)	(0.008)
Trust	0.415***	-0.120***	-0.019	0.105***	0.035***
	(0.030)	(0.011)	(0.014)	(0.008)	(0.005)
Preferences					
Risk Preference	0.361***	-0.105***	-0.017	0.091***	0.030***
	(0.123)	(0.033)	(0.015)	(0.032)	(0.011)
Altruism	-0.646***	0.187***	0.030	-0.164***	-0.054***
	(0.228)	(0.069)	(0.022)	(0.055)	(0.020)
Patience	0.592***	-0.172***	-0.028	0.150***	0.050***
	(0.133)	(0.031)	(0.023)	(0.032)	(0.014)
Pos. Reciprocity	0.681***	-0.198***	-0.032	0.172***	0.057***
	(0.159)	(0.047)	(0.023)	(0.039)	(0.016)
Neg. Reciprocity	-0.619***	0.180***	0.029	-0.157***	-0.052***
	(0.187)	(0.051)	(0.024)	(0.047)	(0.018)
Country Fixed Effects	×	×	×	×	×
Language Family	√	√	√	√	√
Language Genus	√	√	√	√	√
Personal Characteristics	√	√	√	√	√
Immigration History	√	√	√	√	√
Employment Status	√	√	√	√	√
Financial Situation	√	√	√	√	√
Family Situation	√	√	√	√	√
Political Views	√	√	√	√	√
Observations	44617	44617	44617	44617	44617

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. Column I reports coefficients for ordered probit regressions. Column II to V report the marginal effects for each of the four different possible answers given for question V107 of the WVS. The same set of control variables is used as in column III. GPS preference measures (Falk et al., 2018) are used to control for for cultural characteristics and differences of societies.

The second subsample contains only individuals that indicate as the language they speak at home a language that is different from the official language of their country of residence. Table 3.13 presents the regression results for this subsample. The coefficient of the language variable loses its statistical significance, but the sign of the coefficient is still negative. In the regression with cultural preferences, the coefficient additionally becomes very small. The lack of statistical significance could be due to the small sample size, just as in table B.5. The loss of effect size could also be due to the change in composition in the observations reporting a language without politeness distinction. Due to the subsample, this group is now dominated by languages native to Africa. It could

TABLE 3.12: Language Effect on Trust: Immigrants (GPS)

	Country Fixed Effects			Cultural Preferences
	I	II	III	I
Politeness Distinction	-0.510***	-0.520***	-0.390***	-0.346***
	(0.162)	(0.165)	(0.125)	(0.098)
Trust			0.472***	0.482***
			(0.031)	(0.030)
Preferences				
Risk Preference				0.274*
				(0.141)
Altruism				-0.786**
				(0.323)
Patience				0.467***
				(0.165)
Pos. Reciprocity				0.998***
				(0.242)
Neg. Reciprocity				-0.906***
				(0.204)
Country Fixed Effects	√	√	√	×
Language Family	√	√	√	√
Language Genus	√	√	√	√
Personal Characteristics	×	√	√	√
Immigration History	×	×	√	√
Employment Status	×	×	√	√
Financial Situation	×	×	√	√
Family Situation	×	×	√	√
Political Views	×	×	√	√
Observations	4622	4622	4622	4622

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. For all regressions, only people who report being immigrants themselves or report at least one parent as an immigrant are included (Questions V243-245 in WVS). Column I to III report coefficients for ordered probit regressions with different sets of control variables. GPS preference measures (Falk et al., 2018) are used to control for for cultural characteristics and differences of societies instead of country fixed effects in column IV. Marginal Effects are reported in section B.2.3 in the appendix.

be that there are particular factors in African countries or among people from Africa that lead to less trust in foreigners. By omitting the country fixed effects in column IV, the effects of these factors are absorbed by the language variable and counteract the true effect of language. The cultural preferences except for *Risk Preference* and *Patience* also lose their significance. In addition, the direction of the effect of *Risk Preference* and *Altruism* changes.

TABLE 3.13: Language Effect on Trust: Not Speaking Official Language (GPS)

	Country Fixed Effects			Cultural Preferences
	I	II	III	I
Politeness Distinction	-0.597*	-0.554*	-0.507	-0.022
	(0.324)	(0.315)	(0.324)	(0.205)
Trust			0.418***	0.435***
			(0.056)	(0.054)
Preferences				
Risk Preference				-1.220**
				(0.476)
Altruism				0.319
				(0.263)
Patience				0.565***
				(0.147)
Pos. Reciprocity				0.205
				(0.255)
Neg. Reciprocity				-0.205
				(0.328)
Country Fixed Effects	√	√	√	×
Language Family	√	√	√	√
Language Genus	√	√	√	√
Personal Characteristics	×	√	√	√
Immigration History	×	×	√	√
Employment Status	×	×	√	√
Financial Situation	×	×	√	√
Family Situation	×	×	√	√
Political Views	×	×	√	√
Observations	3917	3917	3917	3917

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. Individuals who reported to speak at home the official language of the country they live in have been dropped all regressions. The official language of a country was taken from the CIA factbook Central Intelligence Agency, 2020. Column I to III report coefficients for ordered probit regressions with different sets of control variables. GPS preference measures (Falk et al., 2018) are used to control for for cultural characteristics and differences of societies instead of country fixed effects in column IV. Marginal Effects are reported in chapter B.2.3 in the appendix.

3.7.3 General Trust and Language

One possible idea for the channel through which language influences people's attitudes toward foreigners could be the general trust that a person has in other people. My previous regressions have shown that general trust has a significant positive effect on the level of trust that a person has towards foreigners. To test for this channel, I regress the language variable on the general trust variable from the WVS for the six different datasets from my previous regressions. The results of these regressions are shown in

table 3.14. I find no significant effect of the language variable on the trust that participants place in people in my data. Therefore, the effect of the language variable on attitudes toward foreigners is independent of a person's general trust and is an genuine effect on the attitude toward foreigners.

TABLE 3.14: Language Effect on General Trust

	I	II	III	IV	V	VI
Politeness Distinction	-0.351	-0.368	0.185	-0.535*	-0.508	-0.347
	(0.305)	(0.364)	(0.399)	(0.314)	(0.358)	(0.370)
Country Fixed Effects	√	√	√	√	√	√
Language Family	√	√	√	√	√	√
Language Genus	√	√	√	√	√	√
Personal Characteristics	√	√	√	√	√	√
Observations	41146	4938	1650	44614	4587	3861

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. Column I to III are the samples with Hofstede's cultural dimensions (Hofstede, 2001) and IV to VI with the cultural preferences from GPS (Falk et al., 2018).

3.8 Conclusion

Overall, my results show that language is an important new factor to explain the different attitudes of people towards foreigners. The fact whether a language has a politeness distinction in its second person pronouns has a significant influence on how much trust speakers of that language place in foreigners. People who speak a language with a politeness distinction are more likely to indicate that they do not trust foreigners at all or not very much. On the other hand, they have a lower probability of responding that they trust foreigners completely or at least somewhat. This effect is robust to a variety of control variables and persists even when country fixed effects are omitted to control for cultural differences across societies.

Unfortunately, the variation within my language variable and especially the number of different languages that do not have politeness distinction is limited due to constraints in my data. To address this problem, I look at two different subsamples of my data that include, first, only first and second wave immigrants and, second, only people who do not speak the official language of their country of residence. The subsample results also show a negative effect of the variable on attitudes towards foreigners and support the

results with the full data set. In addition, I use a second source for cultural peculiarities of societies, namely GPS. This gives me the opportunity to test my hypothesis for additional countries and languages. These results also support my initial findings.

For future research on people's attitudes towards foreigners, therefore, the language of the people under investigation should always be included in addition to the economic and non-economic factors that have played a central role in the research so far.

Chapter 4

Language and Gender: How Linguistic Differences Influence a Woman's Labour Market Outcomes

In this chapter, I empirically examine how the grammatical gender system in a language affects the gender norms of an individual and the labour market outcomes for women. I combine data about language features related to gender from the World Atlas of Language Structure and data about individuals from the European Social Survey to investigate cross-linguistic differences within European countries. A gender system based on biological sex is linked to more traditional beliefs about the role of women; they should prioritise their family over their career while men have more rights to jobs. Furthermore, it also leads to worse labour market outcomes for women. They either completely stop participating in the labour market or reduce their hours of work. Women also spend more time on housework, which fits with the greater belief that they should prioritise their family life. The effect of other language features about the number of genders, the system of gender assignment and gender in pronouns is on the other hand ambiguous, which could be due to the fact that their connection to the biological sex is not as direct. Therefore, language features should be considered when talking about factors influencing the labour market outcomes for women.

4.1 Introduction

This chapter examines the impact of certain language features on gender norms and women's real labour market outcomes in 27 European countries. I discover that a gender system based on biological sex in a language is associated with a higher level of agreement on statements that women should put their families before their careers and that men should have more right to a job than women if jobs are scarce. In addition, a language's gender system also influences a women's labour market outcomes in the real world. It is associated with a lower labour supply of women. They are less likely to participate actively in the labour market and if they do, they report working fewer hours per week. However, I observe not only effects on whether a woman works, and if so, what her working hours are, but also in what occupation she works. These women are more likely to work in a occupation that is traditionally considered more feminine, e.g., secretary instead of professor or physician's assistant instead of mechanic.

The idea that a language can influence the beliefs and the behaviour of its speakers comes from linguistics and is called *linguistic relativity*. This theory was developed in the 19th century and has mainly been associated with the American linguists Edward Sapir and Benjamin Lee Whorf since the 1930s (Sapir, 1921; Whorf, 1956). P. Wolff and Holmes (2011) further divide the theory into several sub-branches, with *language as spotlight* being the focus in this work. Language can highlight certain characteristics of reality through special words or constructs and thereby cast a spotlight on aspects of the world which makes them more prominent to the speaker. The use of a sex-based gender system and the emphasis on gender distinction in a language cast a spotlight on the fact that humans are biologically divided into two sexes. This can lead to a reinforcement of gender norms and thus to a change in attitudes and behaviour when it comes to the role of women in the professional life.

The WALS (Dryer and Haspelmath, 2013) lists four different features that are related to gender assignment and gender distinction in languages. The first three features indicate how many genders a language has (Corbett, 2013a), whether they are sex-based (Corbett, 2013b) and how genders are assigned to individual words (Corbett, 2013c). The

fourth feature captures whether the language expresses gender in its independent personal pronouns. From these four features, an index is constructed that reflects the gender intensity of a language, i.e., *Gender Intensity Index* (GII) (Santacreu-Vasut, Shoham, and Gay, 2013). I combine this language data with data from the ESS Round 5 (ESS Round 5, 2010), compiled in 2010, that asks participants in more than thirty nations about their attitudes, beliefs and behaviours. I use ESS Round 5 in particular as it contains not only questions about the respondents' occupations and working lives, but also questions about one's individual beliefs about the role of women in the working life, especially in relation to the conflict between work and family and, in comparison to men. This enables me to observe not only the effect of language features on people's beliefs and gender norms, but also how they directly affect women's lives.

In this chapter, I use a within-country design for my empirical analysis which roughly follows the epidemiological approach and compares individuals who speak different languages but live in the same institutional, legal and socio-economic environment in order to distinguish linguistic effects from confounding factors (Fernández and Fogli, 2009; Blau, Kahn, and Papps, 2011; Fernández, 2011; Blau and Kahn, 2015). Additionally, I include cultural variables from the GPS (Falk et al., 2018) to control for cultural differences. GII has a positive effect on the level of agreement that women should cut down on work in favour of their family and that men should have more right to a job if these are scarce. Therefore, a higher GII strengthens traditional gender roles. However, only the coefficient for the question about cutting down work is statistically significant. I also find negative effects on the labour supply of women; they participate less in the labour market. Furthermore, GII also influences what kind of occupation women work in. Women who speak a language with a higher GII have a greater chance to be employed in a job that is traditionally seen to be more feminine. However, none of these effects are statistically significant. Lastly, in accordance with the belief that women should favour their family over their career, women speaking a language with a higher GII actually seem to reduce their labour supply to take care of their family. They report on average more hours of housework per week. Overall, the directions of GII's effects paint a coherent picture. As some coefficients lack statistical significance however, I examine the components of the index individually to further investigate the effect of

language on women's family and working lifes.

Of the four language variables, the one that indicates whether the gender system of a language is based on biological gender provides the most promising results. Speakers of a language with a sex-based gender system report a higher level of agreement with the statement that women should cut down on their work in favour of their family. In addition, these participants also agree more with the statement that men should have more right to a job when these are scarce. A sex-based gender system is therefore associated with a belief in more traditional gender roles; women stay at home and care for the family while men go to work to provide for them. The effects are quite substantial with an increase in the level of agreement of 0.5 and 1, respectively.

Language does however not only affect the beliefs of its speakers but also their actions in the real world. A sex-based gender system results in a lower labour supply of women. Speakers of a language with a sex-based gender system report on average a lower proportion of women at their workplace with a reduction of almost two answer categories. Correspondingly, it also resulted in a lower probability for women to actively participate in the labour market. The labour supply of women participating in the labour market however is also reduced. They report an average of 12 working hours less per week. The gender system of a language does not only influence the amount of labour women supply to the labour market, it also influences the occupations in which they offer their labour. Women speaking a language with a sex-based gender system have a higher probability to report an occupation as their main job which is traditionally seen as more feminine. Language therefore perpetuates the concentration of women in certain occupational fields that were historically female. They are more likely to be a nurse instead of a doctor or a kindergarten teacher instead of working in construction. In addition, I find results that support the thesis that women reduce their labour supply in favour of their family, as one would expect from the results on participants' beliefs about the role of women. Women speaking a language with a sex-based gender system report about 9 hours more housework per week on average.

The three variables depicting number of genders, system of gender assignment and gender in pronouns also demonstrated an effect on participants' beliefs and women's labour market outcomes. However, these results are more ambiguous overall and do

not paint a picture as clear as those for the sex-based gender system. One possible explanation for these different effects is that these three features are not as clearly linked to biological sex. Therefore, the codification of the different expressions of the features into a binary variable, which on the one hand maps all expressions associated with biological sex and on the other hand maps all others, is difficult.

My results are robust to a wide range of control variables and robustness tests. The effects persist after the inclusion of several variables capturing individual characteristics and a set of cultural preferences from the GPS (Falk et al., 2018). Therefore, it can be concluded that these are direct language effects and that the language variables do not only act via other variables or reflect cultural differences. Furthermore, there is a possibility that potential unobserved variables exist that only affect natives and are not picked up by my control variables. To control for this possibility, I only use immigrant data in a robustness test for my regressions. Reassuringly, I find the same effects of a sex-based gender system in a language. Unfortunately, the questions about an individual's beliefs about the role of a women, the proportion of women at the workplace and the hours of housework in the last week are part of the rolling module about *work, family, and well-being* and are therefore not asked in every round of the ESS. However, I combine data from round 5-9 and my results for the supply of labour of women and their type of job also holds true in this larger sample.

This chapter contributes to the existing literature on the effect of language structures on women's labour market participation by extending the empirical research from analysing only if women participate, to in which intensity they participate and in which occupation. My research findings suggest that the gender system of a language does not only have an impact on women's labour market participation, but also has a negative impact on how many hours per week they work and in which occupation they work. Research by Gay et al. (2018) indicates that language influences female labour market participation by reinforcing gender norms in a society and I can support this relationship by looking at two questions that interrogate an individual's beliefs how women should behave while navigating their career and family and what their role in the professional world is compared to men. Previous research in this area has mainly focused on married women in the US with a migrant background. In my research, I look at women

from 27 European countries, regardless of their marital status or origin. I am able to show that the effect on women's labour market outcomes holds true even in this more general sample from Europe.

There is a growing literature exploring the effect of linguistic structure on people's beliefs and behaviours (M. K. Chen, 2013; Roberts and Winters, 2013; Roberts, Winters, and M. K. Chen, 2015; S. Chen et al., 2017). The presence and intensity of gender in a language is correlated to several real-world outcomes for women. Santacreu-Vasut, Shoham, and Gay (2013) examine the effect of language on female participation in politics. They find that language is linked to the introduction of gender political quota and is even a better predictor than traditional explanatory variables, e.g., economic development, political system, and religion. A sex-based grammatical gender system has a negative impact on the educational attainment of women, as it results in them being less likely to attend college (Galor, Özak, and Sarid, 2020). Hicks, Santacreu-Vasut, and Shoham (2015) finds that people who speak a language that stresses gender in its grammatical structure are more likely to distribute housework based on the sex of the members of a household. Van der Velde, Tyrowicz, and Siwinska (2015) link language to the estimates of the gender wage gap and observe that countries with a more gender neutral language have lower estimates of the gender wage gap. Santacreu-Vasut, Shenkar, and Shoham (2014) show that a more intensive marking of gender in a language leads to a lower participation of females on boards of directors and in senior management. Gay et al. (2018) use an epidemiological approach to examine the relationship between gender in language and the labour market participation of married female immigrants in the United States. Their work indicates that women who speak a language with gender-specific rules are less likely to participate actively in the labour market. They can attribute about two-thirds of these effects to correlated cultural values and at most one-third to a causal language effect.

My work further contributes to the existing literature on the role of women in society and their contribution to the labour market. Over the last decades, the labour market participation rate of women has increased around the world but there are still large differences between countries (Goldin, 2014). Cultural values in particular seem to provide a crucial explanatory contribution to explain these differences (Fernandez, 2007;

Fernández and Fogli, 2009; Alesina, Giuliano, and Nunn, 2013; Farré and Vella, 2013; Fernández, 2013; Jayachandran, 2021).

The chapter is structured as follows. Section 4.2 gives a brief overview about the theory of linguistic relativity which describes how a language can affect the decision making and behaviour of its speakers. Section 4.3 describes my underlying data and gives summary statistics for both, the ESS Round 5 survey data and the language data from the WALS. Section 4.4 presents my empirical model. Section 4.5 presents the empirical results for the impact of gender in language on gender norms and labour market outcomes for women. Section 4.7 concludes my findings.

4.2 Linguistic relativity

The idea that language can influence the behaviour of an individual can be traced back to the 19th century and became more widely known through the works of the linguists Edward Sapir (1921) and Benjamin L. Whorf (1956) in the 20th century. In its moderate interpretation *linguistic relativity*, also called *Sapir-Whorf hypothesis* or *Whorfism*, states that that linguistic categories and usage influence thought and decision making of individuals (Boroditsky, 2018). In recent years, new research in linguistic and psychology has come up with empirical studies supporting this hypothesis (Lucy, 1996; Casasanto and Boroditsky, 2008; Boroditsky, 2001; Winawer et al., 2007; Fausey et al., 2010). There is also a small but growing amount of literature about the effects of language characteristics on economic outcomes (M. K. Chen, 2013; Santacreu-Vasut, Shoham, and Gay, 2013; Roberts, Winters, and M. K. Chen, 2015; Hicks, Santacreu-Vasut, and Shoham, 2015; Gay et al., 2018; Figlio et al., 2019).[1] P. Wolff and Holmes (2011) define several alternative ways how language might influence the thoughts and therefore the behaviour of its speakers. Most relevant for my research is the category of *language as a spotlight*. This means that after exposure to words and constructions that highlight specific properties, attention may linger on those properties. Therefore, language may put a spotlight on certain aspects of the world and make them more salient to the speaker.

[1] For a more comprehensive overview of the history of linguistic relativity and economics, see Mavisakalyan and Weber (2018).

Boroditsky, L. A. Schmidt, and Philipps, p. 65 (2003) argue that *Needing to refer to an object as masculine or feminine may lead people to selectively attend to that object's masculine or feminine qualities, thus making them more salient in the representation*. A language whose gender system is based on biological sex, shines a spotlight on the distinction between the sexes in humans, causing speakers to be more likely to divide the world into male and female and to be more aware of gender differences in general (Konishi, 1993; Sera, Berge, and Pintado, 1994; Phillips and Boroditsky, 2003; Cubelli et al., 2011). This can consequently lead to different beliefs about gender norms (Pérez and Tavits, 2019; De-Franza, H. Mishra, and A. Mishra, 2020) and a different behaviour towards men and women (Galor, Özak, and Sarid, 2020; Hicks, Santacreu-Vasut, and Shoham, 2015; Gay et al., 2018) of the language user.

For my research, I use four different characteristics of languages from the WALS that are related to gender in languages (Dryer and Haspelmath, 2013). I use these measurements for the presence of gender distinctions in languages to test whether the spotlight cast on biological sex through language has an effect on the beliefs of a society regarding the role of a woman in the labour market and family life and further also on the real-world labour market outcomes for women.

4.3 Data

4.3.1 Socio-Economic and Demographic Data

My main source of data is the ESS Round 5 (ESS Round 5, 2010). Round 5 queries not only the respondents' main occupation and labour market situation, but also inquires their general ideas about the role of women in relation to work and family.

My analysis is divided into two parts. The first part of my analysis covers an individual's and accordingly also society's general beliefs about the role of women in work and family life and the relationship between women and men in the labour market. The first dependent variable encodes the support of the interviewee to the question if *[a] woman should be prepared to cut down on her paid work for the sake of her family*. There are five possible answers, that descend from *Agree strongly* to *Disagree strongly*. The second

question is about the conflict between women and men when it comes to the distribution of scarce jobs. It queries the respondent's opinion about the statement that *[w]hen jobs are scarce, men should have more right to a job than women*. As previously, the same five possible answers are given.

In the second part of my analysis, I focus on the effect of language on the labour market outcomes for women. The first dependent variable I use in the second part of my analysis, inquires what percentage of the interviewee's workplace is female. There are seven ascending categories, from *None* to *All*. Following this regression, I restrict my data to only include respondents who reported to be female. The next dependent variable measures whether a woman is actively participating in the labour market. *Labour Market Participation* is 1 if the respondent answered *In paid work* or *Unemployed and actively looking for a job* to the question about their main activity in the last seven days and 0 otherwise. However, I am not only interested in whether a woman participates in the labour market at all, but in particular in the quantity and quality of her participation, i.e., in which field she works and how many hours she works. The *Total Work Hours* reports the average weekly working hours in the main job of the respondent who previously stated that her main activity in the last seven days was *In paid work*. It captures the real workload of their main job per week and not only the contracted hours by also including paid and unpaid overtime. The next dependent variable covers the occupation in which the woman currently works or was last employed in. There are jobs that have traditionally been seen as more feminine, e.g., a nurse compared to a doctor, a secretary compared to a manager or whole professions such as preschool teachers or midwifery. The respondents are asked which name or title their main job currently has or previously had, and the given answers are coded according to the International Standard Classification of Occupations ISCO-88. *Female Occupation* is a dummy variable that is equal to 1 if the main job of the respondent is considered to be traditionally more feminine and 0 otherwise.[2] I follow data from Eurostat and the International Labour Organization, which reports the proportion of women in different occupations, to determine which occupations are female-gendered (Eurostat, 2018; ILO, 2020). The last dependent variable I use, reports the total hours spent on housework per week.

[2] A complete list of occupations considered to be traditionally more feminine can be found in table C.2 in the Appendix.

Housework Hours is a categorical variable that divides the reported hours into groups of 10 hours, i.e, 0 to 9 hours reported, 10 to 19 hours etc., up to 99 hours of housework. The last category contains all individuals who reported more than 100 hours of housework per week. The summary statistics of my variables of interest can be found in table 4.1.

TABLE 4.1: Summary statistics: Dependent Variables

	Women cut Job for Family	Scarce Jobs	Proportion of Women	Labour Market Participation	Female Occupation	Total Work Hours	Housework Hours
Mean	2.744	3.469	3.992	0.482	0.337	36.32	2.882
SD	1.169	1.277	1.774	0.500	0.473	12.84	1.581
Min.	1	1	1	0	0	0	1
Max.	5	5	7	1	1	168	11
Obs.	25,738	25,823	9,983	14,317	12,173	5,887	7,249
Miss.	457	372	16,212	14	2,158	8,444	7,082

Note: Question G4: A woman should be prepared to cut down on her paid work for the sake of her family. Question G5: When jobs are scarce, men should have more right to a job than women. The given answers are ordered from *Agree strongly* to *Disagree strongly*. Question G38: Proportion of women at workplace. There are seven ascending answers ranging from none to all.*Labour Market Participation*: Respondent was *in paid work* or *unemployed and actively looking for a job* in the last seven days. *Female Occupation*: The main occupation of the respondent is seen as traditionally more feminine. *Total Work Hours*: Counts the hours the respondent normally works per week, including paid and unpaid overtime. Only respondents with paid work in the last seven days are included. *Housework Hours* is the total hours of housework reported for the last week. The reported values are grouped into brackets of 10 hours each. Last bracket contains all values above 100.

Table 4.2 presents the socio-economic and demographic characteristics of my whole sample.[3] The typical respondent is more likely to be female and almost 48 years of age. They had 12.5 years of education and are at the very top of the 4th income decile. They report on average a happiness of 7 on a scale from 0 to 10. Politically they are centrist and 67% report to be member of any religious denomination. The average household consists of three people and lives in a suburb of a big city or in a small town.

I also use data from the GPS (Falk et al., 2018) to capture the cultural characteristics of an individual's country of birth and also the cultural characteristics of the country or countries of birth of their parents. The GPS includes the cultural values patience, risk preferences, positive reciprocity, negative reciprocity, and altruism.

Missing answers from the survey questions were imputed using chained imputation with 35 rounds. I have refrained from imputing language characteristics for languages

[3]Summary statistics for the female only subsample can be found in table C.1 in the appendix.

that were reported to be spoken at home for which the WALS do not have data and cultural values for countries missing in the GPS. Overall, I have data for 26,195 individuals who were born in 145 different countries and speak 31 different languages.

TABLE 4.2: Summary Statistics: Descriptives

	Female	Age	Education	Income Class	Happiness
Mean	0.547	47.8	12.53	4.914	6.999
SD	0.498	18.72	3.978	2.744	2.109
Min.	0	14	0	1	0
Max.	1	101	50	10	10
Obs.	26,188	26,147	25,947	20,441	25,976
Missing	7	48	248	5,754	219

	Religious Denomination	Political Views Left/Right	Place of Residence Rural/Urban	Household Size
Mean	0.674	5.132	2.755	2.680
SD	0.469	2.133	1.270	1.418
Min.	0	0	1	1
Max.	1	10	5	19
Obs.	26,011	21,980	26,154	26,181
Missing	184	4,215	41	14

Note: Summary statistics of the socio-economic and demographic characteristics of the individuals.

4.3.2 Language Features

The WALS contains four different features which are explicitly related to gender in languages. I use these four features to create a measure of how intensively a language expresses gender differences in its grammatical system. For the creation of this gender intensity index I follow the approach by Santacreu-Vasut, Shoham, and Gay (2013), Hicks, Santacreu-Vasut, and Shoham (2015) and Gay et al. (2018).

The first measure, *Sex based*, captures whether the gender system of a language is linked to biological sex. The WALS distinguishes three different types of gender systems in languages. There are languages that have no gender at all, and there are languages that have either a sex-based or a non-sex-based gender system. In languages like French or German gender is linked to biological sex, so male or female sex forms the semantic core of the masculine or feminine gender but can also contain other nouns like animals.

On the other hand Fulfulde, a language spoken in Mali, has around twenty genders, which also have a semantic core but sex is not a part of it and Finnish for example has no gender at all (Corbett, 2013b). The variable takes the value 1 if the gender system is based on the biological sex and 0 otherwise.

Languages also differ in how many genders they contain. (Corbett, 2013a). If they contain only two genders, they usually force the speaker to distinguish between feminine and masculine, while languages with three genders may contain a neuter or a non-sex-based gender distinction, which means that the reflection of the two biological genders in the gender system is not as prominent as in the case of only two genders. Therefore, *Number of Genders* equals 1 if a language has two genders and 0 otherwise.

Languages additionally differ in their rules for assigning nouns to genders defined by the gender system. There are two ways to assign a noun, either by its form or by its meaning. Languages might have a semantic assignment system (meaning) or a semantic and formal assignment system but never only a formal assignment system. In Russian language for example nouns are first assigned a gender by their meaning, nouns denoting males are masculine and nouns denoting females are feminine. But the residual nouns do not all belong to a neuter gender like in the Kannada language[4] but are shared between the three genders. So there must be another rule to assign gender to the residual nouns. However, this is not done by further semantic rules. For example, *stul* (chair, masculine), *taburetka* (stool, feminine) and *kreslo* (armchair, neuter) do not share the same gender despite being semantically similar. The residual nouns are assigned a gender according to their inflectional class, i.e. by their morphology (Corbett, 2013c). *Gender Assignment* further divides languages into two categories. It is equal to 1 for languages with a semantic and formal assignment system and 0 for languages with only a semantic assignment system or no gender at all.

Finally, there is also heterogeneity in how languages distinguish gender in pronouns. The variable *Gender in Pronouns* captures these differences. It is equal to 1 for languages with a gender distinction in third-person and in first and/or second-person and 0 if there is no gender distinction in pronouns, only in the third person, only in second person, or only in first person. Table 4.3 presents these differences using the two languages

[4] A language spoken in India.

Hausa[5] and German. In the first case gender is highly prominent in the pronominal system, i.e., in Hausa language gender is expressed in second and third person singular. Therefore, it gets assigned a 1. German on the other hand only expresses gender in its third person pronouns, resulting in a value of 0 (Siewierska, 2013).

TABLE 4.3: Gender Expression in Pronouns

	Hausa		German	
First person singular	nī		ich	
Second person singular	kai	masculine	du	
	kē	feminine		
Third person singular	shī	masculine	er	masculine
	ita	feminine	sie	feminine
			es	neuter

Note: Gender expression in first, second and third person singular pronouns in Hausa and German.

These four characteristics of a language's gender system are combined into a GII. It is defined as:

$$GII = Sex\ based + Number\ of\ Genders + Gender\ Assignment + Gender\ in\ Pronouns. \quad (4.1)$$

The GII is a categorical variable that ranges from 0 to 4 (Santacreu-Vasut, Shoham, and Gay, 2013). For example, the GII for Hebrew is equal to 4 as all four features are present in the language and the score for Finnish is 0 as none of the features are present. The GII ranks the relative intensity of gender across a language's grammar. Table C.3 in the appendix contains the individual values for the four features and the GII for all languages included in my sample. An individual gets assigned the characteristics of the language they report to speak at home.

[5] A language spoken in Nigeria and Niger.

TABLE 4.4: Summary statistics: Language Features

	Sex based	Number of Genders	Gender Assignment	Gender in Pronouns	GII
Mean	0.853	0.228	0.671	0.130	1.882
SD	0.354	0.420	0.470	0.336	1.172
Min.	0	0	0	0	0
Max.	1	1	1	1	4
Obs.	26,195	26,195	26,195	26,195	26,195

Note: Summary statistics of the four language features related to gender and the GII. Summary statistics for the GII per country can be found in table C.4 in the appendix.

Summary statistics for the language variables in my regression sample can be found in table 4.4 and for the GII per country in table C.4 in the appendix.[6]

4.4 Empirical Approach

In my approach, I use country fixed effects to compare outcomes between individuals who speak different languages at home but live in the same institutional, legal, and socio-economic environment. Thereby, I'm able to separate the effects of language on beliefs and labour market outcomes from confounding institutional forces. Furthermore, I also include cultural characteristics of the birth country of an individual and of the birth countries of their parents in my analysis to control for cross-cultural differences. As linguistic and cultural traits tend to coevolve in the course of human history (Galor, ÖZak, and Sarid, 2018), I use these controls for cultural traits to isolate the pure language effect. This approach is similar to the so called epidemiological approach to culture, which uses immigrants to distinguish cultural influences from confounding institutional forces (Blau, Kahn, and Papps, 2011; Blau and Kahn, 2015; Fernández, 2011; Fernández and Fogli, 2009).

Since, unlike the epidemiological approach, I also consider natives in my analysis, there is theoretically the possibility that there are unobserved variables that unilaterally affect only natives and are not picked up by my country fixed effects and the cultural

[6]Regression results for an alternative definition of the GII applied by Gay et al. (2018) can be found in tabe C.6 in the appendix.

variables. Therefore, as a robustness test, I also use a subsample that only includes immigrants.

I use the following OLS specification[7] to estimate my results:

$$Y_{ilcb} = \alpha + \beta_1 X_{il} + \beta_2 Z_i + \beta_3 U_{ib} + \beta_4 V_l + \beta_5 W_{ic} + \varepsilon_{ilcb}, \tag{4.2}$$

where Y_{ilcb} is a measure of an individual *i's* beliefs about the role of women in the labour market or individual *i's* labour market outcomes if they are female. The subscript l indexes the language spoken at home, c country of residence and b country of birth. X_{il} is a vector of features of the language l spoken by individual i. Z_i is a vector of respondent *i's* characteristics. It contains gender, year of birth, years of education, political views on a left/right scale, membership of a religious community, personal happiness in life, household size, household income and the urbanization of the area of residence. To control for the cultural background of an individual, the vector U_{ib} contains the cultural traits of the birth country of individual i and the average cultural traits of their parents' birth countries. The cultural traits are take from the GPS. Languages cannot be assumed to be completely independent from each other, but have common ancestors. To control for this historic relativeness, V_l is a vector of fixed effects for the family and genus of a language (Roberts, Winters, and M. K. Chen, 2015). W_{ic} is a country fixed effect for the country of residence of individual i and ε is the error term. Standard errors are clustered at the language level throughout all of my regressions.

4.5 Main Results

4.5.1 Gender Intensity Index

Table 4.5 presents my empirical findings for the effect of the GII on gender norms in society and labour market outcomes for women. A higher intensity of gender in a language is associated with a higher level of agreement of its speakers with the statement that a woman should cut down on her paid work for the sake of her family. It also leads speakers of such languages to believe that men should be given preference over

[7] Results for non-linear models can be found in table C.5 in the Appendix.

women when jobs are scarce. The gender intensity in a language thus leads to a reinforcement of traditional role models and gender norms among its speakers namely that the woman is responsible for the household and family while the man goes to work. Even though both results draw the same picture about the effect of gender in language, only the effect for the reduction of working hours for the benefit of the family is statistically significant.

TABLE 4.5: Effect of Language on Gender Norms and Labour Market Outcomes

	Cut Down Work	Scarce Jobs	Proportion Women	Labour Market Participation	Female Occupation	Total Work Hours	Housework Hours
	I	II	III	IV	V	VI	VII
GII	-0.159**	-0.034	-0.099	-0.017	0.037	2.252	0.179*
	(0.069)	(0.120)	(0.178)	(0.084)	(0.037)	(1.360)	(0.099)
Female	0.103***	0.322***	1.938***				
	(0.032)	(0.059)	(0.129)				
Years of Education	0.029***	0.053***	0.050***	0.025***	-0.012***	0.059	-0.033***
	(0.005)	(0.004)	(0.007)	(0.003)	(0.002)	(0.113)	(0.008)
Household Size	-0.054***	-0.054***	0.033**	-0.026***	0.016***	-1.355***	0.253***
	(0.014)	(0.013)	(0.014)	(0.006)	(0.005)	(0.461)	(0.028)
Income Class	0.022**	0.041***	-0.012	0.038***	-0.013***	0.722***	-0.063***
	(0.008)	(0.008)	(0.009)	(0.005)	(0.002)	(0.192)	(0.011)
Country Fixed Effects	√	√	√	√	√	√	√
Language Fixed Effects	√	√	√	√	√	√	√
Age	√	√	√	√	√	√	√
Happiness	√	√	√	√	√	√	√
Religious Denomination	√	√	√	√	√	√	√
Political Views	√	√	√	√	√	√	√
Urban/Rural	√	√	√	√	√	√	√
Cultural Values	√	√	√	√	√	√	√
Cultural Values Parents	√	√	√	√	√	√	√
Observations	20622	20622	8154	11266	11266	4782	5679

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. SE in parenthesis. Levels of the GII are standardized. Dependent Variables: Question G4: A woman should be prepared to cut down on her paid work for the sake of her family. Question G5: When jobs are scarce, men should have more right to a job than women. The given answers are ordered from *Agree strongly* to *Disagree strongly*. Therefore, a negative coefficient is associated with a higher agreement with the statement. Question G38: Proportion of women at workplace. *Labour Market Participation*: Respondent was *in paid work* or *unemployed and actively looking for a job* in the last seven days. *Female Occupation*: The main occupation of the respondent is seen as traditionally more feminine. *Total Work Hours*: Counts the hours the respondent normally works per week, including paid and unpaid overtime. Only respondents with paid work in the last seven days are used. *Housework Hours* is the total hours of housework reported for the last week. The reported values are grouped into brackets of 10 hours each. Last bracket contains all values above 100. Missing observations are chained imputed.

These results raise the question of whether the effect of language is also reflected in labour market outcomes for women and they are presented in columns III to VII. A higher intensity of gender leads to participants reporting a lower proportion of women

at the workplace. This might indicate that a lower proportion of women work overall. An impression that is also fostered by the results in column IV which indicate that gender intensity also negatively affects labour market participation of women. Women who speak a language with a higher intensity of gender are less likely to actively participate in the labour market. Encouragingly, this observation is in line with the findings of Gay et al. (2018).

Language does not only influence whether a woman works, but also her line of work. Women who speak a language with a higher gender intensity are more likely to report an occupation as their current or last main job which is traditionally seen as more feminine. In addition to the impact on the type of occupation a woman pursues, I also find an impact on the reported working hours of women who are in paid employment. A higher gender intensity leads to an increase of reported working hours per week for women. Furthermore, I observe a positive effect on the number of reported hours of housework, a result that is also reported by Hicks, Santacreu-Vasut, and Shoham (2015). This increase in hours of housework reflects to some extent the results of the first question on the role of women, that women should take care of the family.

My results show a coherent picture of the effects of gender intensity in a language on gender norms and women's labour market outcomes. Nevertheless, only my results for the statement about the cut down on work by women and the reported hours of housework by women are statistically significant. To further explore the relationship between gender in language and women's labour market role, I investigate the language features that are part of the GII individually in the next section

4.5.2 Individual Language Features

The individual effects of the four language features, sex-based gender system, number of genders, gender assignment and gender in pronouns, that are combined into GII are reported in the tables 4.6, 4.7, 4.8 and 4.9. In this section, I show how the coefficients respond to the introduction of various control variables that are also suspected to be influenced by language; Galor, Özak, and Sarid (2020) for example reports effects of language on the educational attainment of women. Despite these possible links, these variables are used here as I am interested in the direct effect of language rather than

the average causal effect. Moreover, these variables are influenced by other effects and without their inclusion, these effects could be reflected in the language variables and bias my results.

TABLE 4.6: Effect of Individual Language Features: Society's Beliefs

	Cut Down Work					Scarce Jobs				
	I	II	III	IV	V	I	II	III	IV	V
Sex-Based	-0.008	-0.155	-0.185	-0.232*	-0.561***	-0.240	-0.438***	-0.458***	-0.546***	-1.062***
	(0.180)	(0.161)	(0.148)	(0.134)	(0.157)	(0.165)	(0.151)	(0.144)	(0.125)	(0.156)
Nb. of Genders	0.145	0.138	0.148	0.136	0.149	0.356**	0.336**	0.342**	0.319**	0.222***
	(0.107)	(0.106)	(0.100)	(0.096)	(0.097)	(0.137)	(0.134)	(0.131)	(0.121)	(0.068)
Assignment	-0.246**	-0.187**	-0.195**	-0.201**	-0.125	-0.246**	-0.151**	-0.156**	-0.168**	-0.148
	(0.090)	(0.091)	(0.089)	(0.086)	(0.104)	(0.103)	(0.067)	(0.066)	(0.064)	(0.114)
Pronouns	-0.147***	-0.189***	-0.192***	-0.194***	-0.159**	0.113*	0.068	0.066	0.061	0.213***
	(0.040)	(0.031)	(0.029)	(0.029)	(0.061)	(0.058)	(0.069)	(0.067)	(0.058)	(0.062)
Female	0.089***	0.081**	0.081**	0.091***	0.103***	0.311***	0.304***	0.304***	0.324***	0.322***
	(0.031)	(0.032)	(0.032)	(0.030)	(0.032)	(0.053)	(0.054)	(0.054)	(0.052)	(0.059)
Years of Education		0.033***	0.032***	0.027***	0.029***		0.061***	0.060***	0.051***	0.053***
		(0.005)	(0.005)	(0.005)	(0.005)		(0.004)	(0.004)	(0.003)	(0.004)
Household Size			-0.046***	-0.060***	-0.054***			-0.028***	-0.056***	-0.054***
			(0.011)	(0.013)	(0.014)			(0.009)	(0.011)	(0.013)
Income Class				0.023***	0.022**				0.045***	0.041***
				(0.007)	(0.008)				(0.007)	(0.008)
Country Fixed Effects	√	√	√	√	√	√	√	√	√	√
Language Fixed Effects	√	√	√	√	√	√	√	√	√	√
Age	√	√	√	√	√	√	√	√	√	√
Happiness	√	√	√	√	√	√	√	√	√	√
Religious Denomination	√	√	√	√	√	√	√	√	√	√
Political Views	√	√	√	√	√	√	√	√	√	√
Urban/Rural	√	√	√	√	√	√	√	√	√	√
Cultural Values	×	×	×	×	√	×	×	×	×	√
Cultural Values Parents	×	×	×	×	√	×	×	×	×	√
Observations	26195	26154	26154	26154	20622	26195	26154	26154	26154	20622

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. SE in parenthesis. Dependent Variables: Question G4: A woman should be prepared to cut down on her paid work for the sake of her family. Question G5: When jobs are scarce, men should have more right to a job than women. The given answers are ordered from *Agree strongly* to *Disagree strongly*. Therefore, a negative coefficient is associated with a higher agreement with the statement. Independent Variables: *Sex-Based* is equal to 1 if a language has a sex-based gender system. *Nb. of Genders* is equal to 1 if a language only has two genders. *Assignment* is equal to 1 if a language has a semantic and formal assignment system. *Pronouns* is equal to 1 for languages with a gender distinction in third-person and in first and/or second-person. Missing observations are chained imputed.

Table 4.6 presents my findings for the effects of language features on gender norms. A sex-based gender system in a language leads to a higher level of agreement with both statements. The agreement increases by 0.6 and 1 levels, respectively. Both coefficients are statistically significant at the 1% level. The number of genders in a language on the other hand exhibits a positive effect. Only having two genders results in a lower level of agreement to the two statements compared to languages with less or more genders. However, only the coefficient for the statement about the greater right for men to a job compared to woman is statistically significant. The assignment of gender in a language

has a negative effect on both statements, which are however not statistically significant. The results for the gender in pronouns of languages are ambiguous. It leads to a higher level of agreement with the statement that women should cut down on work but to a lower level of agreement with the statement that men should have more rights to a job if jobs are scarce.

TABLE 4.7: Effect of Individual Language Features: Female Labour Market Outcomes I

	Proportion Women					Labour Market Participation				
	I	II	III	IV	V	I	II	III	IV	V
Sex-Based	-1.348***	-1.492***	-1.392***	-1.378***	-1.833***	0.034	-0.029	-0.031	-0.101	-0.155
	(0.215)	(0.258)	(0.235)	(0.241)	(0.330)	(0.083)	(0.101)	(0.100)	(0.093)	(0.148)
Nb. of Genders	0.270*	0.204	0.197	0.201	0.306	-0.044	-0.072	-0.073	-0.092	-0.007
	(0.148)	(0.132)	(0.133)	(0.136)	(0.198)	(0.080)	(0.084)	(0.083)	(0.074)	(0.078)
Assignment	-0.154	-0.086	-0.078	-0.076	0.028	0.019	0.056	0.054	0.043	0.063
	(0.133)	(0.169)	(0.168)	(0.166)	(0.181)	(0.040)	(0.061)	(0.060)	(0.058)	(0.070)
Pronouns	-0.236	-0.257	-0.257	-0.262	-0.301	0.021	-0.035	-0.036	-0.041	-0.178***
	(0.186)	(0.182)	(0.185)	(0.190)	(0.184)	(0.044)	(0.039)	(0.038)	(0.036)	(0.047)
Female	1.977***	1.954***	1.957***	1.953***	1.938***					
	(0.121)	(0.122)	(0.122)	(0.119)	(0.129)					
Years of Education		0.048***	0.048***	0.050***	0.050***		0.032***	0.032***	0.025***	0.025***
		(0.006)	(0.006)	(0.006)	(0.007)		(0.002)	(0.002)	(0.002)	(0.003)
Household Size			0.033**	0.038**	0.033**			-0.006	-0.030***	-0.026***
			(0.014)	(0.015)	(0.014)			(0.006)	(0.007)	(0.006)
Income Class				-0.009	-0.012				0.038***	0.038***
				(0.011)	(0.009)				(0.004)	(0.005)
Country Fixed Effects	√	√	√	√	√	√	√	√	√	√
Language Fixed Effects	√	√	√	√	√	√	√	√	√	√
Age	√	√	√	√	√	√	√	√	√	√
Happiness	√	√	√	√	√	√	√	√	√	√
Religious Denomination	√	√	√	√	√	√	√	√	√	√
Political Views	√	√	√	√	√	√	√	√	√	√
Urban/Rural	√	√	√	√	√	√	√	√	√	√
Cultural Values	×	×	×	×	√	×	×	×	×	√
Cultural Values Parents	×	×	×	×	√	×	×	×	×	√
Observations	9983	9973	9973	9973	8154	14331	14303	14303	14303	11266

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. SE in parenthesis. Dependent Variables: Question G38: Proportion of women at workplace. *Labour Market Participation*: Respondent was *in paid work* or *unemployed and actively looking for a job* in the last seven days. Independent Variables: *Sex-Based* is equal to 1 if a language has a sex-based gender system. *Nb. of Genders* is equal to 1 if a language only has two genders. *Assignment* is equal to 1 if a language has a semantic and formal assignment system. *Pronouns* is equal to 1 for languages with a gender distinction in third-person and in first and/or second-person. Missing observations are chained imputed.

The effect of language features related to gender on the labour market participation of women are depicted in table 4.7. A lower proportion of women at the workplace and a lower probability of active labour market participation by women is found for speakers of languages with a sex-based gender system. Speakers of such languages report on average a proportion of women at their workplace that is almost two categories lower than that of speakers of languages without a sex-based gender system. The same effect is found for gender in pronouns of a language. Having exactly two genders in a

language results in participants reporting a higher proportion of women at the workplace. It is however also associated with less active labour market participation for women. The system of gender assignment in a language has a positive effect on the labour market participation of women, both measured indirectly by the proportion of women reported and directly via the active labour market participation reported by female participants.

TABLE 4.8: Effect of Individual Language Features: Female Labour Market Outcomes II

	Female Occupation					Total Work Hours				
	I	II	III	IV	V	I	II	III	IV	V
Sex-Based	0.241***	0.259***	0.262***	0.289***	0.434***	-1.629	-13.194***	-13.202***	-13.727***	-12.333***
	(0.072)	(0.080)	(0.080)	(0.084)	(0.133)	(1.900)	(2.863)	(2.954)	(3.100)	(3.661)
Nb. of Genders	0.058	0.073	0.074	0.082	0.092	0.104	4.846**	4.555**	4.224**	3.671
	(0.060)	(0.065)	(0.065)	(0.065)	(0.085)	(1.384)	(1.835)	(1.842)	(1.834)	(2.560)
Assignment	0.038	0.026	0.029	0.033	0.009	2.667	0.814	0.396	0.430	0.788
	(0.035)	(0.037)	(0.037)	(0.036)	(0.043)	(1.639)	(2.473)	(2.423)	(2.472)	(1.243)
Pronouns	0.088**	0.110***	0.110***	0.112***	0.081	0.688	3.118***	3.366***	3.755***	4.064***
	(0.041)	(0.037)	(0.037)	(0.036)	(0.062)	(0.846)	(0.758)	(0.747)	(0.853)	(0.682)
Years of Education		-0.015***	-0.014***	-0.012***	-0.012***		0.246**	0.237**	0.105	0.058
		(0.002)	(0.002)	(0.002)	(0.002)		(0.108)	(0.111)	(0.095)	(0.114)
Household Size			0.010**	0.019***	0.016***			-0.906**	-1.357***	-1.356***
			(0.004)	(0.004)	(0.005)			(0.375)	(0.466)	(0.461)
Income Class				-0.015***	-0.013***				0.753***	0.722***
				(0.002)	(0.002)				(0.202)	(0.192)
Country Fixed Effects	√	√	√	√	√	√	√	√	√	√
Language Fixed Effects	√	√	√	√	√	√	√	√	√	√
Age	√	√	√	√	√	√	√	√	√	√
Happiness	√	√	√	√	√	√	√	√	√	√
Religious Denomination	√	√	√	√	√	√	√	√	√	√
Political Views	√	√	√	√	√	√	√	√	√	√
Urban/Rural	√	√	√	√	√	√	√	√	√	√
Cultural Values	×	×	×	×	√	×	×	×	×	√
Cultural Values Parents	×	×	×	×	√	×	×	×	×	√
Observations	14331	14303	14303	14303	11266	11426	5879	5879	5879	4782

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. SE in parenthesis. Dependent Variables: *Female Occupation*: The main occupation of the respondent is seen as traditionally more feminine. *Total Work Hours*: Counts the hours the respondent normally works per week, including paid and unpaid overtime. Only respondents with paid work in the last seven days are used. Independent Variables: *Sex-Based* is equal to 1 if a language has a sex-based gender system. *Nb. of Genders* is equal to 1 if a language only has two genders. *Assignment* is equal to 1 if a language has a semantic and formal assignment system. *Pronouns* is equal to 1 for languages with a gender distinction in third-person and in first and/or second-person. Missing observations are chained imputed.

Table 4.8 reports the results for the effects of the language features on the type of job women report as their current or last main occupation and on the number of work hours in the last seven days for women in paid work. Women who speak a language with a sex-based gender system have a higher probability to report an occupation as their main job which is traditionally seen as more feminine. The effect size is quite large with 0.4 and is significant at the 1% level. The coefficients of the three other language

features also have a positive sign but are not statistically significant. Additionally, a sex-based gender system also exhibits a strong effect on the work hours reported by women with paid work. On average, they report 12 hours less work than women who speak a language without a sex-based gender system. The pronoun variable on the other hand however has a positive effect. Women speaking a language with a gender distinction in third-person and in first and/or second-person report on average 4 work hours more. This effect is also statistically significant at the 1% level. The number of genders and the system of gender assignment also lead to a higher number of work hours reported but both effects are not statistically significant.

The last variable I look at in detail, is the number of hours of housework that women report. A sex-based gender system in a language has a positive effect on the amount of housework reported. The number of hours spent on housework increases on average by 1 bracket. This is in line with the previous results for the sex-based variable. It is associated with a higher level of agreement of individuals to the statement that women should cut down on their work for the sake of their family and a partial or complete reduction in labour market participation of women. Women seem to reduce their labour supply in favour of their family, as this is expected of them by the reinforced gender norms. The system of gender assignment also has a positive effect on the amount of housework reported. The number of genders and pronoun variables have a negative effect, but it is not statistically significant.

Throughout all of my regressions, the sex-based variable draws the most coherent picture. A sex-based gender system in a language leads to a strengthening of traditional gender norms in an individual. It implies that women should prioritise their family over their career and men have more rights to a job than women as work and income generation are not part of the traditional role of women in a family. This effect on gender norms is also reflected in real world labour market outcomes for women. It leads to a reduction in labour supply by women. They are less likely to actively participate in the labour market and if they do, they do it to a lesser extent, i.e., they report less working hours per week. However, not only the quantity of their labour supply is affected but also the quality. Women speaking a language with a sex-based gender system have a higher probability to report an occupation that is traditionally seen as more feminine

TABLE 4.9: Effect of Individual Language Features: Female Labour Market Outcomes III

	Housework Hours				
	I	II	III	IV	V
Sex-Based	-0.107	0.467*	0.911***	1.191***	0.931***
	(0.255)	(0.231)	(0.230)	(0.229)	(0.242)
Nb. of Genders	0.015	-0.089	-0.107	-0.058	-0.127
	(0.192)	(0.160)	(0.131)	(0.129)	(0.120)
Assignment	0.263***	0.126	0.188*	0.198*	0.315*
	(0.078)	(0.101)	(0.100)	(0.107)	(0.151)
Pronouns	-0.400	-0.224	-0.165	-0.172	-0.205
	(0.240)	(0.213)	(0.129)	(0.127)	(0.128)
Years of Education		-0.050***	-0.049***	-0.035***	-0.032***
		(0.006)	(0.006)	(0.007)	(0.008)
Household Size			0.237***	0.259***	0.254***
			(0.024)	(0.024)	(0.029)
Income Class				-0.069***	-0.064***
				(0.008)	(0.011)
Country Fixed Effects	√	√	√	√	√
Language Fixed Effects	√	√	√	√	√
Age	√	√	√	√	√
Happiness	√	√	√	√	√
Religious Denomination	√	√	√	√	√
Political Views	√	√	√	√	√
Urban/Rural	√	√	√	√	√
Cultural Values	×	×	×	×	√
Cultural Values Parents	×	×	×	×	√
Observations	7249	7239	7239	7239	5679

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. SE in parenthesis. Dependent Variables: *Housework Hours* is the total hours of housework reported for the last week. The reported values are grouped into brackets of 10 hours each. Last bracket contains all values above 100. Independent Variables: *Sex-Based* is equal to 1 if a language has a sex-based gender system. *Nb. of Genders* is equal to 1 if a language only has two genders. *Assignment* is equal to 1 if a language has a semantic and formal assignment system. *Pronouns* is equal to 1 for languages with a gender distinction in third-person and in first and/or second-person. Missing observations are chained imputed.

as their main job. Lastly, it also results on average in women reporting more hours of housework done. Taken together, this paints a picture that a sex-based gender system in a language leads to a more traditional role of women in the family.

The effects of the other three language features are less coherent, depending on the outcome variable, positive or negative effects on women can be found. A possible explanation for these ambiguous results is that the codification and their connection to

biological sex is not as clear as in the case of the sex-based variable. The English language for example knows three genders and therefore the number of genders variable is 0. The masculine and feminine gender however are exclusive to humans. It is a priori not so clear why the additional third gender, which is never associated with humans, should lead to different effects for women compared to French, which only has two genders. It might be the case that it is more important if the gender system itself is connected to biological sex than how much gender it expresses, and if genders are assigned to objects. Therefore, sex-based might be the variable to focus on to investigate the effect of gender in languages on labour market outcomes for women, due to its clear connection to biological sex.

Of the control variables that might be influenced by language structures themselves, the cultural variables in particular have a strong effect on the results of my sex-based variable. This is not surprising, as different cultures vary greatly in the role they assign to women in family and professional life. Therefore, it is particularly important to control for these cultural differences, even though links between my language and cultural variables cannot be ruled out and are to some extent likely. Nevertheless, my sex-based variable shows a robust and significant effect on gender norms and women's labour market outcomes even after controlling for cultural differences. This points towards a genuine and direct language effect.

4.6 Robustness Tests

4.6.1 Immigrants

I use country fixed effects to control for different institutional, legal, and socio-economic environments to separate the effects of language on beliefs and labour market outcomes from confounding institutional forces. In addition, I use a wide range of cultural variables to account for differences between cultures in the perceived role of women in family and professional life. However, I cannot completely rule out that there are additional unobserved variables that influence only natives. In particular, an influence on the labour market outcomes for native-born women might be possible. In this section, I only look at immigrants in my dataset to control for this potential unobserved

variables. I use two different definitions to determine who counts as a migrant.

TABLE 4.10: Effect of Individual Language Features: Immigrant Subsample I

	Cut Down Work	Scarce Jobs	Proportion Women	Labour Market Participation	Female Occupation	Total Work Hours	Housework Hours
	I	II	III	IV	V	VI	VII
Sex-Based	-0.570***	-1.171***	-0.584	-0.076	0.417***	-14.469***	1.439***
	(0.176)	(0.169)	(0.347)	(0.171)	(0.142)	(2.674)	(0.332)
Nb. of Genders	0.070	0.395***	0.614***	0.054	0.123	-1.435	-0.224
	(0.086)	(0.075)	(0.193)	(0.097)	(0.089)	(4.385)	(0.221)
Assignment	-0.190	-0.192*	-0.179	-0.006	-0.003	4.032**	0.084
	(0.125)	(0.106)	(0.161)	(0.063)	(0.057)	(1.460)	(0.204)
Pronouns	0.025	0.215	-0.452**	-0.196***	0.083	0.405	-0.128
	(0.098)	(0.126)	(0.188)	(0.055)	(0.088)	(1.488)	(0.144)
Female	0.117***	0.273***	2.022***				
	(0.033)	(0.051)	(0.253)				
Country Fixed Effects	√	√	√	√	√	√	√
Language Fixed Effects	√	√	√	√	√	√	√
Age	√	√	√	√	√	√	√
Happiness	√	√	√	√	√	√	√
Religious Denomination	√	√	√	√	√	√	√
Political Views	√	√	√	√	√	√	√
Urban/Rural	√	√	√	√	√	√	√
Education	√	√	√	√	√	√	√
Household Size	√	√	√	√	√	√	√
Household Income	√	√	√	√	√	√	√
Cultural Values	√	√	√	√	√	√	√
Cultural Values Parents	√	√	√	√	√	√	√
Observations	2185	2185	866	1297	1297	550	693

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. SE in parenthesis. Immigrant is anyone who states that they were not born in the country where they currently live. Dependent Variables: Question G4: A woman should be prepared to cut down on her paid work for the sake of her family. Question G5: When jobs are scarce, men should have more right to a job than women. The given answers are ordered from *Agree strongly* to *Disagree strongly*. Therefore, a negative coefficient is associated with a higher agreement with the statement. Question G38: Proportion of women at workplace. *Labour Market Participation*: Respondent was *in paid work* or *unemployed and actively looking for a job* in the last seven days. *Female Occupation*: The main occupation of the respondent is seen as traditionally more feminine. *Total Work Hours*: Counts the hours the respondent normally works per week, including paid and unpaid overtime. Only respondents with paid work in the last seven days are used. *Housework Hours* is the total hours of housework reported for the last week. The reported values are grouped into brackets of 10 hours each. Last bracket contains all values above 100. Independent Variables: *Sex-Based* is equal to 1 if a language has a sex-based gender system. *Nb. of Genders* is equal to 1 if a language only has two genders. *Assignment* is equal to 1 if a language has a semantic and formal assignment system. *Pronouns* is equal to 1 for languages with a gender distinction in third-person and in first and/or second-person. Missing observations are chained imputed.

First, I only include individuals who report not to be born in they country the currently live in. The results of these regressions are reported in table 4.10. The sex-based variable shows the same effects as before. It leads participants to continue to believe more in a role for women that focuses on family life, and it also has a negative impact on women's labour market outcomes.

TABLE 4.11: Effect of Individual Language Features: Immigrant Subsample II

	I	II	III	IV	V	VI	VII
	Cut Down Work	Scarce Jobs	Proportion Women	Labour Market Participation	Female Occupation	Total Work Hours	Housework Hours
Sex-Based	-0.557***	-1.154***	-1.340***	-0.096	0.444***	-14.194***	1.103***
	(0.170)	(0.113)	(0.398)	(0.137)	(0.145)	(4.090)	(0.346)
Nb. of Genders	0.058	0.308***	0.375*	0.014	0.115	2.729	-0.096
	(0.100)	(0.077)	(0.181)	(0.077)	(0.088)	(3.066)	(0.190)
Assignment	-0.113	-0.160	0.043	0.025	0.003	0.165	0.274
	(0.123)	(0.098)	(0.189)	(0.050)	(0.053)	(2.101)	(0.164)
Pronouns	-0.024	0.232***	-0.255	-0.158***	0.073	5.616***	-0.198*
	(0.058)	(0.067)	(0.155)	(0.043)	(0.071)	(1.166)	(0.101)
Female	0.133***	0.313***	2.029***				
	(0.026)	(0.038)	(0.209)				
Country Fixed Effects	√	√	√	√	√	√	√
Language Fixed Effects	√	√	√	√	√	√	√
Age	√	√	√	√	√	√	√
Happiness	√	√	√	√	√	√	√
Religious Denomination	√	√	√	√	√	√	√
Political Views	√	√	√	√	√	√	√
Urban/Rural	√	√	√	√	√	√	√
Education	√	√	√	√	√	√	√
Household Size	√	√	√	√	√	√	√
Household Income	√	√	√	√	√	√	√
Cultural Values	√	√	√	√	√	√	√
Cultural Values Parents	√	√	√	√	√	√	√
Observations	3996	3996	1596	2291	2291	991	1198

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. SE in parenthesis. Immigrant is anyone who was either not born in the country in which they currently live or anyone of whom at least one parent was not born in the country in which the respondent currently lives. Dependent Variables: Question G4: A woman should be prepared to cut down on her paid work for the sake of her family. Question G5: When jobs are scarce, men should have more right to a job than women. The given answers are ordered from *Agree strongly* to *Disagree strongly*. Therefore, a negative coefficient is associated with a higher agreement with the statement. Question G38: Proportion of women at workplace. *Labour Market Participation*: Respondent was *in paid work* or *unemployed and actively looking for a job* in the last seven days. *Female Occupation*: The main occupation of the respondent is seen as traditionally more feminine. *Total Work Hours*: Counts the hours the respondent normally works per week, including paid and unpaid overtime. Only respondents with paid work in the last seven days are used. *Housework Hours* is the total hours of housework reported for the last week. The reported values are grouped into brackets of 10 hours each. Last bracket contains all values above 100. Independent Variables: *Sex-Based* is equal to 1 if a language has a sex-based gender system. *Nb. of Genders* is equal to 1 if a language only has two genders. *Assignment* is equal to 1 if a language has a semantic and formal assignment system. *Pronouns* is equal to 1 for languages with a gender distinction in third-person and in first and/or second-person. Missing observations are chained imputed.

My second definition of immigrants uses a broader approach. Even when an individual is born in a country, it can be the case that they are not seen as native born because one or both of their parents were immigrants. Therefore, they may be treated as immigrants and the unobserved variables do not affect them in the same way as natives whose parents were also born in the country. To account for this possible effect in my second set of regressions, I define everyone as an immigrant who is not born in the country

they live in or at least one of their parents is not born there. I observe the same results for my sex-based variable again, which supports my findings on the effect of gender in language on the beliefs about the role of women, and labour market outcomes for women.

4.6.2 Wider Sample

As an additional robustness check, I combine the data from ESS round 5 with the data from rounds 6 to 9 (ESS Round 6, 2012; ESS Round 7, 2014; ESS Round 8, 2016; ESS Round 9, 2018) to increase my sample size. Unfortunately, the questions about an individual's beliefs about the role of a women, the proportion of women at the workplace and the hours of housework in the last week are not part of round 6 to 9 of the ESS as they are part of the rolling module about *work, family, and well-being*.

Columns I to III in table 4.12 report the results for the regressions with the complete sample. The results for the subsample with narrower and broader definitions for immigrants are shown in columns IV to VI and VII to IX, respectively. The results are in line with my previous findings. A sex-based gender system in a language is associated with lower active labour market participation among women. In addition, women are more likely to report an occupation, which is traditionally seen as more feminine as their main job. Finally, women who have paid jobs also report fewer working hours on average. These results also support the findings in my main regressions.

TABLE 4.12: Effect of Individual Language Features: Female Labour Market Outcomes ESS Round 5-9

	Full Sample			Immigrant I			Immigrant II		
	Labour Market Participation	Female Occupation	Total Work Hours	Labour Market Participation	Female Occupation	Total Work Hours	Labour Market Participation	Female Occupation	Total Work Hours
	I	II	III	I	II	III	I	II	III
Sex-Based	-0.132***	0.024	-10.288***	-0.039	0.155***	-13.224***	-0.082**	0.193***	-13.131***
	(0.030)	(0.038)	(1.021)	(0.045)	(0.045)	(1.536)	(0.038)	(0.046)	(1.506)
Nb. of Genders	-0.031	0.035	1.538	0.004	0.063*	-0.396	-0.018	0.057	0.039
	(0.022)	(0.029)	(1.838)	(0.034)	(0.031)	(1.789)	(0.025)	(0.035)	(1.964)
Assignment	0.062***	0.035	0.821	0.030	0.003	2.209*	0.053**	0.019	1.615
	(0.015)	(0.037)	(0.608)	(0.029)	(0.040)	(1.161)	(0.020)	(0.038)	(1.082)
Pronouns	-0.001	0.024	0.308	-0.029	0.021	2.362**	-0.026	0.017	2.446*
	(0.048)	(0.039)	(0.906)	(0.067)	(0.032)	(1.036)	(0.054)	(0.039)	(1.261)
Country Fixed Effects	√	√	√	√	√	√	√	√	√
Language Fixed Effects	√	√	√	√	√	√	√	√	√
Age	√	√	√	√	√	√	√	√	√
Happiness	√	√	√	√	√	√	√	√	√
Religious Denomination	√	√	√	√	√	√	√	√	√
Political Views	√	√	√	√	√	√	√	√	√
Urban/Rural	√	√	√	√	√	√	√	√	√
Education	√	√	√	√	√	√	√	√	√
Household Size	√	√	√	√	√	√	√	√	√
Household Income	√	√	√	√	√	√	√	√	√
Cultural Values	√	√	√	√	√	√	√	√	√
Cultural Values Parents	√	√	√	√	√	√	√	√	√
Observations	48779	48779	23704	5625	5625	2587	10031	10031	4964

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. SE in parenthesis. In columns IV-VI an immigrant is anyone who states that they were not born in the country where they currently live. In columns VII-IX an immigrant is anyone who was either not born in the country in which they currently live or anyone of whom at least one parent was not born in the country in which the respondent currently lives. Dependent Variables: *Labour Market Participation*: Respondent was *in paid work* or *unemployed and actively looking for a job* in the last seven days. *Female Occupation*: The main occupation of the respondent is seen as traditionally more feminine. *Total Work Hours*: Counts the hours the respondent normally works per week, including paid and unpaid overtime. Only respondents with paid work in the last seven days are used. Independent Variables: *Sex-Based* is equal to 1 if a language has a sex-based gender system. *Nb. of Genders* is equal to 1 if a language only has two genders. *Assignment* is equal to 1 if a language has a semantic and formal assignment system. *Pronouns* is equal to 1 for languages with a gender distinction in third-person and in first and/or second-person. Missing observations are chained imputed.

4.7 Conclusion

Based on the theory of linguistic relativity, which states that grammatical and linguistic characteristics of a language can influence the thoughts and behaviour of its speakers, I show in this chapter that language has an effect on people's beliefs regarding the role of women in the working world and women's working lives. Thus, my work contributes to the existing literature on the relationship between grammatical characteristics of languages and people's behaviour by examining people's opinions towards the role of women in the conflict between work and family and in comparison, to men and actual labour market outcomes such as participation, weekly working hours and type of occupation.

I use data from ESS Round 5 for my analysis, which asks participants about their opinion whether a woman should prioritise her family over her job, and whether men should enjoy privileges over women when jobs are scarce. In addition, it also contains data on the real labour market outcomes for female participants. These are combined with data on the gender system of languages from the WALS. Using these datasets, I show that the gender system in a language has an impact on a person's attitude and labour market outcomes for women.

A higher GII is associated with a higher level of agreement with traditional gender roles and with worse labour market outcomes for women, i.e., lower labour market participation, lower working hours and different types of occupation. When I take a closer look at the language features that make up the GII, I find that especially the question of whether the gender system of a language is based on biological sex has a negative effect on women. A sex-based gender system leads its speakers to a higher level of agreement with the statements that a woman should be prepared to cut down on work in favour of family and that men should have more rights to a job when these are scarce. However, it does not only affect an individual's beliefs about the role of women in family and work life but it also has effect on the labour market outcomes for women. A sex-based gender system in a language leads to a lower labour market participation of women and those that still work, report fewer working hours per week. Furthermore, it does not only influence the quantity of women's labour supply, but also the quality. Women who speak a language with a sex-based gender system are

more likely to work in a profession that is traditionally seen as feminine. In support of these findings about the beliefs about a woman's role and their labour supply, I additionally find a sex-based gender system results in a higher number of reported hours of housework by women.

The connection with the other three language features, number of genders, system of gender assignment and gender in pronouns, is ambiguous. One possible reason for this could be that these three characteristics are not directly linked to biological sex and therefore coding them into a binary variable is not straightforward and leaves room for discussion.

The effects for the sex-based gender system are robust to a number of control variables and robustness tests. I use a within country design for my analysis to compare people who live in the same institutional and social environment but speak different languages that differ in how they assign gender. Although it can never be completely ruled out that language only represents a deeper cultural effect, my results are stable even with the addition of a wide range of cultural measures. This applies both to the cultural values of the participants' country of birth and to the cultural values of the parents that they can pass on to their children. I find the same results in a subsample that contains only immigrants, in order to account for possible unobserved variables that only affect natives and are not represented by the country fixed effects or cultural variables. In summary, my observed effects are real language effects and show how features of languages affect women's lives.

My findings show language has an effect on the attitudes and behaviour of its speakers through the spotlight it shines on certain circumstances in the world through grammatical idiosyncrasies. The gender system of a language and especially if this system is linked to biological sex has an impact on women's labour market outcomes and on gender norms in society and must therefore be considered by policy makers who want to change the situation of women in the labour market.

Appendix A

Language and the Future

A.1 Company descriptives

TABLE A.1: Observations per Country (R&D expenditures)

	Overall		Between	
Country	Frequency	Percentage	Frequency	Percentage
Austria	10	0.21	2	0.23
Belgium	67	1.39	14	1.61
Denmark	40	0.83	12	1.38
France	2084	43.15	358	41.24
Germany	537	11.12	89	10.25
Ireland	31	0.64	7	0.81
Luxembourg	12	0.25	3	0.35
Netherlands	1	0.02	1	0.12
Sweden	270	5.59	61	7.03
United Kingdom	1778	36.81	321	36.98
Total	4830	100.00	868	100.00

TABLE A.2: Observations per Sector (R&D Expenditures)

	Overall		Between	
Sector	Frequency	Percentage	Frequency	Percentage
Aerospace & Defence	108	2.24	16	1.84
Automobiles & Parts	151	3.13	22	2.53
Beverages	69	1.43	10	1.15
Business Services	202	4.18	39	4.49
Chemicals	215	4.45	33	3.80
Clothing & Personal Products	123	2.55	21	2.42
Construction & Building Materials	204	4.22	31	3.57
Consumer Services	11	0.23	3	0.35
Diversified Industrials	73	1.51	14	1.61
Electricity	42	0.87	5	0.58
Electronic & Electrical Equipment	271	5.61	49	5.65
Engineering & Machinery	311	6.44	54	6.22
Food & Drug Retailers	48	0.99	7	0.81
Food Producers & Processors	175	3.62	29	3.34
Forestry & Paper	15	0.31	2	0.23
Health	325	6.73	61	7.03
Household Products	83	1.72	16	1.84
Information Technology Hardware	217	4.49	43	4.95
Leisure Goods	39	0.81	8	0.92
Leisure & Hotels	100	2.07	16	1.84
Media & Entertainment	180	3.73	35	4.03
Steel & Other Metals	45	0.93	7	0.81
Mining	34	0.70	5	0.58
Oil & Gas	97	2.01	19	2.19
Containers & Packaging	64	1.33	9	1.04
Pharmaceuticals & Biotechnology	423	8.76	94	10.83
Real Estate	150	3.11	30	3.46
General Retailers	77	1.59	13	1.50
Renewable Energy	76	1.57	18	2.07
Software & Computer Services	645	13.35	120	13.82
Telecommunication Services	93	1.93	17	1.96
Tobacco	11	0.23	2	0.23
Transport	92	1.90	12	1.38
Utilities	61	1.26	8	0.92
Total	4830	100.00	868	100.00

TABLE A.3: Observations per Country (Intangible Assets Growth Rate)

	Overall		Between	
Country	Frequency	Percentage	Frequency	Percentage
Austria	75	0.75	16	0.84
Belgium	171	1.72	36	1.90
Croatia	12	0.12	2	0.11
Czech Republic	11	0.11	2	0.11
Denmark	56	0.56	24	1.27
Finland	198	1.99	36	1.90
France	1683	16.88	323	17.04
Germany	1142	11.46	233	12.30
Gibraltar	7	0.07	1	0.05
Greece	114	1.14	19	1.00
Hungary	11	0.11	4	0.21
Iceland	1	0.01	1	0.05
Ireland	105	1.05	23	1.21
Italy	437	4.38	86	4.54
Luxembourg	37	0.37	9	0.47
Monaco	2	0.02	1	0.05
Netherlands	99	0.99	31	1.64
Norway	115	1.15	30	1.58
Poland	9	0.09	2	0.11
Portugal	124	1.24	25	1.32
Russia	49	0.49	11	0.58
Spain	451	4.52	82	4.33
Sweden	533	5.35	94	4.96
United Kingdom	4526	45.41	804	42.43
Total	9968	100.00	1895	100.00

TABLE A.4: Observations per Sector (Intangible Assets Growth Rate)

	Overall		Between	
Sector	Frequency	Percentage	Frequency	Percentage
Aerospace & Defence	116	1.16	17	0.90
Automobiles & Parts	235	2.36	38	2.01
Beverages	137	1.37	28	1.48
Business Services	724	7.26	127	6.70
Chemicals	253	2.54	47	2.48
Clothing & Personal Products	197	1.98	38	2.01
Construction & Building Materials	599	6.01	94	4.96
Consumer Services	32	0.32	8	0.42
Diversified Industrials	221	2.22	36	1.90
Education	1	0.01	1	0.05
Electricity	163	1.64	27	1.42
Electronic & Electrical Equipment	388	3.89	67	3.54
Engineering & Machinery	552	5.54	99	5.22
Food & Drug Retailers	134	1.34	24	1.27
Food Producers & Processors	273	2.74	51	2.69
Forestry & Paper	87	0.87	16	0.84
Health	377	3.78	83	4.38
Household Products	166	1.67	29	1.53
Information Technology Hardware	211	2.21	46	2.43
Leisure Goods	58	0.58	12	0.63
Leisure & Hotels	438	4.39	82	4.33
Media & Entertainment	641	6.43	114	6.02
Steel & Other Metals	131	1.31	26	1.37
Mining	185	1.86	47	2.48
Oil & Gas	500	5.02	99	5.22
Containers & Packaging	77	0.77	14	0.74
Pharmaceuticals & Biotechnology	510	5.12	115	6.07
Publishing	24	0.24	4	0.21
Real Estate	415	4.16	89	4.70
General Retailers	343	3.44	66	3.48
Renewable Energy	178	1.79	38	2.01
Software & Computer Services	882	8.85	175	9.23
Telecommunication Services	291	2.92	60	3.17
Tobacco	25	0.25	3	0.16
Transport	255	2.56	53	2.80
Utilities	148	1.48	21	1.11
Wholesale Trade	1	0.01	1	0.05
Total	9968	100.00	1895	100.00

A.2 Additional Regressions

A.2.1 Intangible Asset Growth Rate

TABLE A.5: Effect on Intangible Asset Growth Rate $< 200\%$: Mean

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	-0.002	0.029	0.039	0.036	0.037	0.038	0.029	0.029
	(0.034)	(0.037)	(0.036)	(0.036)	(0.036)	(0.036)	(0.038)	(0.042)
Inflectional-FTR	-0.032	-0.088	-0.142	-0.157	-0.151	-0.149	-0.159	-0.159
	(0.036)	(0.190)	(0.192)	(0.190)	(0.189)	(0.189)	(0.190)	(0.193)
Financials								
EBIT Margin			0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share			-0.010	-0.010	-0.011	-0.010	-0.014	-0.014
			(0.021)	(0.021)	(0.021)	(0.021)	(0.022)	(0.023)
Mean Age				-0.002*	-0.002**	-0.002**	-0.002**	-0.002**
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female Share				0.012	0.015	0.013	0.013	0.013
				(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Enter					-0.043*	-0.043*	-0.043*	-0.043*
					(0.025)	(0.026)	(0.025)	(0.025)
Exit					-0.067**	-0.066**	-0.067**	-0.067**
					(0.032)	(0.032)	(0.032)	(0.032)
Total Assets						0.000	0.000	0.000
						(0.000)	(0.000)	(0.000)
Cultur								
Patience							-0.058	-0.058
							(0.082)	(0.094)
Risk Taking								-0.000
								(0.199)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	10262	10262	10262	10262	10262	10262	10262	10262
Clusters	1911	1911	1911	1911	1911	1911	1911	1911

Note: All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Growth rates in the regressions are restricted to be smaller than 200%. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

TABLE A.6: Effect on Intangible Asset Growth Rate $< 150\%$: Mean

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	-0.002	0.037	0.046	0.044	0.045	0.046	0.047	0.036
	(0.033)	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)	(0.037)	(0.040)
Inflectional-FTR	-0.049	-0.055	-0.139	-0.148	-0.143	-0.141	-0.141	-0.156
	(0.034)	(0.196)	(0.211)	(0.211)	(0.209)	(0.209)	(0.210)	(0.216)
Financials								
EBIT Margin			0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share			-0.004	-0.004	-0.005	-0.004	-0.004	-0.002
			(0.019)	(0.019)	(0.019)	(0.019)	(0.020)	(0.020)
Mean Age				-0.001	-0.001*	-0.001*	-0.001*	-0.001*
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female Share				0.016	0.019	0.016	0.016	0.016
				(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Enter					-0.039*	-0.039*	-0.039*	-0.039*
					(0.023)	(0.023)	(0.023)	(0.023)
Exit					-0.056*	-0.055*	-0.055*	-0.055*
					(0.029)	(0.029)	(0.029)	(0.029)
Total Assets						0.000*	0.000*	0.000*
						(0.000)	(0.000)	(0.000)
Cultur								
Patience							0.003	-0.036
							(0.056)	(0.078)
Risk Taking								0.114
								(0.168)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	10163	10163	10163	10163	10163	10163	10163	10163
Clusters	1906	1906	1906	1906	1906	1906	1906	1906

Note: All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Growth rates in the regressions are restricted to be smaller than 150%. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

TABLE A.7: Effect on Intangible Asset Growth Rate $< 75\%$: Mean

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	-0.011	0.014	0.020	0.020	0.021	0.023	0.025	0.014
	(0.029)	(0.030)	(0.031)	(0.031)	(0.031)	(0.031)	(0.033)	(0.035)
Inflectional-FTR	-0.028	0.017	-0.044	-0.046	-0.040	-0.038	-0.036	-0.050
	(0.029)	(0.163)	(0.155)	(0.156)	(0.156)	(0.155)	(0.155)	(0.157)
Financials								
EBIT Margin			0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share			-0.013	-0.012	-0.014	-0.012	-0.011	-0.009
			(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Mean Age				-0.000	-0.001	-0.001	-0.001	-0.001
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female Share				0.027	0.030	0.026	0.026	0.025
				(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Enter					-0.039**	-0.039**	-0.039**	-0.039**
					(0.018)	(0.018)	(0.018)	(0.018)
Exit					-0.037	-0.036	-0.036	-0.036
					(0.023)	(0.023)	(0.023)	(0.023)
Total Assets						0.000***	0.000***	0.000***
						(0.000)	(0.000)	(0.000)
Cultur								
Patience							0.009	-0.027
							(0.049)	(0.066)
Risk Taking								0.103
								(0.152)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	9774	9774	9774	9774	9774	9774	9774	9774
Clusters	1888	1888	1888	1888	1888	1888	1888	1888

Note: All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Growth rates in the regressions are restricted to be smaller than 75%. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

TABLE A.8: Effect on Intangible Asset Growth Rate $< 200\%$: Median

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	0.038	0.064*	0.067**	0.066*	0.067**	0.068**	0.064*	0.065*
	(0.031)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.035)	(0.036)
Inflectional-FTR	-0.084***	-0.119***	-0.128***	-0.125***	-0.123***	-0.123***	-0.120***	-0.121***
	(0.032)	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)	(0.046)	(0.045)
Financials								
EBIT Margin			0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share			-0.006	-0.006	-0.007	-0.006	-0.009	-0.010
			(0.021)	(0.021)	(0.021)	(0.021)	(0.022)	(0.023)
Mean Age				-0.002*	-0.002**	-0.002**	-0.002**	-0.002**
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female Share				0.012	0.014	0.012	0.012	0.012
				(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Enter					-0.043*	-0.043*	-0.043*	-0.043*
					(0.025)	(0.025)	(0.025)	(0.025)
Exit					-0.067**	-0.066**	-0.067**	-0.067**
					(0.031)	(0.031)	(0.031)	(0.031)
Total Assets						0.000	0.000	0.000
						(0.000)	(0.000)	(0.000)
Cultur								
Patience							-0.049	-0.041
							(0.079)	(0.085)
Risk Taking								-0.025
								(0.187)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	10262	10262	10262	10262	10262	10262	10262	10262
Clusters	1911	1911	1911	1911	1911	1911	1911	1911

Note: All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Growth rates in the regressions are restricted to be smaller than 200%. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

TABLE A.9: Effect on Intangible Asset Growth Rate $< 150\%$: Median

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	0.035	0.068**	0.072**	0.072**	0.072**	0.073**	0.074**	0.068**
	(0.030)	(0.032)	(0.032)	(0.032)	(0.033)	(0.033)	(0.033)	(0.034)
Inflectional-FTR	-0.095***	-0.135***	-0.143***	-0.141***	-0.139***	-0.138***	-0.139***	-0.138***
	(0.032)	(0.044)	(0.045)	(0.044)	(0.044)	(0.045)	(0.044)	(0.044)
Financials								
EBIT Margin			0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share			-0.000	-0.000	-0.001	0.000	0.000	0.002
			(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Mean Age				-0.001	-0.001*	-0.001*	-0.001*	-0.001*
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female Share				0.015	0.018	0.015	0.015	0.015
				(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Enter					-0.039*	-0.039*	-0.039*	-0.039*
					(0.023)	(0.023)	(0.023)	(0.023)
Exit					-0.056*	-0.055*	-0.055*	-0.055*
					(0.029)	(0.029)	(0.029)	(0.029)
Total Assets						0.000*	0.000*	0.000*
						(0.000)	(0.000)	(0.000)
Cultur								
Patience							0.008	-0.023
							(0.054)	(0.070)
Risk Taking								0.098
								(0.155)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	10163	10163	10163	10163	10163	10163	10163	10163
Clusters	1906	1906	1906	1906	1906	1906	1906	1906

Note: All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Growth rates in the regressions are restricted to be smaller than 150%. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

TABLE A.10: Effect on Intangible Asset Growth Rate $< 75\%$: Median

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	0.027	0.049*	0.052*	0.052*	0.052*	0.054*	0.055*	0.051*
	(0.027)	(0.028)	(0.029)	(0.029)	(0.029)	(0.029)	(0.030)	(0.031)
Inflectional-FTR	-0.069***	-0.103***	-0.108***	-0.107***	-0.105***	-0.104***	-0.105***	-0.104***
	(0.026)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
Financials								
EBIT Margin			0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share			-0.009	-0.009	-0.010	-0.008	-0.007	-0.006
			(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)
Mean Age				-0.000	-0.001	-0.001	-0.001	-0.001
				(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female Share				0.027	0.030	0.025	0.025	0.025
				(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Enter					-0.038**	-0.038**	-0.039**	-0.039**
					(0.018)	(0.018)	(0.018)	(0.018)
Exit					-0.037	-0.037	-0.036	-0.036
					(0.023)	(0.023)	(0.023)	(0.023)
Total Assets						0.000***	0.000***	0.000***
						(0.000)	(0.000)	(0.000)
Cultur								
Patience							0.015	-0.010
							(0.047)	(0.060)
Risk Taking								0.079
								(0.142)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	9774	9774	9774	9774	9774	9774	9774	9774
Clusters	1888	1888	1888	1888	1888	1888	1888	1888

Note: All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Growth rates in the regressions are restricted to be smaller than 75%. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

A.2.2 R&D Expenditures

TABLE A.11: Effect on R&D Expenditures Proportions < 1: Mean

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	0.015	0.021	0.012	0.009	0.009	0.008	0.014	-0.005
	(0.015)	(0.016)	(0.017)	(0.018)	(0.018)	(0.018)	(0.017)	(0.020)
Inflectional-FTR	-0.013	0.206***	0.189	0.219	0.229	0.237	0.246	0.233
	(0.015)	(0.041)	(0.152)	(0.162)	(0.164)	(0.164)	(0.164)	(0.165)
Financials								
EBIT Margin			0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share				-0.008	-0.007	-0.008	-0.006	-0.003
				(0.012)	(0.012)	(0.012)	(0.011)	(0.011)
Mean Age				-0.000	-0.000	-0.000	-0.000	-0.000
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female Share				-0.018	-0.018	-0.018	-0.017	-0.018
				(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Enter					0.005	0.005	0.005	0.005
					(0.010)	(0.010)	(0.010)	(0.010)
Exit					-0.019*	-0.019*	-0.019	-0.019*
					(0.011)	(0.011)	(0.011)	(0.011)
Total Assets						-0.000	-0.000	-0.000
						(0.000)	(0.000)	(0.000)
Cultur								
Patience							0.060*	-0.003
							(0.034)	(0.044)
Risk Taking								0.220**
								(0.107)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	4910	4910	4910	4910	4910	4910	4910	4910
Clusters	873	873	873	873	873	873	873	873

Note: All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Negative proportions and proportions over 1 are excluded from the regression. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

TABLE A.12: Effect on R&D Expenditures Proportions < 0.75: Mean

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	0.021*	0.028**	0.020	0.018	0.018	0.017	0.022	0.007
	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.015)
Inflectional-FTR	-0.023**	0.163***	0.353***	0.365***	0.365***	0.373***	0.371***	0.363***
	(0.011)	(0.035)	(0.127)	(0.135)	(0.136)	(0.135)	(0.135)	(0.135)
Financials								
EBIT Margin			0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share				-0.001	-0.001	-0.002	-0.000	0.002
				(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
Mean Age				-0.000	-0.000	-0.000	-0.000	-0.000
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female Share				-0.020*	-0.019*	-0.019*	-0.018*	-0.019*
				(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Enter					-0.002	-0.002	-0.003	-0.002
					(0.007)	(0.007)	(0.007)	(0.007)
Exit					-0.007	-0.007	-0.006	-0.007
					(0.008)	(0.008)	(0.008)	(0.008)
Total Assets						-0.000**	-0.000**	-0.000**
						(0.000)	(0.000)	(0.000)
Cultur								
Patience							0.048	-0.000
							(0.032)	(0.045)
Risk Taking								0.167
								(0.106)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	4885	4885	4885	4885	4885	4885	4885	4885
Clusters	872	872	872	872	872	872	872	872

Note: All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Negative proportions and proportions over 0.75 are excluded from the regression. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

TABLE A.13: Effect on R&D Expenditures Proportions < 0.25: Mean

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	0.007	0.007	0.002	0.002	0.002	0.001	0.004	-0.004
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.011)
Inflectional-FTR	-0.021**	0.048**	0.162**	0.161**	0.159**	0.159**	0.148**	0.145*
	(0.009)	(0.022)	(0.070)	(0.075)	(0.075)	(0.074)	(0.074)	(0.074)
Financials								
EBIT Margin			0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share				0.001	0.001	0.001	0.002	0.003
				(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Mean Age				-0.000	-0.000	-0.000	-0.000	-0.000
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female Share				-0.006	-0.005	-0.005	-0.005	-0.005
				(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Enter					-0.004	-0.004	-0.004	-0.004
					(0.004)	(0.004)	(0.004)	(0.004)
Exit					-0.001	-0.001	-0.001	-0.001
					(0.004)	(0.004)	(0.004)	(0.004)
Total Assets						-0.000***	-0.000***	-0.000***
						(0.000)	(0.000)	(0.000)
Cultur								
Patience							0.036	0.009
							(0.026)	(0.037)
Risk Taking								0.095
								(0.084)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	4659	4659	4659	4659	4659	4659	4659	4659
Clusters	854	854	854	854	854	854	854	854

Note: All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Negative proportions and proportions over 0.25 are excluded from the regression. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

TABLE A.14: Effect on R&D Expenditures Median Proportions < 1: Median

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	0.006	0.009	0.005	0.003	0.002	0.002	0.004	-0.010
	(0.012)	(0.012)	(0.015)	(0.016)	(0.016)	(0.016)	(0.016)	(0.018)
Inflectional-FTR	-0.001	0.005	0.007	0.011	0.011	0.011	0.010	0.016
	(0.011)	(0.012)	(0.013)	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)
Financials								
EBIT Margin			0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share				-0.008	-0.008	-0.008	-0.007	-0.004
				(0.012)	(0.012)	(0.012)	(0.011)	(0.011)
Mean Age				-0.000	-0.000	-0.000	-0.000	-0.000
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female Share				-0.018	-0.019	-0.018	-0.018	-0.019
				(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Enter					0.005	0.005	0.005	0.005
					(0.010)	(0.010)	(0.010)	(0.010)
Exit					-0.019*	-0.019*	-0.019	-0.019*
					(0.011)	(0.011)	(0.011)	(0.011)
Total Assets						-0.000	-0.000	-0.000
						(0.000)	(0.000)	(0.000)
Cultur								
Patience							0.056	-0.006
							(0.034)	(0.044)
Risk Taking								0.232**
								(0.100)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	4910	4910	4910	4910	4910	4910	4910	4910
Clusters	873	873	873	873	873	873	873	873

Note: All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Negative proportions and proportions over 1 are excluded from the regression. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

TABLE A.15: Effect on R&D Expenditures Median Proportions < 0.75: Median

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	0.013	0.016	0.011	0.010	0.010	0.010	0.012	-0.000
	(0.011)	(0.012)	(0.014)	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)
Inflectional-FTR	-0.008	0.002	0.005	0.007	0.007	0.008	0.007	0.012
	(0.009)	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)
Financials								
EBIT Margin			0.000	0.000	0.000	0.000	0.000	0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share				-0.001	-0.001	-0.001	-0.000	0.002
				(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
Mean Age				-0.000	-0.000	-0.000	-0.000	-0.000
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female Share				-0.021*	-0.020*	-0.019*	-0.019*	-0.020*
				(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Enter					-0.002	-0.002	-0.002	-0.002
					(0.007)	(0.007)	(0.007)	(0.007)
Exit					-0.007	-0.007	-0.006	-0.007
					(0.008)	(0.008)	(0.008)	(0.008)
Total Assets						-0.000**	-0.000**	-0.000**
						(0.000)	(0.000)	(0.000)
Cultur								
Patience							0.042	-0.010
							(0.032)	(0.041)
Risk Taking								0.194*
								(0.100)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	4885	4885	4885	4885	4885	4885	4885	4885
Clusters	872	872	872	872	872	872	872	872

Note: All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Negative proportions and proportions over 0.75 are excluded from the regression. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

TABLE A.16: Effect on R&D Expenditures Median Proportions < 0.25: Median

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Strong-FTR	0.002	0.002	-0.003	-0.003	-0.002	-0.003	-0.002	-0.008
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Inflectional-FTR	-0.015*	-0.009	-0.004	-0.004	-0.004	-0.004	-0.004	-0.002
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)
Financials								
EBIT Margin			0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
EBITDA Margin			-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Board Characteristics								
Country Share				0.001	0.001	0.001	0.002	0.003
				(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
Mean Age				-0.000	-0.000	-0.000	-0.000	-0.000
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female Share				-0.006	-0.005	-0.005	-0.005	-0.005
				(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Enter					-0.004	-0.004	-0.004	-0.004
					(0.004)	(0.004)	(0.004)	(0.004)
Exit					-0.001	-0.001	-0.001	-0.001
					(0.004)	(0.004)	(0.004)	(0.004)
Total Assets						-0.000***	-0.000***	-0.000***
						(0.000)	(0.000)	(0.000)
Cultur								
Patience							0.035	0.007
							(0.025)	(0.034)
Risk Taking								0.104
								(0.075)
Country	√	√	√	√	√	√	√	√
Country × Year	√	√	√	√	√	√	√	√
Language Family	×	√	√	√	√	√	√	√
Language Genus	×	√	√	√	√	√	√	√
Sector Fixed Effects	×	×	√	√	√	√	√	√
Sector × Year	×	×	√	√	√	√	√	√
Attrition Control	√	√	√	√	√	√	√	√
Observations	4659	4659	4659	4659	4659	4659	4659	4659
Clusters	854	854	854	854	854	854	854	854

Note: All Regressions are random effect models with the Swamy-Arora estimator of the variance components. Standard errors clustered at company level are reported in parenthesis. Companies from the financial sector are excluded from all regressions. Negative proportions and proportions over 0.25 are excluded from the regression. *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level.

Appendix B

Language and Xenophobia

B.1 Regressions with SE Clustered at Country Level

TABLE B.1: Language Effect on Trust: SE clustered at country level

	I	II	III	Do not trust at all	Do not trust very much	Trust somewhat	Trust completely
Politeness Distinction	-0.442**	-0.441**	-0.293*	0.079*	0.018	-0.074*	-0.024*
	(0.205)	(0.198)	(0.160)	(0.042)	(0.014)	(0.041)	(0.013)
Trust			0.435***	-0.117***	-0.027**	0.109***	0.035***
			(0.031)	(0.008)	(0.012)	(0.009)	(0.005)
Immigrant			0.131***	-0.035***	-0.008**	0.033***	0.011***
			(0.043)	(0.013)	(0.004)	(0.011)	(0.003)
Immigration Status Parents							
One Immigrant			0.127***	-0.033***	-0.010**	0.032***	0.011***
			(0.028)	(0.007)	(0.004)	(0.007)	(0.003)
Both Immigrants			0.028	-0.008	-0.002	0.007	0.002
			(0.047)	(0.012)	(0.003)	(0.012)	(0.004)
Country Fixed Effects	√	√	√	√	√	√	√
Language Family	√	√	√	√	√	√	√
Language Genus	√	√	√	√	√	√	√
Personal Characteristics	×	√	√	√	√	√	√
Employment Status	×	×	√	√	√	√	√
Financial Situation	×	×	√	√	√	√	√
Family Situation	×	×	√	√	√	√	√
Political Views	×	×	√	√	√	√	√
Observations	41152	41152	41152	41152	41152	41152	41152

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. Column I to III report coefficients for ordered probit regressions with different sets of control variables. Column IV to VII report the marginal effects for each of the four different possible answers given for question V107 of the WVS. The same set of control variables is used as in column III. Standard errors are clustered at the country level.

TABLE B.2: Language Effect on Trust: Cultural Dimensions and SE clustered at country level

	I	Do not trust at all	Do not trust very much	Trust somewhat	Trust completely
Politeness Distinction	-0.333**	0.092**	0.021	-0.085**	-0.027***
	(0.130)	(0.036)	(0.013)	(0.035)	(0.010)
Trust	0.416***	-0.115***	-0.026**	0.107***	0.034***
	(0.033)	(0.010)	(0.012)	(0.009)	(0.005)
Culture					
Masculinity	-0.003	0.001	0.000	-0.001	-0.000
	(0.003)	(0.001)	(0.000)	(0.001)	(0.000)
Power Distance	-0.012***	0.003***	0.001*	-0.003***	-0.001***
	(0.003)	(0.001)	(0.000)	(0.001)	(0.000)
Immigrant	0.122***	-0.034**	-0.008**	0.031**	0.010***
	(0.047)	(0.014)	(0.004)	(0.012)	(0.004)
Immigration Status Parents					
One Immigrant	0.121***	-0.032***	-0.009**	0.031***	0.011***
	(0.033)	(0.009)	(0.004)	(0.008)	(0.003)
Both Immigrants	0.023	-0.006	-0.001	0.006	0.002
	(0.050)	(0.014)	(0.003)	(0.013)	(0.004)
Country Fixed Effects	×	×	×	×	×
Language Family	√	√	√	√	√
Language Genus	√	√	√	√	√
Personal Characteristics	√	√	√	√	√
Employment Status	√	√	√	√	√
Financial Situation	√	√	√	√	√
Family Situation	√	√	√	√	√
Political Views	√	√	√	√	√
Observations	41152	41152	41152	41152	41152

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. Column I reports coefficients for ordered probit regressions. Column II to V report the marginal effects for each of the four different possible answers given for question V107 of the WVS. The same set of control variables is used as in column III. Hofstede's cultural dimensions (Hofstede, 2001) are used to control for for cultural characteristics and differences of societies. Standard errors are clustered at the country level.

B.2 Marginal Effects

B.2.1 Immigrants

TABLE B.3: Language Effect on Trust: Immigrants (Marginal Effects)

	I	II	III	Do not trust at all	Do not trust very much	Trust somewhat	Trust completely
Politeness Distinction	-0.456**	-0.489**	-0.364**	0.071**	0.052**	-0.081**	-0.041**
	(0.187)	(0.191)	(0.150)	(0.035)	(0.020)	(0.037)	(0.017)
Trust			0.490***	-0.095***	-0.069***	0.109***	0.056***
			(0.040)	(0.014)	(0.016)	(0.014)	(0.008)
Country Fixed Effects	√	√	√	√	√	√	√
Language Family	√	√	√	√	√	√	√
Language Genus	√	√	√	√	√	√	√
Personal Characteristics	×	√	√	√	√	√	√
Immigration History	×	×	√	√	√	√	√
Employment Status	×	×	√	√	√	√	√
Financial Situation	×	×	√	√	√	√	√
Family Situation	×	×	√	√	√	√	√
Political Views	×	×	√	√	√	√	√
Observations	4957	4957	4957	4957	4957	4957	4957

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. For all regressions, only people who report being immigrants themselves or report at least one parent as an immigrant are included (Questions V243-245 in WVS). Column I to III report coefficients for ordered probit regressions with different sets of control variables. Column IV to VII report the marginal effects for each of the four different possible answers given for question V107 of the WVS. The same set of control variables is used as in column III.

TABLE B.4: Language Effect on Trust: Immigrants and Cultural Dimensions (Marginal Effects)

	I	Do not trust at all	Do not trust very much	Trust somewhat	Trust completely
Politeness Distinction	-0.242**	0.048**	0.036**	-0.056**	-0.028**
	(0.103)	(0.021)	(0.017)	(0.026)	(0.011)
Trust	0.503***	-0.101***	-0.074***	0.116***	0.059***
	(0.037)	(0.015)	(0.015)	(0.014)	(0.008)
Culture					
Masculinity	-0.000	0.000	0.000	-0.000	-0.000
	(0.004)	(0.001)	(0.001)	(0.001)	(0.000)
Power Distance	-0.010***	0.002***	0.001***	-0.002***	-0.001***
	(0.002)	(0.001)	(0.000)	(0.001)	(0.000)
Country Fixed Effects	×	×	×	×	×
Language Family	√	√	√	√	√
Language Genus	√	√	√	√	√
Personal Characteristics	√	√	√	√	√
Immigration History	√	√	√	√	√
Employment Status	√	√	√	√	√
Financial Situation	√	√	√	√	√
Family Situation	√	√	√	√	√
Political Views	√	√	√	√	√
Observations	4957	4957	4957	4957	4957

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. For all regressions, only people who report being immigrants themselves or report at least one parent as an immigrant are included (Questions V243-245 in WVS). Column I reports coefficients for ordered probit regressions. Column II to V report the marginal effects for each of the four different possible answers given for question V107 of the WVS. The same set of control variables is used as in column I. Hofstede's cultural dimensions (Hofstede, 2001) are used to control for for cultural characteristics and differences of societies.

B.2.2 Official Language

TABLE B.5: Language Effect on Trust: Not Speaking Official Language (Marginal Effects)

	I	II	III	Do not trust at all	Do not trust very much	Trust somewhat	Trust completely
Politeness Distinction	-0.644*	-0.613*	-0.565	0.129	0.049	-0.142	-0.037
	(0.354)	(0.336)	(0.351)	(0.084)	(0.044)	(0.090)	(0.023)
Trust			0.547***	-0.125***	-0.048	0.137***	0.036***
			(0.092)	(0.011)	(0.036)	(0.030)	(0.009)
Country Fixed Effects	√	√	√	√	√	√	√
Language Family	√	√	√	√	√	√	√
Language Genus	√	√	√	√	√	√	√
Personal Characteristics	×	√	√	√	√	√	√
Immigration History	×	×	√	√	√	√	√
Employment Status	×	×	√	√	√	√	√
Financial Situation	×	×	√	√	√	√	√
Family Situation	×	×	√	√	√	√	√
Political Views	×	×	√	√	√	√	√
Observations	1688	1688	1688	1688	1688	1688	1688

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. Individuals who reported to speak at home the official language of the country they live in have been dropped all regressions. The official language of a country was taken from the CIA World Factbook (Central Intelligence Agency, 2020). Column I to III report coefficients for ordered probit regressions with different sets of control variables. Column IV to VII report the marginal effects for each of the four different possible answers given for question V107 of the WVS. The same set of control variables is used as in column III.

TABLE B.6: Language Effect on Trust: Not Speaking Official Language and Cultural Dimensions (Marginal Effects)

	I	Do not trust at all	Do not trust very much	Trust somewhat	Trust completely
Politeness Distinction	-0.327*	0.076	0.029	-0.083*	-0.022*
	(0.186)	(0.048)	(0.022)	(0.046)	(0.013)
Trust	0.574***	-0.133***	-0.050	0.145***	0.038***
	(0.089)	(0.016)	(0.036)	(0.030)	(0.008)
Culture					
Masculinity	-0.003	0.001	0.000	-0.001	-0.000
	(0.003)	(0.001)	(0.000)	(0.001)	(0.000)
Power Distance	-0.016***	0.004***	0.001	-0.004***	-0.001***
	(0.004)	(0.001)	(0.001)	(0.001)	(0.000)
Country Fixed Effects	×	×	×	×	×
Language Family	√	√	√	√	√
Language Genus	√	√	√	√	√
Personal Characteristics	√	√	√	√	√
Immigration History	√	√	√	√	√
Employment Status	√	√	√	√	√
Financial Situation	√	√	√	√	√
Family Situation	√	√	√	√	√
Political Views	√	√	√	√	√
Observations	1688	1688	1688	1688	1688

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. Individuals who reported to speak at home the official language of the country they live in have been dropped all regressions. The official language of a country was taken from the CIA World Factbook (Central Intelligence Agency, 2020). Column I reports coefficients for ordered probit regressions. Column II to V report the marginal effects for each of the four different possible answers given for question V107 of the WVS. The same set of control variables is used as in column I. Hofstede's cultural dimensions (Hofstede, 2001) are used to control for for cultural characteristics and differences of societies.

B.2.3 GPS

TABLE B.7: Language Effect on Trust: Immigrants (GPS) (Marginal Effects)

	I	II	III	Do not trust at all	Do not trust very much	Trust somewhat	Trust completely
Politeness Distinction	-0.510***	-0.520***	-0.390***	0.080***	0.049**	-0.087***	-0.042***
	(0.162)	(0.165)	(0.125)	(0.029)	(0.019)	(0.029)	(0.015)
Trust			0.472***	-0.097***	-0.059***	0.106***	0.051***
			(0.031)	(0.013)	(0.016)	(0.011)	(0.007)
Country Fixed Effects	√	√	√	√	√	√	√
Cultural Preferences	×	×	×	×	×	×	×
Language Family	√	√	√	√	√	√	√
Language Genus	√	√	√	√	√	√	√
Personal Characteristics	×	√	√	√	√	√	√
Immigration History	×	×	√	√	√	√	√
Employment Status	×	×	√	√	√	√	√
Financial Situation	×	×	√	√	√	√	√
Family Situation	×	×	√	√	√	√	√
Political Views	×	×	√	√	√	√	√
Observations	4622	4622	4622	4622	4622	4622	4622

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. For all regressions, only people who report being immigrants themselves or report at least one parent as an immigrant are included (Questions V243-245 in WVS). Column I to III report coefficients for ordered probit regressions with different sets of control variables. Column IV to VII report the marginal effects for each of the four different possible answers given for question V107 of the WVS. The same set of control variables is used as in column III.

TABLE B.8: Language Effect on Trust: Immigrants and Cultural Preferences (GPS) (Marginal Effects)

	I	Do not trust at all	Do not trust very much	Trust somewhat	Trust completely
Politeness Distinction	-0.346***	0.073***	0.045***	-0.080***	-0.038***
	(0.098)	(0.024)	(0.016)	(0.025)	(0.011)
Trust	0.482***	-0.101***	-0.063***	0.111***	0.053***
	(0.030)	(0.015)	(0.015)	(0.011)	(0.008)
Country Fixed Effects	×	×	×	×	×
Cultural Preferences	√	√	√	√	√
Language Family	√	√	√	√	√
Language Genus	√	√	√	√	√
Personal Characteristics	√	√	√	√	√
Employment Status	√	√	√	√	√
Financial Situation	√	√	√	√	√
Family Situation	√	√	√	√	√
Political Views	√	√	√	√	√
Observations	4622	4622	4622	4622	4622

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. For all regressions, only people who report being immigrants themselves or report at least one parent as an immigrant are included (Questions V243-245 in WVS). Column I reports coefficients for ordered probit regressions. Column II to V report the marginal effects for each of the four different possible answers given for question V107 of the WVS. The same set of control variables is used as in column III. GPS preference measures (Falk et al., 2018) are used to control for for cultural characteristics and differences of societies.

TABLE B.9: Language Effect on Trust: Not Speaking Official Language (GPS) (Marginal Effects)

	I	II	III	Do not trust at all	Do not trust very much	Trust somewhat	Trust completely
Politeness Distinction	-0.597*	-0.554*	-0.507	0.135	0.034	-0.126	-0.044
	(0.324)	(0.315)	(0.324)	(0.091)	(0.036)	(0.079)	(0.029)
Trust			0.418***	-0.111***	-0.029	0.104***	0.036***
			(0.056)	(0.019)	(0.024)	(0.018)	(0.005)
Country Fixed Effects	√	√	√	√	√	√	√
Cultural Preferences	×	×	×	×	×	×	×
Language Family	√	√	√	√	√	√	√
Language Genus	√	√	√	√	√	√	√
Personal Characteristics	×	√	√	√	√	√	√
Employment Status	×	×	√	√	√	√	√
Financial Situation	×	×	√	√	√	√	√
Family Situation	×	×	√	√	√	√	√
Political Views	×	×	√	√	√	√	√
Observations	3917	3917	3917	3917	3917	3917	3917

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. Individuals who reported to speak at home the official language of the country they live in have been dropped all regressions. The official language of a country was taken from the CIA factbook Central Intelligence Agency, 2020. Column I to III report coefficients for ordered probit regressions with different sets of control variables. Column IV to VII report the marginal effects for each of the four different possible answers given for question V107 of the WVS. The same set of control variables is used as in column III. The coefficients and marginal effects of the control variables can be found in the appendix. GPS preference measures (Falk et al., 2018) are used to control for for cultural characteristics and differences of societies.

TABLE B.10: Language Effect on Trust: Not Speaking Official Language and Cultural Preferences (GPS) (Marginal Effects)

	I	Do not trust at all	Do not trust very much	Trust somewhat	Trust completely
Politeness Distinction	-0.022	0.006	0.001	-0.005	-0.002
	(0.205)	(0.055)	(0.017)	(0.051)	(0.018)
Trust	0.435***	-0.116***	-0.030	0.108***	0.038***
	(0.054)	(0.019)	(0.025)	(0.018)	(0.005)
Country Fixed Effects	×	×	×	×	×
Cultural Preferences	√	√	√	√	√
Language Family	√	√	√	√	√
Language Genus	√	√	√	√	√
Personal Characteristics	√	√	√	√	√
Employment Status	√	√	√	√	√
Financial Situation	√	√	√	√	√
Family Situation	√	√	√	√	√
Political Views	√	√	√	√	√
Observations	3917	3917	3917	3917	3917

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. Individuals who reported to speak at home the official language of the country they live in have been dropped all regressions. The official language of a country was taken from the CIA factbook Central Intelligence Agency, 2020. Column I reports coefficients for ordered probit regressions. Column II to V report the marginal effects for each of the four different possible answers given for question V107 of the WVS. The same set of control variables is used as in column III. GPS preference measures (Falk et al., 2018) are used to control for for cultural characteristics and differences of societies.

B.3 Summary Statistics

TABLE B.11: Politeness Distinction by Country (GPS)

Country	Mean	Std. Dev.	Freq.	Country	Mean	Std. Dev.	Freq.
Algeria	0.017	0.128	1197	Morocco	0.006	0.076	1199
Argentina	1	0	1021	Netherlands	0.995	0.067	1768
Australia	0.036	0.187	1434	Nigeria	0.224	0.417	1646
Brazil	1	0	1486	Pakistan	0.961	0.194	847
Chile	1	0	1000	Peru	0.993	0.081	1210
China	1	0	2300	Philippines	1	0	512
Colombia	0.995	0.068	1505	Poland	1	0	963
Estonia	1	0	496	Romania	1	0	1498
Georgia	1	0	1200	Russia	1	0	2343
Germany	0.990	0.099	2027	Rwanda	0.622	0.492	37
Ghana	0.004	0.063	251	South Africa	0	0	1224
Haiti	0.955	0.213	22	Zimbabwe	0	0	65
India	1	0	1871	Sweden	0.989	0.106	1142
Iraq	0	0	955	Thailand	1	0	1152
Japan	1	0	2443	Turkey	1	0	1498
Kazakhstan	1	0	736	Ukraine	1	0	737
Jordan	0.001	0.029	1200	Egypt	0	0	1523
Mexico	0.974	0.159	1936	United States	0.072	0.259	2173

Overall

	Mean	Std. Dev.	Freq.
Total	0.722	0.448	44617

Note: Distribution of the politeness distinction variable of the language individuals speak at home grouped by their country of living. Frequency is the total number of observations from one country. Data for the dataset with GPS preference measures instead of Hofstede cultural dimensions.

TABLE B.12: List of Languages

Full Dataset			Only Immigrants			Official language		
Country	Politeness Distinction	Freq.	Country	Politeness Distinction	Freq.	Country	Politeness Distinction	Freq.
Albanian	No	5	Albanian	No	5	Albanian	No	5
Arabic	No	1150	Arabic	No	27	Arabic	No	23
Armenian	Yes	1	Armenian	Yes	1	Armenian	Yes	1
Aymara	No	7	Berber	No	1	Aymara	No	7
Berber	No	66	Dutch	Yes	341	Berber	No	66
Brahui	No	33	English	No	1506	English	No	150
Dutch	Yes	1747	Finnish	Yes	2	Finnish	Yes	3
English	No	6128	French	Yes	14	French	Yes	31
Finnish	Yes	3	German	Yes	279	German	Yes	18
French	Yes	31	Greek	Yes	5	Greek	Yes	5
German	Yes	1948	Hindi	Yes	111	Hungarian	Yes	131
Greek	Yes	5	Hungarian	Yes	7	Indonesian	Yes	1
Hindi	Yes	973	Indonesian	Yes	1	Italian	Yes	8
Hungarian	Yes	131	Italian	Yes	7	Japanese	Yes	4
Indonesian	Yes	1	Japanese	Yes	3	Kashmiri	Yes	2
Italian	Yes	8	Kannada	Yes	1	Korean	Yes	2
Japanese	Yes	2447	Kashmiri	Yes	2	Mandarin	Yes	356
Kannada	Yes	144	Korean	Yes	2	Maori	No	2
Kashmiri	Yes	2	Malayalam	Yes	1	Nepali	Yes	6
Korean	Yes	2	Mandarin	Yes	470	Panjabi	Yes	2
Malayalam	Yes	192	Marathi	Yes	1	Persian	Yes	1
Mandarin	Yes	4659	Nepali	Yes	2	Polish	Yes	10
Maori	No	2	Panjabi	Yes	119	Portuguese	Yes	1
Marathi	Yes	250	Pashto	Yes	86	Quechua	Yes	44
Nepali	Yes	6	Persian	Yes	1	Russian	Yes	526
Panjabi	Yes	744	Polish	Yes	51	Sinhala	Yes	2
Pashto	Yes	233	Portuguese	Yes	83	Spanish	Yes	163
Persian	Yes	1	Quechua	Yes	3	Tagalog	Yes	6
Polish	Yes	970	Romanian	Yes	19	Tamil	Yes	63
Portuguese	Yes	1487	Russian	Yes	483	Turkish	Yes	39
Quechua	Yes	44	Sinhala	Yes	2	Vietnamese	Yes	10
Romanian	Yes	1375	Spanish	Yes	835			
Russian	Yes	2869	Swedish	Yes	202			
Sinhala	Yes	2	Tagalog	Yes	9			
Spanish	Yes	8871	Tamil	Yes	38			
Swedish	Yes	1114	Thai	Yes	25			
Tagalog	Yes	518	Turkish	Yes	199			
Tamil	Yes	160	Urdu	Yes	4			
Thai	Yes	1150	Vietnamese	Yes	9			
Turkish	Yes	1537						
Urdu	Yes	126						
Vietnamese	Yes	10						

Note: Number of participants who reported speaking a language at home. The politeness distinction column indicates whether the language has a politeness distinction or not.

TABLE B.13: List of Languages (GPS)

Full Dataset			Only Immigrants			Official language		
Country	Politeness Distinction	Freq.	Country	Politeness Distinction	Freq.	Country	Politeness Distinction	Freq.
Albanian	No	5	Albanian	No	5	Albanian	No	5
Arabic	No	5823	Arabic	No	616	Arabic	No	23
Armenian	Yes	13	Armenian	Yes	3	Armenian	Yes	13
Aymara	No	7	Berber	No	16	Aymara	No	7
Berber	No	246	Dutch	Yes	341	Berber	No	66
Brahui	No	33	English	No	808	Brahui	No	33
Dutch	Yes	1747	Ewe	No	11	English	No	81
English	No	4326	Finnish	Yes	2	Ewe	No	167
Ewe	No	167	French	Yes	38	Finnish	Yes	3
Finnish	Yes	3	Georgian	Yes	47	French	Yes	46
French	Yes	73	German	Yes	278	German	Yes	9
Georgian	Yes	1180	Greek	Yes	5	Greek	Yes	5
German	Yes	1939	Hausa	No	25	Hausa	No	28
Greek	Yes	5	Hindi	Yes	111	Hungarian	Yes	126
Hausa	No	615	Hungarian	Yes	4	Indonesian	Yes	1
Hindi	Yes	973	Igbo	No	3	Italian	Yes	7
Hungarian	Yes	126	Indonesian	Yes	1	Japanese	Yes	3
Igbo	No	338	Italian	Yes	7	Kannada	Yes	144
Indonesian	Yes	1	Japanese	Yes	2	Kashmiri	Yes	2
Italian	Yes	7	Kannada	Yes	1	Korean	Yes	2
Japanese	Yes	2446	Kashmiri	Yes	2	Malayalam	Yes	192
Kannada	Yes	144	Korean	Yes	2	Mandarin	Yes	15
Kashmiri	Yes	2	Malayalam	Yes	1	Marathi	Yes	250
Korean	Yes	2	Mandarin	Yes	15	Nepali	Yes	6
Malayalam	Yes	192	Marathi	Yes	1	Panjabi	Yes	744
Mandarin	Yes	2315	Nepali	Yes	2	Pashto	Yes	233
Marathi	Yes	250	Panjabi	Yes	119	Persian	Yes	1
Nepali	Yes	6	Pashto	Yes	86	Polish	Yes	10
Panjabi	Yes	744	Persian	Yes	1	Portuguese	Yes	1
Pashto	Yes	233	Polish	Yes	51	Quechua	Yes	44
Persian	Yes	1	Portuguese	Yes	83	Romanian	Yes	12
Polish	Yes	970	Quechua	Yes	3	Russian	Yes	1256
Portuguese	Yes	1487	Romanian	Yes	21	Sinhala	Yes	2
Quechua	Yes	44	Russian	Yes	846	Spanish	Yes	179
Romanian	Yes	1387	Sinhala	Yes	2	Tagalog	Yes	6
Russian	Yes	4333	Spanish	Yes	582	Tamil	Yes	17
Sinhala	Yes	2	Swahili	No	3	Turkish	Yes	41
Spanish	Yes	6735	Swedish	Yes	202	Urdu	Yes	126
Swahili	No	12	Tagalog	Yes	9	Vietnamese	Yes	10
Swedish	Yes	1114	Thai	Yes	25	Zulu	No	1
Tagalog	Yes	518	Turkish	Yes	199			
Tamil	Yes	17	Urdu	Yes	4			
Thai	Yes	1150	Vietnamese	Yes	9			
Turkish	Yes	1539	Yoruba	Yes	12			
Urdu	Yes	126	Zulu	No	18			
Vietnamese	Yes	10						
Yoruba	Yes	369						
Zulu	No	842						

Note: Number of participants who reported speaking a language at home. The politeness distinction column indicates whether the language has a politeness distinction or not. Data for the dataset with GPS preference measures instead of Hofstede cultural dimensions.

Appendix C

Language and Gender

C.1 Data

TABLE C.1: Summary statistics: Descriptives (Female Subsample)

	Age	Education	Income Class	Happiness
Mean	48.5	12.5	4.66	6.973
SD	18.71	4.04	2.707	2.136
Min.	15	0	1	0
Max.	101	50	10	10
Obs.	14,301	14,202	11,169	14,203
Missing	30	129	3,162	128

	Religious Denomination	Political Views Left/Right	Place of Residence Rural/Urban	Household Size
Mean	0.717	5.104	2.715	2.658
SD	0.451	2.082	1.268	1.437
Min.	0	0	1	1
Max.	1	10	5	14
Obs.	14,240	11,619	14,303	14,323
Missing	91	2,712	28	8

Note: Summary statistics of the socio-economic and demographic characteristics of female individuals.

TABLE C.2: List of Feminine Occupations

ISCO-88	Occupation
2230	Nursing and midwifery professionals
2331	Primary education teaching professionals
2332	Pre-primary educ teaching professionals
3221	Medical assistants
3225	Dental assistants
3227	Veterinary assistants
3228	Pharmaceutical assistants
3231	Nursing associate professionals
3232	Midwifery associate professionals
3310	Primary education teaching associate professionals
3320	Pre-primary education teaching associate professionals
3431	Administrative secretaries, related associate professionals
3471	Decorators and commercial designers
4111	Stenographers and typists
4112	Word-processor and related operators
4115	Secretaries
4211	Cashiers and ticket clerks
4222	Receptionists and information clerks
4223	Telephone switchboard operators
5121	Housekeepers and related workers
5123	Waiters, waitresses and bartenders
5131	Child-care workers
5132	Institution-based personal care workers
5133	Home-based personal care workers
5139	Personal care, related workers not else class
5141	Hairdresser, barber, beautician, related workers
5210	Fashion and other models
7432	Weavers, knitters and related workers
7433	Tailors, dressmakers and hatters
9131	Domestic helpers and cleaners
9132	Helper, cleaner in office, hotel, other establishments
9133	Hand-launderers and pressers

Note: Occupations considered to be traditionally more feminine. I follow data from Eurostat and the International Labour Organization, which reports the proportion of women in different occupations, to determine which occupations are female-gendered (Eurostat, 2018; ILO, 2020).

TABLE C.3: List of Languages and its Features

Language	Gender Intensity Index	Sex-based	Number Genders	Gender Assignment	Gender in Pronouns
Amharic	4	1	1	1	1
Armenian	0	0	0	0	0
Basque	0	0	0	0	0
Berber	4	1	1	1	1
English	1	1	0	0	0
Filipino	2	1	1	0	0
Finnish	0	0	0	0	0
French	3	1	1	1	0
Georgian	0	0	0	0	0
German	2	1	0	1	0
Greek	2	1	0	1	0
Guarani	0	0	0	0	0
Hebrew	4	1	1	1	1
Hindi	3	1	1	1	0
Hungarian	0	0	0	0	0
Indonesian	0	0	0	0	0
Khmer	0	0	0	0	0
Latvian	3	1	1	1	0
Lezgian	0	0	0	0	0
Malagasy	0	0	0	0	0
Oromo	3	1	1	1	0
Persian	0	0	0	0	0
Russian	2	1	0	1	0
Spanish	4	1	1	1	1
Tagalog	2	1	1	0	0
Thai	0	0	0	0	0
Turkish	0	0	0	0	0
Vietnamese	0	0	0	0	0
Yoruba	0	0	0	0	0

Note: Values of the GII and the individual language features for each language in my sample.

TABLE C.4: Summary Statistics: GII per Country

Country	Mean	SD	Variance	Min.	Max.	Obs.
Belgium	2.920	0.477	0.227	0	4	651
Bulgaria	0.000	0.000	0.000	0	0	248
Croatia	1.000	0.707	0.500	0	2	9
Cyprus	1.993	0.091	0.008	0	2	1,072
Czechia	2.333	0.577	0.333	2	3	3
Denmark	0.833	0.618	0.382	0	2	18
Estonia	1.990	0.125	0.016	0	2	381
Finnland	0.024	0.230	0.053	0	4	1,759
France	2.970	0.296	0.088	0	4	1,636
Germany	1.973	0.258	0.066	0	4	2,805
Greece	1.999	0.039	0.001	0	2	2,675
Hungary	0.003	0.076	0.006	0	2	1,560
Ireland	1.021	0.219	0.048	0	4	2,413
Israel	3.707	0.751	0.565	0	4	1,898
Lithuania	1.988	0.112	0.013	1	2	80
Netherlands	1.281	1.350	1.822	0	4	32
Norway	1.410	1.044	1.090	0	4	39
Poland	2.000	.	.	2	2	1
Portugal	2.100	0.994	0.989	1	3	10
Russia	1.997	0.080	0.006	0	2	2,500
Slovakia	0.028	0.236	0.056	0	2	143
Slovenia	0.600	0.966	0.933	0	2	10
Spain	3.972	0.308	0.095	0	4	1,746
Sweden	0.865	1.273	1.620	0	4	37
Switzerland	2.190	0.517	0.267	0	4	1,322
Ukraine	1.993	0.117	0.014	0	2	872
United Kingdom	1.016	0.212	0.045	0	4	2,275

Note: Summary statistics for the GII for each country in my sample.

C.2 Non-Linear Models

Results for non-linear regression models are presented in table C.5. I use a ordered logit model in columns I-III and VII, a logit model in columns IV and V and a poisson model in column VI.

TABLE C.5: Non-Linear Models

	Cut Down Work	Scarce Jobs	Proportion Women	Labour Market Participation	Female Occupation	Total Work Hours	Housework Hours
	I	II	III	IV	V	VI	VII
Sex-Based	-0.879***	-2.050***	-2.481***	-1.620**	0.188	-0.336***	0.948**
	(0.287)	(0.228)	(0.433)	(0.799)	(0.449)	(0.104)	(0.378)
Nb. of Genders	0.248	0.328**	0.326	0.059	0.397	0.106	-0.130
	(0.192)	(0.128)	(0.204)	(0.410)	(0.407)	(0.071)	(0.168)
Assignment	-0.167	-0.279*	0.079	0.315	0.046	0.024	0.620***
	(0.170)	(0.169)	(0.231)	(0.433)	(0.163)	(0.037)	(0.204)
Pronouns	-0.252**	0.371***	-0.462	-0.963***	0.383	0.105***	-0.396**
	(0.120)	(0.126)	(0.296)	(0.263)	(0.239)	(0.020)	(0.200)
Female	0.164***	0.572***	2.277***				
	(0.055)	(0.070)	(0.135)				
Country Fixed Effects	√	√	√	√	√	√	√
Language Fixed Effects	√	√	√	√	√	√	√
Age	√	√	√	√	√	√	√
Happiness	√	√	√	√	√	√	√
Religious Denomination	√	√	√	√	√	√	√
Political Views	√	√	√	√	√	√	√
Urban/Rural	√	√	√	√	√	√	√
Education	√	√	√	√	√	√	√
Household Size	√	√	√	√	√	√	√
Household Income	√	√	√	√	√	√	√
Cultural Values	√	√	√	√	√	√	√
Cultural Values Parents	√	√	√	√	√	√	√
Observations	20622	20622	8154	11253	9664	4782	5679

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. SE in parenthesis. Dependent Variables: Question G4: A woman should be prepared to cut down on her paid work for the sake of her family. Question G5: When jobs are scarce, men should have more right to a job than women. The given answers are ordered from *Agree strongly* to *Disagree strongly*. Therefore, a negative coefficient is associated with a higher agreement with the statement. Question G38: Proportion of women at workplace. *Labour Market Participation*: Respondent was *in paid work* or *unemployed and actively looking for a job* in the last 7 days. *Female Occupation*: The main occupation of the respondent is seen as traditionally more feminine. *Total Work Hours*: Counts the hours the respondent normally works per week, including paid and unpaid overtime. Only respondents with paid work in the last 7 days are used. *Housework Hours* is the total hours of housework reported for the last week. The reported values are grouped into brackets of 10 hours each. Last bracket contains all values above 100. Independent Variables: *Sex-Based* is equal to 1 if a language has a sex-based gender system. *Nb. of Genders* is equal to 1 if a language only has 2 genders. *Assignment* is equal to 1 if a language has a semantic and formal assignment system. *Pronouns* is equal to 1 for languages with a gender distinction in third-person and in first and/or second-person. Columns I-III and VII are ordered logit models. Columns IV and V are logit models. Column VI is a poisson model. Missing observations are chained imputed.

C.3 Alternative GII

Gay et al. (2018) use a different calculation method for the index to measure the intensity of gender in languages. Instead of using the sum of all 4 language characteristics the author use the following formula to derive their version of the *Gender Intensity Index*. This method of calculation reflects the point I made earlier that the assessment and effect of the other 3 features is strongly related to whether the gender system of a language is fundamentally based on biological sex.

$$GII2 = Sex\ based \times (Number\ of\ Genders + Gender\ Assignment + Gender\ in\ Pronouns) \tag{C.1}$$

TABLE C.6: Effect of Language and Gender Norms and Labour Market Outcomes (Alternative GII)

	Cut Down Work	Scarce Jobs	Proportion Women	Labour Market Participation	Female Occupation	Total Work Hours	Housework Hours
	I	II	III	IV	V	VI	VII
GII2	-0.132**	-0.028	-0.082	-0.014	0.031	1.869	0.149*
	(0.058)	(0.099)	(0.148)	(0.069)	(0.031)	(1.128)	(0.082)
Female	0.103***	0.322***	1.938***				
	(0.032)	(0.059)	(0.129)				
Country Fixed Effects	√	√	√	√	√	√	√
Language Fixed Effects	√	√	√	√	√	√	√
Age	√	√	√	√	√	√	√
Happiness	√	√	√	√	√	√	√
Religious Denomination	√	√	√	√	√	√	√
Political Views	√	√	√	√	√	√	√
Urban/Rural	√	√	√	√	√	√	√
Education	√	√	√	√	√	√	√
Household Size	√	√	√	√	√	√	√
Household Income	√	√	√	√	√	√	√
Cultural Values	√	√	√	√	√	√	√
Cultural Values Parents	√	√	√	√	√	√	√
Observations	20622	20622	8154	11266	11266	4782	5679

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. SE in parenthesis. Levels of the GII2 are standardized. Dependent Variables: Question G4: A woman should be prepared to cut down on her paid work for the sake of her family. Question G5: When jobs are scarce, men should have more right to a job than women. The given answers are ordered from *Agree strongly* to *Disagree strongly*. Therefore, a negative coefficient is associated with a higher agreement with the statement. Question G38: Proportion of women at workplace. *Labour Market Participation*: Respondent was *in paid work* or *unemployed and actively looking for a job* in the last 7 days. *Female Occupation*: The main occupation of the respondent is seen as traditionally more feminine. *Total Work Hours*: Counts the hours the respondent normally works per week, including paid and unpaid overtime. Only respondents with paid work in the last 7 days are used. *Housework Hours* is the total hours of housework reported for the last week. Missing observations are chained imputed.

C.4 Sex-based Only Dataset

WALS has information about the connection of the gender system of a language to biological sex for some languages for which the information about at least on of the other 3 features connected to gender is missing. Therefore, the number of observations in my sample increases when I only use the sex-based language variable. Regression results are presented in tables C.7 and C.8.

TABLE C.7: Effect of Language and Gender Norms and Labour Market Outcomes: Only Sex-based

	Cut Down Work	Scarce Jobs	Proportion Women	Labour Market Participation	Female Occupation	Total Work Hours	Housework Hours
	I	II	III	IV	V	VI	VII
Sex-Based	-0.730***	-1.223***	-1.813***	-0.101	0.433***	-11.830***	1.479***
	(0.108)	(0.122)	(0.317)	(0.072)	(0.096)	(3.254)	(0.256)
Female	0.108***	0.326***	1.956***				
	(0.031)	(0.056)	(0.123)				
Country Fixed Effects	√	√	√	√	√	√	√
Language Fixed Effects	√	√	√	√	√	√	√
Age	√	√	√	√	√	√	√
Happiness	√	√	√	√	√	√	√
Religious Denomination	√	√	√	√	√	√	√
Political Views	√	√	√	√	√	√	√
Urban/Rural	√	√	√	√	√	√	√
Education	√	√	√	√	√	√	√
Household Size	√	√	√	√	√	√	√
Household Income	√	√	√	√	√	√	√
Cultural Values	√	√	√	√	√	√	√
Cultural Values Parents	√	√	√	√	√	√	√
Observations	21718	21718	8484	11939	11939	5254	5964

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. SE in parenthesis. Immigrant is anyone who was either not born in the country in which they currently live or anyone of whom at least one parent was not born in the country in which the respondent currently lives. Dependent Variables: Question G4: A woman should be prepared to cut down on her paid work for the sake of her family. Question G5: When jobs are scarce, men should have more right to a job than women. The given answers are ordered from *Agree strongly* to *Disagree strongly*. Therefore, a negative coefficient is associated with a higher agreement with the statement. Question G38: Proportion of women at workplace. *Labour Market Participation*: Respondent was *in paid work* or *unemployed and actively looking for a job* in the last 7 days. *Female Occupation*: The main occupation of the respondent is seen as traditionally more feminine. *Total Work Hours*: Counts the hours the respondent normally works per week, including paid and unpaid overtime. Only respondents with paid work in the last 7 days are used. *Housework Hours* is the total hours of housework reported for the last week. The reported values are grouped into brackets of 10 hours each. Last bracket contains all values above 100. Independent Variables: *Sex-Based* is equal to 1 if a language has a sex-based gender system. Missing observations are chained imputed.

TABLE C.8: Effect of Language and Gender Norms and Labour Market Outcomes: Only Sex-based ESS Round 5-9

	Labour Market Participation	Female Occupation	Total Work Hours
	I	II	III
Sex-Based	-0.072**	0.057***	-9.129***
	(0.029)	(0.016)	(0.744)
Country Fixed Effects	√	√	√
Language Fixed Effects	√	√	√
Age	√	√	√
Happiness	√	√	√
Religious Denomination	√	√	√
Political Views	√	√	√
Urban/Rural	√	√	√
Education	√	√	√
Household Size	√	√	√
Household Income	√	√	√
Cultural Values	√	√	√
Cultural Values Parents	√	√	√
Observations	50319	50319	24230

Note: *** Significant at the 1% level. ** Significant at the 5% level. * Significant at the 10% level. SE in parenthesis. Immigrant is anyone who was either not born in the country in which they currently live or anyone of whom at least one parent was not born in the country in which the respondent currently lives. Dependent Variables: Question G4: A woman should be prepared to cut down on her paid work for the sake of her family. Question G5: When jobs are scarce, men should have more right to a job than women. The given answers are ordered from *Agree strongly* to *Disagree strongly*. Therefore, a negative coefficient is associated with a higher agreement with the statement. Question G38: Proportion of women at workplace. *Labour Market Participation*: Respondent was *in paid work* or *unemployed and actively looking for a job* in the last 7 days. *Female Occupation*: The main occupation of the respondent is seen as traditionally more feminine. *Total Work Hours*: Counts the hours the respondent normally works per week, including paid and unpaid overtime. Only respondents with paid work in the last 7 days are used. *Housework Hours* is the total hours of housework reported for the last week. The reported values are grouped into brackets of 10 hours each. Last bracket contains all values above 100. Independent Variables: *Sex-Based* is equal to 1 if a language has a sex-based gender system. Missing observations are chained imputed.

Bibliography

Alesina, Alberto and Paola Giuliano (June 2010). "The power of the family". In: *Journal of Economic Growth* 15.2, pp. 93–125. DOI: 10.1007/s10887-010-9052-z.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn (May 2013). "On the Origins of Gender Roles: Women and the Plough *". In: *The Quarterly Journal of Economics* 128.2, pp. 469–530. DOI: 10.1093/qje/qjt005.

Athanasopoulos, Panos and Emanuel Bylund (Mar. 2013). "Does Grammatical Aspect Affect Motion Event Cognition? A Cross-Linguistic Comparison of English and Swedish Speakers". In: *Cognitive Science* 37.2, pp. 286–309. DOI: 10.1111/cogs.12006.

Athanasopoulos, Panos and Aina Casaponsa (Aug. 2020). "The Whorfian brain: Neuroscientific approaches to linguistic relativity". In: *Cognitive Neuropsychology* 37.5-6, pp. 393–412. DOI: 10.1080/02643294.2020.1769050.

Baltagi, Badi H. (2013). *Econometric Analysis of Panel Data*. 5th. Chichester: John Wiley & Sons Ltd.

Bernhofer, Juliana, Francesco Costantini, and Matija Kovacic (May 2021). "Risk Attitudes, Investment Behavior and Linguistic Variation". In: *Journal of Human Resources*, 0119–9999R2. DOI: 10.3368/jhr.59.2.0119-9999R2.

Blau, Francine D. and Lawrence M. Kahn (2015). "Substitution between individual and source country characteristics: Social capital, culture, and US labor market outcomes among immigrantwomen". In: *Journal of Human Capital* 9.4, pp. 439–482. DOI: 10.1086/683542.

Blau, Francine D., Lawrence M. Kahn, and Kerry L Papps (Feb. 2011). "Gender, Source Country Characteristics, and Labor Market Assimilation among Immigrants". In: *Review of Economics and Statistics* 93.1. Ed. by Anne C. Gielen and Klaus F. Zimmermann, pp. 43–58. DOI: 10.1162/REST_a_00064.

Bloom, Nick, Rachel Griffith, and John Van Reenen (2002). "Do R & D tax credits work? Evidence from a panel of countries 1979-1997". In: *Journal of Public Economics* 85.1, pp. 1–31. DOI: 10.1016/S0047-2727(01)00086-X.

Boroditsky, Lera (Aug. 2001). "Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time". In: *Cognitive Psychology* 43.1, pp. 1–22. DOI: 10.1006/cogp.2001.0748.

– (Jan. 2006). "Linguistic Relativity". In: *Encyclopedia of Cognitive Science*. Chichester: John Wiley & Sons, Ltd, pp. 158–174. DOI: 10.1002/0470018860.s00567.

– (2018). "How Language Shapes Thought". In: *Scientific American* 304.2, pp. 62–65.

Boroditsky, Lera and Alice Gaby (Nov. 2010). "Remembrances of Times East". In: *Psychological Science* 21.11, pp. 1635–1639. DOI: 10.1177/0956797610386621.

Boroditsky, Lera and Webb Philipps (2003). "Can Quirks of Grammar Affect the Way You Think? Grammatical Gender and Object Concepts". In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 25.25.

Boroditsky, Lera, Lauren A. Schmidt, and Webb Philipps (2003). "Sex, syntax and semantics". In: *Language in mind: Advances in the study of language and thought*. Ed. by D. Gentner and S. Goldin-Meadow. Cambridge, MA: MIT Press, pp. 61–79.

Boutonnet, Bastien, Panos Athanasopoulos, and Guillaume Thierry (Oct. 2012). "Unconscious effects of grammatical gender during object categorisation". In: *Brain Research* 1479, pp. 72–79. DOI: 10.1016/j.brainres.2012.08.044.

Bowerman, Melissa (1996). "The origins of children's spatial semantic categories: Cognitive versus linguistic determinants". In: *Rethinking Linguistic Relativity*. Ed. by John J. Gumperz and Stephen C. Levinson. Cambridge, MA: Cambridge University Press, pp. 145–176.

Brown, R. and A. Gilman (1960). "The Pronouns of Power and Solidarity". In: *Style in Language*. Ed. by T. A. Sebeok. MIT Press. Chap. 12, pp. 253–276.

Bylund, Emanuel, Panos Athanasopoulos, and Marcelyn Oostendorp (Jan. 2013). "Motion event cognition and grammatical aspect: Evidence from Afrikaans". In: *Linguistics* 51.5. DOI: 10.1515/ling-2013-0033.

Cantwell, John and Lucia Piscitello (Feb. 2005). "Recent Location of Foreign-owned Research and Development Activities by Large Multinational Corporations in the European Regions: The Role of Spillovers and Externalities". In: *Regional Studies* 39.1, pp. 1–16. DOI: 10.1080/0034340052000320824.

Casasanto, Daniel (2008). "Who's afraid of the big bad Whorf? crosslinguistic differences in temporal language and thought". In: *Language Learning* 58.1, pp. 63–79. DOI: 10.1111/j.1467-9922.2008.00462.x.

Casasanto, Daniel and Lera Boroditsky (Feb. 2008). "Time in the mind: Using space to think about time". In: *Cognition* 106.2, pp. 579–593. DOI: 10.1016/j.cognition.2007.03.004.

Central Intelligence Agency (2020). *The World Factbook 2020*. URL: https://www.cia.gov/the-world-factbook/ (visited on 01/31/2022).

Chen, M. Keith (2012). *Whorfian Economics*. URL: https://languagelog.ldc.upenn.edu/nll/?p=3792 (visited on 02/11/2019).

– (2013). "The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets". In: *American Economic Review* 103.2, pp. 690–731. DOI: 10.1257/aer.103.2.690.

Chen, Shimin et al. (Oct. 2017). "Languages and corporate savings behavior". In: *Journal of Corporate Finance* 46, pp. 320–341. DOI: 10.1016/j.jcorpfin.2017.07.009.

Chi, Jianxin Daniel et al. (June 2020). "Is language an economic institution? Evidence from R&D investment". In: *Journal of Corporate Finance* 62, p. 101578. DOI: 10.1016/j.jcorpfin.2020.101578.

Corbett, Greville G. (2013a). "Number of Genders". In: *The World Altlas of Language Structures Online*. Ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL: https://wals.info/chapter/30.

– (2013b). "Sex-based and Non-sex-based Gender Systems". In: *The World Altlas of Language Structures Online*. Ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL: https://wals.info/chapter/31.

– (2013c). "Systems of Gender Assignment". In: *The World Altlas of Language Structures Online*. Ed. by Matthew S. Dryer and Martin Hasplemath. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL: https://wals.info/chapter/32.

Cubelli, Roberto et al. (2011). "The effect of grammatical gender on object categorization". In: *Journal of Experimental Psychology: Learning Memory and Cognition* 37.2, pp. 449–460. DOI: 10.1037/a0021965.

Dahl, Östen (2000). "The Grammar of Future Time Reference in European Languages". In: *Tense and Aspect in the Languages of Europe*. Ed. by Östen Dahl. Berlin: Mouton de Gruyter, pp. 309–328.

– (2009). *Stuck in the futureless Zone*. URL: https://dlc.hypotheses.org/360 (visited on 07/01/2019).

Dahl, Östen and Viveka Velupillai (2013). "Tense and Aspect". In: *The World Altlas of Language Structures Online*. Ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL: https://wals.info/chapter/s7.

Davis, Lewis S. and Farangis Abdurazokzoda (2016). "Language, culture and institutions: Evidence from a new linguistic dataset". In: *Journal of Comparative Economics* 44.3, pp. 541–561. DOI: 10.1016/j.jce.2015.10.015.

DeFranza, David, Himanshu Mishra, and Arul Mishra (July 2020). "How language shapes prejudice against women: An examination across 45 world languages." In: *Journal of Personality and Social Psychology* 119.1, pp. 7–22. DOI: 10.1037/pspa0000188.

Dryer, Matthew and Martin Haspelmath, eds. (2013). *The World Atlas of Language Structure Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL: http://wals.info.

Ernst, Christof, Katharina Richter, and Nadine Riedel (2014). "Corporate taxation and the quality of research and development". In: *International Tax and Public Finance* 21.4, pp. 694–719. DOI: 10.1007/s10797-014-9315-2.

Ernst, Christof and Christoph Spengel (2011). "Taxation, R&D Tax Incentives and Patent Application in Europe". In: *SSRN Electronic Journal* 11. DOI: 10.2139/ssrn.1805762.

ESS Round 5 (2010). *European Social Survey Round 5 Data, Data file edition 3.4*. Norway - Data Archive and distributor of ESS data for ESS ERIC: NSD - Norwegian Centre for Research Data. DOI: 10.21338/NSD-ESS5-2010.

ESS Round 6 (2012). *European Social Survey Round 6 Data, Data file edition 2.4*. Norway - Data Archive and distributor of ESS data for ESS ERIC: NSD - Norwegian Centre for Research Data. DOI: 10.21338/NSD-ESS6-2012.

ESS Round 7 (2014). *European Social Survey Round 7 Data, Data file edition 2.2*. Norway - Data Archive and distributor of ESS data for ESS ERIC: NSD - Norwegian Centre for Research Data. DOI: 10.21338/NSD-ESS7-2014.

ESS Round 8 (2016). *European Social Survey Round 8 Data, Data file edition 2.2*. Norway - Data Archive and distributor of ESS data for ESS ERIC: NSD - Norwegian Centre for Research Data. DOI: 10.21338/NSD-ESS8-2016.

ESS Round 9 (2018). *European Social Survey Round 9 Data, Data file edition 3.1*. Norway - Data Archive and distributor of ESS data for ESS ERIC: NSD - Norwegian Centre for Research Data. DOI: 10.21338/NSD-ESS9-2018.

Eurostat (2018). *Jobs still split along gender lines*. URL: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20180307-1 (visited on 01/31/2022).

Facchini, Giovanni and Anna Maria Mayda (Feb. 2012). "Individual Attitudes Towards Skilled Migration: An Empirical Analysis Across Countries". In: *The World Economy* 35.2, pp. 183–196. DOI: 10.1111/j.1467-9701.2011.01427.x.

Facchini, Giovanni, Anna Maria Mayda, and Prachi Mishra (Sept. 2011). "Do interest groups affect US immigration policy?" In: *Journal of International Economics* 85.1, pp. 114–128. DOI: 10.1016/j.jinteco.2011.05.006.

Falk, Armin et al. (2016). "The preference survey module: A validated instrument for measuring risk, time, and social preferences". In: *IZA Discussion Paper* 9674.

Falk, Armin et al. (Nov. 2018). "Global Evidence on Economic Preferences". In: *The Quarterly Journal of Economics* 133.4, pp. 1645–1692. DOI: 10.1093/qje/qjy013.

Farré, Lídia and Francis Vella (Apr. 2013). "The Intergenerational Transmission of Gender Role Attitudes and its Implications for Female Labour Force Participation". In: *Economica* 80.318, pp. 219–247. DOI: 10.1111/ecca.12008.

Fasan, Marco et al. (2016). "Language FTR and Earnings Management: International Evidence". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2763922.

Fausey, Caitlin M. et al. (2010). "Constructing agency: the role of language". In: *Frontiers in Psychology* 1.OCT. DOI: 10.3389/fpsyg.2010.00162.

Fernandez, Raquel (2007). "Alfred Marshall Lecture : Women, Work, and Culture". In: *Journal of the European Economic Association* 5.2, pp. 305–332.

Fernández, Raquel (2011). "Does culture matter?" In: *Handbook of Social Economics* 1.1 B, pp. 481–510. DOI: 10.1016/B978-0-444-53187-2.00011-5.

– (Feb. 2013). "Cultural Change as Learning: The Evolution of Female Labor Force Participation over a Century". In: *American Economic Review* 103.1, pp. 472–500. DOI: 10.1257/aer.103.1.472.

Fernández, Raquel and Alessandra Fogli (Jan. 2009). "Culture: An Empirical Investigation of Beliefs, Work, and Fertility". In: *American Economic Journal: Macroeconomics* 1.1, pp. 146–177. DOI: 10.1257/mac.1.1.146.

Figlio, David et al. (Nov. 2019). "Long-Term Orientation and Educational Performance". In: *American Economic Journal: Economic Policy* 11.4, pp. 272–309. DOI: 10.1257/pol.20180374.

Fuchs-Schündeln, Nicola, Paolo Masella, and Hannah Paule-Paludkiewics (Aug. 2020). "Cultural Determinants of Household Saving Behavior". In: *Journal of Money, Credit and Banking* 52.5, pp. 1035–1070. DOI: 10.1111/jmcb.12659.

Galor, Oded, Ömer ÖZak, and Assaf Sarid (2018). "Geographical Origins of Language Structures". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3097220.

Galor, Oded, Ömer Özak, and Assaf Sarid (2016). "Geographical Origins and Economic Consequences of Language Structures". In: *IZA Discussion Paper*. IZA Discussion Paper 10379. DOI: 10.2139/ssrn.2820889.

– (May 2020). "Linguistic Traits and Human Capital Formation". In: *AEA Papers and Proceedings* 110, pp. 309–313. DOI: 10.1257/pandp.20201069.

Gang, Ira N., Francisco L. Rivera-Batiz, and Myeng-Su Yun (May 2013). "Economic Strain, Education and Attitudes towards Foreigners in the European Union". In: *Review of International Economics* 21.2, pp. 177–190. DOI: 10.1111/roie.12029.

Gay, Victor et al. (Dec. 2018). "Decomposing culture: an analysis of gender, language, and labor supply in the household". In: *Review of Economics of the Household* 16.4, pp. 879–909. DOI: 10.1007/s11150-017-9369-x.

Goldin, Claudia (Apr. 2014). "A Grand Gender Convergence: Its Last Chapter". In: *American Economic Review* 104.4, pp. 1091–1119. DOI: 10.1257/aer.104.4.1091.

Guin, Benjamin (2016). "Culture and Household Saving". In: *SSRN Electronic Journal* October. DOI: 10.2139/ssrn.2698872.

Gumperz, John J. et al. (1996). *Rethinking Language Relativity*. Ed. by John J. Gumperz and Stephen C Levinson. Cambridge: Cambridge University Press.

Hall, Bronwyn and John Van Reenen (2000). "How effective are fiscal incentives for R&D? a review of the evidence". In: *Research Policy* 29.4-5, pp. 449–469. DOI: 10.1016/S0048-7333(99)00085-2.

Heider, Eleanor R. (1972). "Universals in color naming and memory." In: *Journal of Experimental Psychology* 93.1, pp. 10–20. DOI: 10.1037/h0032606.

Helmbrecht, Johannes (2013). "Politeness Distinction in Pronouns". In: *The World Atlas of Language Structures Online*. Ed. by Matthew Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL: http://wals.info/chapter/45.

Hicks, Daniel L., Estefania Santacreu-Vasut, and Amir Shoham (Feb. 2015). "Does mother tongue make for women's work? Linguistics, household labor, and gender identity". In: *Journal of Economic Behavior & Organization* 110, pp. 19–44. DOI: 10.1016/j.jebo.2014.11.010.

Hjerm, Mikael (1998). "National Identities, National Pride and Xenophobia: A Comparison of Four Western Countries". In: *Acta Sociologica* 41.4, pp. 335–347. DOI: 10.1177/000169939804100403.

Hofstede, Geert (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. 2nd. Thousand Oaks, CA: Sage.

ILO (2020). *These occupations are dominated by women*. URL: https://ilostat.ilo.org/these-occupations-are-dominated-by-women/ (visited on 01/31/2022).

Inglehart, Ronald et al. (2014). *World Values Survey: Round Six - Country-Pooled Datafile Version*. Madrid.

Jayachandran, Seema (Sept. 2021). "Social Norms as a Barrier to Women's Employment in Developing Countries". In: *IMF Economic Review* 69.3, pp. 576–595. DOI: 10.1057/s41308-021-00140-w.

Kim, Jaehyeon, Yongtae Kim, and Jian Zhou (Apr. 2017). "Languages and earnings management". In: *Journal of Accounting and Economics* 63.2-3, pp. 288–306. DOI: 10.1016/j.jacceco.2017.04.001.

Kim, Jungbu (Nov. 2011). "Political Institutions and Public R&D Expenditures in Democratic Countries". In: *International Journal of Public Administration* 34.13, pp. 843–857. DOI: 10.1080/01900692.2011.615051.

Konishi, Toshi (Sept. 1993). "The semantics of grammatical gender: A cross-cultural study". In: *Journal of Psycholinguistic Research* 22.5, pp. 519–534. DOI: 10.1007/BF01068252.

Kovacic, Matija and Cristina Elisa Orso (2016). "Why Do Some Countries Fear Immigration More than Others? Evidence from Europe". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2739274.

Leong, Chan Hoong and Colleen Ward (2006). "Cultural values and attitudes toward immigrants and multiculturalism: The case of the Eurobarometer survey on racism and xenophobia". In: *International Journal of Intercultural Relations* 30.6, pp. 799–810. DOI: 10.1016/j.ijintrel.2006.07.001.

Levinson, Stephen C. (1996). "Frames of reference and Molyneux's question: Crosslinguistic evidence." In: *Language and space*. Ed. by Paul Bloom et al. MIT Press, pp. 109–169.

Li, Peggy and Lila Gleitman (Apr. 2002). "Turning the tables: language and spatial reasoning". In: *Cognition* 83.3, pp. 265–294. DOI: 10.1016/S0010-0277(02)00009-4.

Lucy, John A. (July 1992). *Language Diversity and Thought: A Reformulation of the Linguistic Relativity Hypothesis*. Cambridge University Press. DOI: 10.1017/CBO9780511620843.

– (1996). *Grammatical Categories and Cognition*. Cambridge: Cambridge University Press.

– (Oct. 1997). "Linguistic Relativity". In: *Annual Review of Anthropology* 26.1, pp. 291–312. DOI: 10.1146/annurev.anthro.26.1.291.

– (2016). "Recent Advances in the Study of Linguistic Relativity in Historical Context: A Critical Assessment". In: *Language Learning* 66.3, pp. 487–515. DOI: 10.1111/lang.12195.

Malmkjaer, Kirsten, ed. (Dec. 2009). *The Routledge Linguistics Encyclopedia*. Vol. 48. 02. Routledge, pp. 48–0611–48–0611. DOI: 10.4324/9780203874950.

Mavisakalyan, Astghik and Clas Weber (July 2018). "LINGUISTIC STRUCTURES AND ECONOMIC OUTCOMES". In: *Journal of Economic Surveys* 32.3, pp. 916–939. DOI: 10.1111/joes.12247.

Mayda, Anna Maria (Aug. 2006). "Who Is Against Immigration? A Cross-Country Investigation of Individual Attitudes toward Immigrants". In: *Review of Economics and Statistics* 88.3, pp. 510–530. DOI: 10.1162/rest.88.3.510.

McDonough, Laraine, Soonja Choi, and Jean M. Mandler (May 2003). "Understanding spatial relations: Flexible infants, lexical adults". In: *Cognitive Psychology* 46.3, pp. 229–259. DOI: 10.1016/S0010-0285(02)00514-5.

Ostapczuk, Martin, Jochen Musch, and Morten Moshagen (Oct. 2009). "A randomized-response investigation of the education effect in attitudes towards foreigners". In: *European Journal of Social Psychology* 39.6, pp. 920–931. DOI: 10.1002/ejsp.588.

Pérez, Efrén O. and Margit Tavits (July 2017). "Language Shapes People's Time Perspective and Support for Future-Oriented Policies". In: *American Journal of Political Science* 61.3, pp. 715–727. DOI: 10.1111/ajps.12290.

– (Jan. 2019). "Language Influences Public Attitudes toward Gender Equality". In: *The Journal of Politics* 81.1, pp. 81–93. DOI: 10.1086/700004.

Phillips, Webb and Lera Boroditsky (Sept. 2003). "Can Quirks of Grammar Affect the Way You Think? Grammatical Gender and Object Concepts". In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 25.

Pullum, Geoffrey K. (2012). *Keith Chen, Whorfian economist*. URL: http://languagelog.ldc.upenn.edu/nll/?p=3756 (visited on 02/11/2019).

Raijman, R., Moshe Semyonov, and Peter Schmidt (Sept. 2003). "Do Foreigners Deserve Rights? Determinants of Public Views Towards Foreigners in Germany and Israel". In: *European Sociological Review* 19.4, pp. 379–392. DOI: 10.1093/esr/19.4.379.

Rapp, Marc S., Philipp D. Schaller, and Michael Wolff (2012). "Fördern aktienkursbasierte Vergütungsinstrumente langfristig orientierte Unternehmensentscheidungen? Lehren aus der Kreditkrise". In: *ZfB - Zeitschrift für Betriebswirtschaft* 82.10, pp. 1057–1087. DOI: 10.1007/s11573-011-0615-z.

Roberts, Seán G. and James Winters (Aug. 2013). "Linguistic Diversity and Traffic Accidents: Lessons from Statistical Studies of Cultural Traits". In: *PLoS ONE* 8.8. Ed. by Frank Emmert-Streib, e70902. DOI: 10.1371/journal.pone.0070902.

Roberts, Seán G., James Winters, and M. Keith Chen (July 2015). "Future Tense and Economic Decisions: Controlling for Cultural Evolution". In: *PLoS ONE* 10.7. Ed. by Ramesh Balasubramaniam. DOI: 10.1371/journal.pone.0132145.

Romer, Paul M. (1990). "Endogenous Technological Change". In: *Journal of Political Economy* 98.5, Part 2, pp. 71–102. DOI: 10.1086/261725.

Rosch, Eleanor (Sept. 1975). "Cognitive representations of semantic categories." In: *Journal of Experimental Psychology: General* 104.3, pp. 192–233. DOI: 10.1037/0096-3445.104.3.192.

– (1978). "Principles of categorization". In: *Cognition and categorization*. Ed. by Eleanor Rosch and Barbara Llyod. Hillsdale, NJ: Lawrence Erlbaum, pp. 27–48.

Santacreu-Vasut, Estefania, Oded Shenkar, and Amir Shoham (2014). "Linguistic gender marking and its international business ramifications". In: *Journal of International Business Studies* 45.9, pp. 1170–1178. DOI: 10.1057/jibs.2014.5.

Santacreu-Vasut, Estefania, Amir Shoham, and Victor Gay (Mar. 2013). "Do female/male distinctions in language matter? Evidence from gender political quotas". In: *Applied Economics Letters* 20.5, pp. 495–498. DOI: 10.1080/13504851.2012.714062.

Sapir, Edward (1921). *Language*. New York: Harcourt Brace.

– (Dec. 1949). *Selected Writings of Edward Sapir in Language, Culture, and Personality*. Ed. by David G. Mandelbaum. Vol. 1. 4. University of California Press, p. 361. DOI: 10.1525/9780520324077.

Scheve, Kenneth F. and Matthew J. Slaughter (2001). "Labor market competition and individual preferences over immigration policy". In: *Review of Economics and Statistics* 83.1, pp. 133–145. DOI: 10.1162/003465301750160108.

Scott, Amanda (Dec. 1989). "The vertical dimension and time in mandarin". In: *Australian Journal of Linguistics* 9.2, pp. 295–314. DOI: 10.1080/07268608908599424.

Sera, Maria D., Christian A.H. Berge, and Javier del Castillo Pintado (July 1994). "Grammatical and conceptual forces in the attribution of gender by English and Spanish speakers". In: *Cognitive Development* 9.3, pp. 261–292. DOI: 10.1016/0885-2014(94)90007-8.

Siewierska, Anna (2013). "Gender Distinction in Independent Personal Pronouns". In: *The World Altlas of Language Structures Online*. Ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL: https://wals.info/chapter/44.

Slobin, Dan I. (Sept. 1987). "Thinking for Speaking". In: *Annual Meeting of the Berkeley Linguistics Society* 13.January, p. 435. DOI: 10.3765/bls.v13i0.1826.

Slobin, Dan I. (1996). “From "Thought and Language" to "Thinking for Speaking"”. In: *Rethinking Linguistic Relativity*. Ed. by J. J. Gumperz and S. C. Levinson. Cambridge: Cambridge University Press, pp. 70–96.

Solow, Robert (1956). “Technical Change and the Aggregate Production Function”. In: *The Review of Economics and Statistics* 39.3, pp. 312–320.

Stutterheim, Christiane von and Ralf Nüse (Jan. 2003). “Processes of conceptualization in language production: language-specific perspectives and event construal”. In: *Linguistics* 41.5. DOI: 10.1515/ling.2003.028.

Stutterheim, Christiane von et al. (Jan. 2012). “How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance”. In: *Linguistics* 50.4, pp. 833–867. DOI: 10.1515/ling-2012-0026.

Swamy, P. A. V. B. and S. S. Arora (Mar. 1972). “The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models”. In: *Econometrica* 40.2, p. 261. DOI: 10.2307/1909405.

The World Bank (n.d.). *World Development Indicators*. Washington, D.C. URL: http://data.worldbank.org/data-catalog/world-development-indicators.

Thieroff, Rolf (2000). “On the Areal Distribution of Tense-Aspect Categories in Europe”. In: *Tense and Aspect in the Languages of Europe*. Ed. by Östen Dahl. Berlin: Mouton de Gruyter, pp. 309–328.

Van der Velde, Lucas, Joanna Tyrowicz, and Joanna Siwinska (2015). “Language and (the estimates of) the gender wage gap”. In: *Economics Letters* 136, pp. 165–170. DOI: 10.1016/j.econlet.2015.08.014.

Whorf, Benjamin Lee (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. Ed. by John B. Carroll. Cambridge, MA: MIT Press, p. 278.

Winawer, Jonathan et al. (May 2007). “Russian blues reveal effects of language on color discrimination”. In: *Proceedings of the National Academy of Sciences* 104.19, pp. 7780–7785. DOI: 10.1073/pnas.0701644104.

Wolff, Phillip and Kevin J. Holmes (May 2011). “Linguistic relativity”. In: *Wiley Interdisciplinary Reviews: Cognitive Science* 2.3, pp. 253–265. DOI: 10.1002/wcs.104.

Wooldridge, Jeffrey M. (2010). *Econometric Analysis of Cross Section and Panel Data*. 2nd. Cambridge, MA: MIT Press.

Declaration for Admission to the Doctoral Examination

I confirm

- that the dissertation "Linguistic Relativity and Economic Outcomes - How Language Influences Thought and Behaviour of Individuals" that I submitted was produced independently without assistance from external parties, and not contrary to high scientific standards and integrity,
- that I have adhered to the examination regulations, including upholding a high degree of scientific integrity, which includes the strict and proper use of citations so that the inclusion of other ideas in the dissertation are clearly distinguished,
- that in the process of completing this doctoral thesis, no intermediaries were compensated to assist me neither with the admissions or preparation processes, and in this process,
  - No remuneration or equivalent compensation were provided
  - No services were engaged that may contradict the purpose of producing a doctoral thesis
- that I have not submitted this dissertation or parts of this dissertation elsewhere.

I am aware that false claims (and the discovery of those false claims now, and in the future) with regards to the declaration for admission to the doctoral examination can lead to the invalidation or revoking of the doctoral degree.

Signature: 

Date: 21.02.2022

Author Contributions

The main part of the thesis builds on three research papers. My contributions to the three papers are as follows:

Paper 1: *Language and the Future: Board Members and the Investment in the Future*

- This study is single-authored.

Paper 2: *Language and Xenophobia: The Effect of Politeness Distinction in Pronouns on your Attitude towards Foreigners*

- This study is single-authored.

Paper 3: *Language and Gender: How Linguistic Differences Influence a Woman's Labour Market Outcomes*

- This study is single-authored.

Signature: