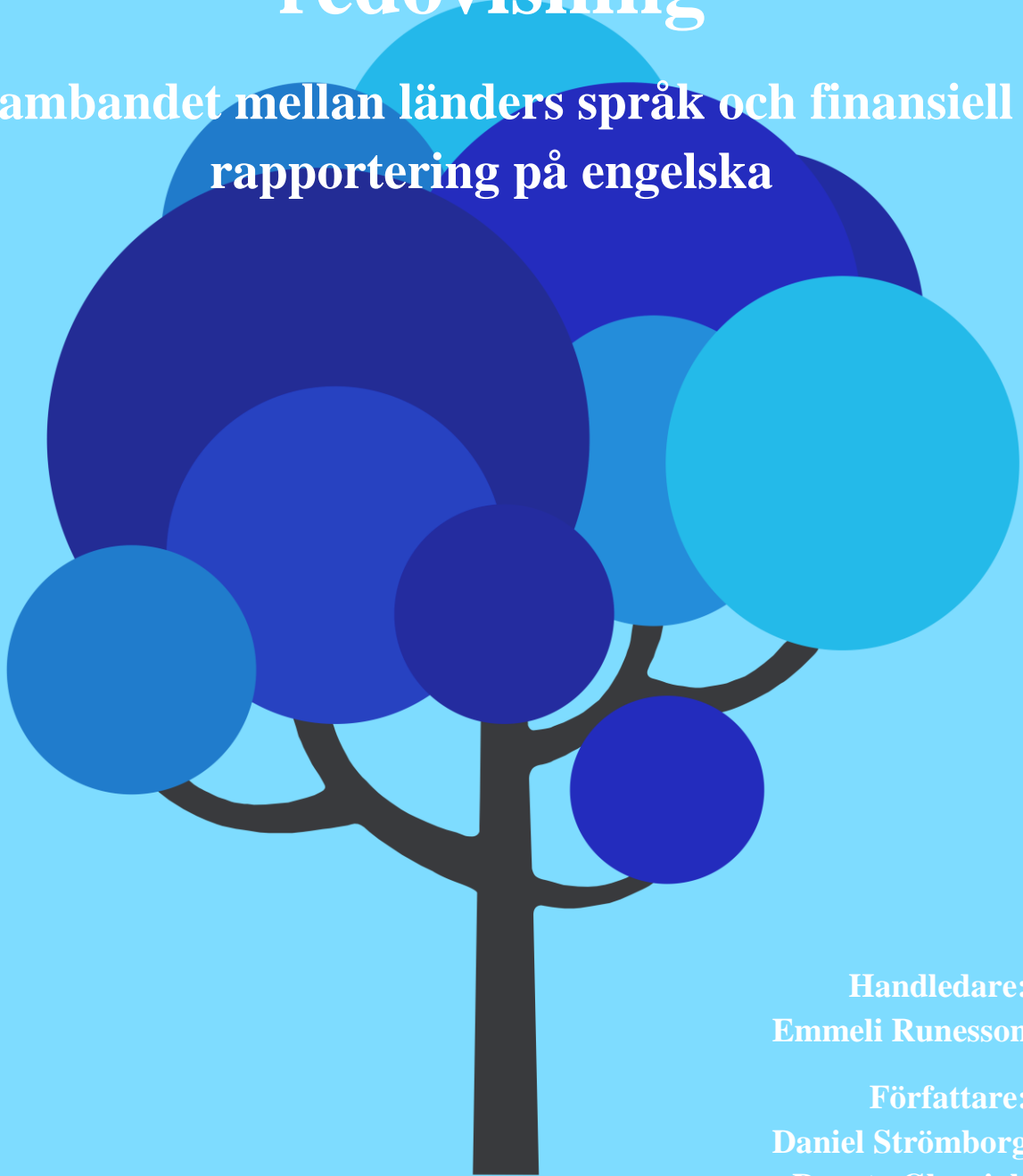


# Läsbarhet av internationell redovisning

Sambandet mellan länders språk och finansiell  
rapportering på engelska



Handledare:  
Emmeli Runesson

Författare:  
Daniel Strömborg  
Dorota Glosniak

Magisteruppsats inom externredovisning  
Handelshögskolan vid Göteborgs Universitet  
Vårterminen 2018



# Förord

I början av terminen var vi två förvirrade studenter utan någon idé om vad vår uppsats skulle handla om. Ett tjugotal potentiella uppsatsämnen avvägdes och förkastades i sökandet efter något som både var roligt och givande att skriva om. Veckorna gick och uppsatstiden närmade sig med stormsteg, men vi hade fortfarande inte kommit närmare att välja ett ämne. Då fann vi det. I en beskrivning av tillgängliga handledare föreslog universitetslektorn, Emmeli Runesson, läsbarhet som ett möjligt forskningsområde. Inom en dag var vi båda överens - vår uppsats skulle handla om läsbarheten av årsredovisningar. Vi tog kontakt med Emmeli i samband med en metod-föreläsning som hon höll för framtida uppsatsskrivare och frågade henne om råd. Hon har varit ett stort stöd i uppsatsskrivandet ända sedan dess. Vi vill därför först och främst tacka vår handledare för det engagemang hon har visat under uppsatsskrivandet och för att hon alltid har varit där med goda råd när vi har varit vilse.

Vi vill även rikta ett stort tack till de andra uppsatsskrivande grupper som under seminarier och vid informella konversationer kommit med feedback på uppsatsen och bollat idéer med oss. Ni vet vilka ni är.

Slutligen vill vi även tacka Anne-Charlotte Strömborg för att hon vid två tillfällen under uppsatsskrivandets gång tog tid ur sin vardag för att korrekturläsa vår uppsats och komma med förslag på förbättringar.

Utan er alla hade aldrig denna uppsats varit möjlig. Tack!

Göteborg, Maj 2018

Daniel Strömborg och Dorota Glosniak

PS: Denna Uppsats har fått ett Fog Index på 15,9 och ett Flesch-Kincaid Index på 14,9. Om läsbarhetsmåttén även fungerar för det svenska språket så innebär det att uppsatsen är läsbar för en student i slutet av sin kandidatutbildning.

# Abstract

**Master thesis in Accounting (one year), Gothenburg School of Business, Economics and Law, Business Department, spring term 2018**

**Authors:** Daniel Strömborg and Dorota Glosniak

**Advisor:** Emmeli Runesson

**Title:** Readability in International Accounting - The Relationship Between Countries' Languages and Financial Reporting in English

**Background & problem:** One of the main purposes of financial accounting is to provide investors with decision-relevant information. One way that investors gain access to this information is through the written disclosures included with the financial statements. How readable those disclosures are has been shown to affect how investors value companies. Further, increased internationalization of the stock markets has led to an increasing number of companies choosing to report their financials in English. Studies, however, have indicated that there is a difference between countries in the readability of these reports.

**Purpose:** The purpose of this study is to investigate whether the majority language of a country determines the readability of the notes included in companies' English-language financial reports.

**Delimitations:** The study has been delimited to only include listed IFRS-compliant European companies established in the energy sector. Further, the study only analyzes one note found within the financial reports, the note *Accounting Principles*.

**Method:** 419 consolidated financial statements covering the financial years 2014-2016 have been collected. Their readability has been graded using three measures: Fog Index, Flesch-Kincaid Index and Total Number of Words. Further, the different languages of Europe have been divided into seven language groups based on their genetic relationships to each other. Analysis of covariance (ANCOVA) and regression analysis have been used to determine whether there is a difference in readability between language groups, and whether this difference is linked to how closely related the languages are to English.

**Conclusion:** The results indicate that there is a difference in the readability of financial reports between companies that is dependent on the languages of companies' home countries. Furthermore, we find evidence that companies from countries whose languages are less related to English write in a way that is less readable.

**Proposals for future research:** A similar study that also includes the effects of borrowing between languages could serve to increase the understanding of the role languages play in determining the readability of English financial reports. Further, a similar study that also examines syntax could serve to corroborate, or contest, the findings of this study.

**Keywords:** Readability, Textual Analysis, Linguistics, Communication, Language Barriers, Disclosures

# Sammanfattning

**Examensarbete i företagsekonomi, Handelshögskolan vid Göteborgs universitet, Magisteruppsats i Externredovisning, VT 18**

**Författare:** Daniel Strömborg och Dorota Glosniak

**Handledare:** Emmeli Runesson

**Titel:** Läsbarhet av internationell redovisning - Sambandet mellan länders språk och finansiell rapportering på engelska.

**Bakgrund och Problem:** Ett av redovisningens syften är att förse investerare med beslutsrelevant information. Ett sätt som investerare får tillgång till denna information är genom de skriftliga upplysningar som inkluderas i de finansiella rapporterna. Hur lättlästa upplysningar är har visat sig påverka hur investerare värderar företag. En ökad internationalisering av aktiemarknaden har samtidigt lett till att allt fler företag väljer att redovisa på engelska. Det har dock visat sig att det finns skillnader mellan länder på läsbarheten av de engelska rapporterna.

**Syfte:** Att undersöka om majoritetsspråket i ett land är avgörande för läsbarheten av noterna i företags engelskspråkiga finansiella rapporter.

**Avgränsningar:** Studien har avgränsats till noterade europeiska företag inom energisektorn som följer IFRS. Vidare har studien även begränsats till endast en not i de finansiella rapporterna, noten *Redovisningsprinciper*.

**Metod:** 419 koncernredovisningar har samlats in för räkenskapsåren 2014–2016. Deras läsbarhet har graderats med tre mått: Fog Index, Flesch-Kincaid Index och Totalt antal ord. Vidare har Europas språk delats in i sju språkgrupper baserat på deras genetiska släktskap till varandra. Kovariansanalys (ANCOVA) och regressionsanalys har använts för att se om det finns någon skillnad på läsbarheten mellan språkgrupper och om detta har ett samband med hur nära besläktade språkgrupperna är till det engelska språket.

**Slutsatser:** Resultaten indikerar att det finns en skillnad på läsbarheten av finansiella rapporter på engelska mellan företag som beror på språket i företagets hemland. Vidare finner vi bevis på att företag från länder vars språk är mindre besläktat med engelska skriver mindre läsbart.

**Förslag till vidare forskning:** En liknande studie som inkluderar effekten av lånande mellan språk hade tjänat att öka förståelsen av den roll som språk spelar för läsbarheten av finansiell rapportering på engelska. Vidare hade en liknande studie som även undersöker syntax tjänat att bekräfta eller bestrida upptäckterna i denna studie.

**Nyckelord:** Readability, textanalys, lingvistik, kommunikation, språkbarriärer, upplysningar

# Förkortningar

IAS	International Accounting Standards
IASB	International Accounting Standards Board
ICB	Industry Classification Benchmark
IFRS	International Financial Reporting Standards
ISIN	International Securities Identification Number
SEC	United States Securities and Exchange Commission

# Innehållsförteckning

1. Inledning .....	1
1.1 Bakgrund.....	1
1.2 Problemdiskussion .....	3
1.3 Studiens syfte & hypoteser .....	4
1.4 Urval och metod.....	4
1.5 Resultat .....	5
1.6 Studiens svagheter .....	5
1.7 Studiens bidrag.....	6
2. Litteraturoversikt och hypotesutveckling .....	7
2.1 Läsbarhet.....	7
2.1.1 Definitionen på läsbarhet .....	7
2.1.2 Traditionella läsbarhetsmått.....	8
2.1.3 Kritik mot traditionella mått .....	8
2.1.4 Dokumentlängd - Ett alternativt läsbarhetsmått .....	9
2.1.5 Tidigare forskning.....	10
2.2 Historisk Lingvistik .....	11
2.2.1 Trädmodellen .....	12
2.2.2 Lånande.....	12
2.2.3 De europeiska språkfamiljerna .....	13
2.3 Hypotesutveckling .....	15
3. Data och Variabler .....	17
3.1 Informationsmedium.....	17
3.2 Urval och Insamling.....	17
3.3 Statistiska analysmetoder.....	20
3.3.1 ANCOVA (Hypotes 1) .....	20
3.3.2 Post-hoc tester (Hypotes 1).....	20
3.3.3 Regressionsanalys (Hypotes 2) .....	21
3.3.4 Heteroskedasticitet och robust regression (Hypotes 2).....	21
3.3.5 Signifikansnivå .....	22
3.4 Variabler .....	22
3.4.1 Beroende variabler - Läsbarhet.....	22
3.4.2 Oberoende variabler - Språk .....	24

3.4.3 Kontrollvariabler och Kovariater .....	25
4. Resultat och analys .....	26
4.1 Beskrivande statistik .....	26
4.2 Hypotes 1 - Finns det en skillnad på läsbarheten mellan språkgrupper? .....	27
4.2.1 Post-hoc tester .....	28
4.3 Hypotes 2 - Finns det ett samband mellan släktskap till engelska och läsbarhet? .....	31
4.3.1 Kontrollvariablerna .....	34
4.4 Företagsspecifika fasta effekter .....	34
5. Diskussion .....	35
6. Slutsatser .....	36
7. Förslag till fortsatt forskning .....	37
8. Källförteckning .....	38
Böcker .....	38
Vetenskapliga tidskrifter .....	38
Övriga källor .....	41
Bilaga 1 – Validitetstest av Readable.io .....	i
Bilaga 2 – Beskrivande statistik per språkgrupp .....	ii
Bilaga 3 – Beskrivande statistik per land .....	vi
Bilaga 4 – Parvis korrelation .....	viii
Bilaga 5 – Icke-robusta regressioner .....	ix
Bilaga 6 – H1 och H2, utan 2015-2016 .....	x

# 1. Inledning

## 1.1 Bakgrund

Syftet med redovisning är att förse intressenter med information som är användbar för deras beslutsfattande. Bland annat ska resultatet och den finansiella ställning som presenteras i redovisningen hjälpa investeraren att ta beslut när denne överväger att köpa aktier i ett företag (Marton, 2013). Det är dock inte bara informationen i resultat-, balansräkning och kassaflödesanalys som är av vikt för beslutsfattandet. Även de skriftliga upplysningar som presenteras i de finansiella rapporterna påverkar investeraren. Därför är det viktigt att noter och övriga delar av rapporterna är utformade så att investeraren enkelt kan ta till sig informationen i dessa texter. Enligt Hoogervorst (IFRS Foundation, 2017a), ordförande på International Accounting Standards Board (IASB), kallas detta för effektiv kommunikation. Motsatsen, dåligt utformade texter och ineffektiv kommunikation, kan istället leda till att investeraren förbiser relevant information eller helt enkelt inte förstår informationen som presenteras. Hoogervorst varnar för att detta kan leda till att investeraren blir osäker om företagets framtidsutsikt, vilket leder till högre kapitalkostnader för företaget.

Både i praktiken och inom forskningen diskuteras effektiviteten av att kommunicera information i text utifrån begreppet läsbarhet (SEC, 1998; Li, 2008; Miller, 2010). Läsbarhet har definierats av Dave och Chall (1948) som “... *den totala summan av (inklusive interaktioner mellan) alla beståndsdelar i ett specifikt tryckt material vilket påverkar framgången en grupp läsare har att förstå den.*” [egen översättning] (Citerad i Tekfi, 1987). Baserat på denna definition har Loughran och McDonald (2014) utifrån ett redovisningsbaserat perspektiv senare definierat läsbarhet som “... *förmågan hos individuella investerare och analytiker att assimilera värderrelevant information från en finansiell upplysning*” [egen översättning].

I samstämmighet med Hoogervorsts utlåtande (IFRS Foundation, 2017a) har läsbarheten av finansiella rapporter visat sig påverka investerares bedömningar. Lawrence (2013) finner att investerare är mer troliga att investera i företag med årsredovisningarna som innehåller kortare texter och mindre komplicerade meningar. Det har även visat sig att mer svårlästa årsredovisningar är associerade med mindre handel av ett företags aktier. Detta på grund av att små investerare blir mindre villiga att riskera sina pengar i företaget (Miller, 2010). Kontrollerade experiment finner även att läsbarheten av finansiella dokument påverkar hur starkt investerare reagerar på informationen i texten (Rennekamp, 2012; Tan, Wang & Zhou, 2015). Exempelvis finner Rennekamp (2012) att små investerare reagerar starkare när upplysningar är mer lättlästa, och därför värderar de ett företag högre vid positiva nyheter och lägre vid negativa nyheter när texterna är mer lättlästa. Detta ser hon som stöd för att läsbarhet påverkar investerares förmåga att bearbeta texters innehåll.

Även företagen som upprättar de finansiella rapporterna verkar ha en förståelse för de skriftliga upplysningarnas betydelse för investerares beslutsfattande. Tidigare studier visar



att det finns ett samband mellan redovisade siffror och läsbarheten av texter i företags årsredovisningar (Li, 2008; Lo, Ramos & Rogo, 2017). Li (2008) finner bevis för att företag med lägre redovisat resultat tenderar att ha årsredovisningar som är mer svårlästa. Samtidigt visar han att företag med mer lättlästa årsredovisningar även har högre *earnings persistence*<sup>1</sup>. Dessa upptäckter har förklarats på två sätt. Enligt Bloomfield (2008) behöver dåligt presterande företag mer text och längre meningar för att kunna förklara situationen för deras investerare. Enligt Li (2008) försöker istället företagsledare opportunistiskt utforma årsredovisningarna för att gömma negativ information från investerarna.

Både i USA och Europa har standardsättare och reglerare riktat en allt större uppmärksamhet mot läsbarheten av upplysningar. I USA bemötte tillsynsmyndigheten United States Securities and Exchange Commission (SEC) på 1990-talet problemet med att dessa texter många gånger var för svårlästa. SEC:s åtgärd var att ge ut riktlinjer och rekommendationer i ett projekt som kom att kallas "*Plain Writing Initiative*" (SEC, 2016). I Europa står istället standardsättaren IASB bakom försöken att göra finansiell rapportering mer läsbar. Deras projekt påbörjades 2013 och heter "*Disclosure Initiative*" (Deloitte, u.å.).

År 1998 publicerade SEC handboken "*A Plain English Handbook: How to Create Clear SEC Disclosure Documents*" för att hjälpa företag att minska mängden svårförståeligt fackspråk i dokumenten som de ger ut. Detta sätt att uttrycka sig på kallas för *enkel engelska* (plain English). Enligt SEC är enkel engelska inte ett sätt att ta bort komplicerad information för att göra texter mer förståeliga. Enkel engelska är istället lätt att förstå för att den är klar och koncis. I handboken presenteras tio vanliga problem som gör engelskan i upplysningar mindre enkel: långa meningar, passiv röst, svaga verb, överflödiga ord, juridiskt och finansiellt fikonspråk, en otalig mängd definitioner, abstrakta ord, onödiga detaljer och oläsbar design och layout (SEC, 1998). 1998 införde SEC också ett krav på att vissa avsnitt i prospekten vid börsnotering måste skrivas med enkel engelska. SEC har på senare år även sagt att de vill införa detta krav på fler av de upplysningar som företag ger ut (SEC, 2016). Likt effektiv kommunikation har begreppet enkel engelska kommit att bli nästan synonymt med läsbarhet inom den amerikanska finanslitteraturen. Därmed har läsbarhetsmått kommit att användas för att mäta enkel engelska (Li, 2008; Loughran & McDonald, 2014; Brochet, Naranjo & Yu, 2016).

Startskottet för Disclosure Initiative var 2012, då IASB påbörjade ett korttidsinitiativ. Målet var att utforska möjligheter för hur företag som följer International Financial Reporting Standards (IFRS) kunde förbättra upplysningarna som de lämnar enligt standardens upplysningskrav (Deloitte, u.å.). Detta ledde till att IASB sedan startade projektet "*Disclosure Initiative*" 2013. Centralt i detta projekt var hur kommunikationen i finansiella rapporters noter kunde förbättras. Speciellt berördes hur noternas utformning och språk kunde förenklas, då många läsare hade svårt att förstå noternas innehåll (IFRS Foundation, 2017b).

---

<sup>1</sup> Earnings persistence definieras av Bloomfield (2008) som kontinuiteten och hållbarheten av resultatet från period till period. Högre earnings persistence innebär förmågan att behålla nuvarande resultat över flera perioder.

Att både forskare och praktiker har riktat allt mer uppmärksamhet åt läsbarheten av upplysningar från företag, tyder på att läsbarhet får en allt större betydelse inom både finans och redovisning.

## 1.2 Problemdiskussion

I takt med den ökande internationaliseringen har det engelska språket blivit allt mer av ett lingua franca såväl externt och internt för företag (Nickerson, 2005). Samtidigt har aktiemarknaden i takt med globaliseringen ökat i betydelse som en finansieringskälla i länder där det tidigare inte har funnits betydande kapitalmarknader. Detta har inneburit ett ökat tryck mot en mer harmoniserad redovisning, i hopp om att underlätta jämförelser mellan företag. Inom EU har detta visat sig i form av införandet av lagkrav på att alla företag som är noterade på reglerade kapitalmarknader måste följa IFRS. Allt fler företag i Europa redovisar därför idag enligt IFRS (Marton, Lundqvist & Pettersson, 2016, s.5). Det finns även stöd för att allt fler företag, från icke engelskspråkiga länder, väljer att frivilligt använda engelska i sina finansiella rapporter. Dessa företag motiverar valet att redovisa på engelska med ökad utlandsförsäljning och bättre tillgång till investerare på den internationella kapitalmarknaden (Jeanjean, Lesage & Stolowy, 2010). Positiva följder för företagen har varit ett ökat utländskt ägande, lägre kapitalkostnader och ett ökat följande från analytiker (Jeanjean, Stolowy, Erkens & Yohn, 2014).

I en studie som tar nytta av denna översättningstrend fokuserar Lundholm, Rogo och Zhang (2014) på läsbarheten av utländska företags<sup>2</sup> obligatoriska (årsredovisningar) och frivilliga upplysningar (pressmeddelanden). De antar och finner att utländska företag, för att motarbeta investerarens partiskhet till hemlandet, skapar mer läsbara upplysningar ju längre bort företagen är från USA gällande redovisningsstandarder, investerarskyddslagar och geografiskt avstånd. Överlag finner Lundholm et al (2014) också att företag i icke engelsktalande länder generellt skriver mer läsbart. Denna skillnad finns inte bara mellan engelsktalande och icke-engelsktalande företag. Det finns även bevis på en skillnad av läsbarheten mellan de icke-engelsktalande företagen. När Courtis och Hassan (2002) undersöker skillnader i läsbarheten av årsredovisningar översatta från kinesiska till engelska, respektive malaysiska till engelska, finner författarna att de malaysiska årsredovisningarna är mer läsbara än de kinesiska.

Rollen som språket i ett företag spelar för effektiviteten av engelsk kommunikation har även undersökts djupare i en studie av frivilliga muntliga upplysningar. Brochet et al (2016) undersöker hur språkbarriärer påverkar kapitalmarknadens reaktioner på upplysningar som lämnas i konferenssamtal. De antar och finner att samtal med chefer från länder där språket är mindre besläktat med engelska är mer troliga att innehålla engelska som är *mindre enkel* (non-plain English) och som innehåller fler oriktiga uttryck. Brochet et al (2016) finner att dessa två faktorer påverkar kapitalmarknaden. Kapitalmarknadens reaktioner på *mindre enkel*

---

<sup>2</sup> Företag som är noterade i USA trots att de är från länder där engelska inte talas som förstaspråk av en majoritet av befolkningen.

*engelska* och mer oriktiga uttryck är mer negativ när företaget är lokaliserat i ett icke engelsktalande land.

Djupare än så har dock inte relationen mellan upplysningars läsbarhet och engelsk rapportering från icke engelskspråkiga företag studerats. Frågan om ett lands språk har en inverkan på läsbarheten av de obligatoriska upplysningarna i engelska årsredovisningar är därför fortfarande obesvarad.

Parallellt med redovisningsforskningen, och mer generellt ekonomivetenskapen, har språkvetenskapen historisk lingvistik sedan 1800-talet analyserat och sorterat världens olika språk in i språkfamiljer. Inom dessa familjer är alla medlemmar mer eller mindre besläktade med varandra (Beekes, 1995, s.7; Anttila, 1989, s.300). Tre huvudsakliga språkfamiljer kan urskiljas till vilket nästan alla kontemporära europeiska språk hör: Indo-Europeiska, Finsk-Ugriska och Afro-Asiatiska (Ruhlen, 1994, s.15; Anttila, 1989, s.300). Genom att föra samman redovisningslitteraturen med upptäckter inom den historiska lingvistikens finns möjligheten att få svar på vilken roll som majoritetsspråket i ett land spelar för läsbarheten av företags finansiella rapporter.

### 1.3 Studiens syfte & hypoteser

Syftet i denna studie är att undersöka om majoritetsspråket i ett land inom Europa är avgörande för läsbarheten av företagets engelska finansiella rapporter. Baserat på historisk lingvistik samt tidigare litteratur om läsbarhet inom finans och redovisning har vi tagit fram två hypoteser som prövas. Hypoteserna baseras på språkgrupper och släktskap som har identifierats mellan språk av den historiska lingvistikens.

**Hypotes 1:** Det finns en skillnad på läsbarheten av den engelska årsredovisningen mellan olika språkgrupper.

**Hypotes 2:** Det finns ett samband mellan språkens släktskap till engelska och läsbarheten av den engelska årsredovisningen.

### 1.4 Urval och metod

För att testa dessa hypoteser samlar vi in 419 koncernredovisningar för räkenskapsåren 2014–2016 från noterade europeiska företag inom energisektorn. Från koncernredovisningarna extraheras sedan noten *Redovisningsprinciper* till textfiler. **Hypotes 1** testas med hjälp av en kovariansanalys (ANCOVA), där medelvärdet av läsbarheten för varje språkgrupp jämförs för att se om de signifikant skiljer sig åt. Fem kovariater inkluderas i analysen för att ta hänsyn till bland annat företagets storlek, lönsamhet och investeringsmöjligheter. **Hypotes 2** testas istället med hjälp av linjär regressionsanalys, där släktskapet till det engelska språket används som oberoende variabel för att förklara skillnader i textfilers läsbarhet mellan företag. Här inkluderas fem kontrollvariabler, vilka är samma som kovariaterna i hypotes 1.

Vi graderar varje textfils läsbarhet med hjälp av molntjänsten Readable.io. Läsbarheten mäts med tre variabler. Den första är *Fog Index*, vilken statistiskt kombinerar antalet ord per

mening och antalet stavelser per ord för att skapa ett mått på läsbarhet. Den andra är *Flesch-Kincaid Index* vilken, likt Fog Index, kombinerar antalet ord per mening och antalet stavelser per ord för att skapa ett läsbarhetsmått. Både Fog Index och Flesch-Kincaid Index föreslår, allt annat lika, att mer stavelser per ord eller mer ord per mening gör en text svårare att läsa (DuBay, 2004). Det tredje måttet är *dokumentlängd* (antalet ord i texten). Resonemanget bakom detta mått är att längre dokument är mer avskräckande att läsa och att det behövs mer tid och resurser för att bearbeta informationen i texten (Loughran & McDonald, 2014).

## 1.5 Resultat

Vid jämförelsen av läsbarheten på noten *Redovisningsprinciper* indikerar alla tre mått att det finns en skillnad mellan språkgrupper. Detta är bevis för att det finns en skillnad på hur läsbart företagen skriver sina engelska finansiella rapporter som beror på språket i företagens hemland. När språkgrupperna sedan parvist jämförs mot varandra finner vi dock att språkgrupperna bara skriver mer eller mindre läsbart än varandra i högst hälften av fallen. Det går därför inte alltid att påstå att företag från en språkgrupp skriver mer läsbart än företag från en annan. Efterföljande regressionsanalyser indikerar att företag vars språk är mer avlägset besläktat med engelska skriver mindre läsbart. De nyttjar både längre meningar och ord med fler stavelser, samt använder fler ord i texten totalt. Detta kan förklaras med att dessa företag har det svårare att uttrycka sig på engelska eller att deras egna språk påverkar hur de skriver på engelska, vilket försämrar rapportens läsbarhet.

## 1.6 Studiens svagheter

Vi vill även belysa svagheter i denna studie. Det finns speciellt fyra svagheter som läsaren bör känna till:

De mjukvaror som värderar texters läsbarhet har olika algoritmer för att identifiera stavelser i ord och hur lång en mening faktiskt är. Detta innebär från vår erfarenhet att läsbarhetsmättet av en och samma text kommer att variera mellan mjukvarorna. Då studien använder sig av den kommersiellt tillgängliga tjänsten Readable.io finns ingen information om dess algoritmer tillgängligt. Hur läsbarheten faktiskt mäts ligger därför utanför studiens kontroll. Detta är av betydelse för både studiens validitet (att det faktiskt är läsbarheten som mäts) och dess reliabilitet (att studiens resultat är tillförlitliga och kan återskapas med andra mjukvaror).

Många studier som undersöker läsbarheten av årsredovisningar belyser omöjligheten i att helt separera den underliggande komplexiteten i företagets verksamhet från själva rapportens komplexitet (Li, 2008; Loughran & McDonald, 2016; Lundholm et al, 2014). Denna studie är inget undantag. Trots att vi försöker lindra problemet genom att avgränsa studien till endast en sektor och inkludera kontrollvariabler i analysen, kan inte problemet undvikas helt.

Vidare, trots att historisk lingvistik är ett moget område inom språkvetenskapen är området fortfarande under ständig utveckling. Nya upptäckter görs ständigt som leder till att existerande teorier och språkfamiljer målas om för att sammanfalla med de nya rönen. Samtidigt är fortfarande många frågor inom historisk lingvistik hett debatterade. För denna

studie innebär det att de språkgrupper och släktskap som utgås från i studien kan visa sig vara felaktiga i framtiden.

Slutligen undersöker vi endast noten *Redovisningsprinciper* i årsredovisningar hos europeiska företag inom energisektorn som följer IFRS. Andra möjliga kontexter inkluderar andra branscher, finansiella upplysningar eller regioner med fler språk. Risken finns att sambanden som vi finner i studien försvinner när någon av dessa tre kontexter ändras.

## 1.7 Studiens bidrag

För det första bidrar denna studie till det växande fältet av forskning om icke-numerisk information i finansiell rapportering, specifikt textanalys med fokus på texters läsbarhet (Li, 2008; Lundholm et al, 2014; Miller, 2010). Den visar på att språkskillnader mellan länder fortfarande är avgörande för skillnader årsredovisningar emellan trots att företag väljer att redovisa på samma språk; engelska.

För det andra ger denna studie även två praktiska bidrag. För upprättare av redovisning upplyser studien om vikten av att upprättarna tar hänsyn till hur skillnader mellan deras modersmål och det engelska språket kan ha en inverka på läsbarheten av den slutliga rapporten. För reglerare visar studien på att det fortfarande är viktigt att ta hänsyn till språkskillnader mellan länder vid framtagandet av standarder och rekommendationer, trots att det är allt fler företag som rapporterar på engelska.

## 2. Litteraturöversikt och hypotesutveckling

### 2.1 Läsbarhet

Forskningen inom finans- och redovisningslitteraturen som använder läsbarhet och läsbarhetsmått faller inom ett område kallat textanalys. Enligt Loughran och McDonald (2016) är textanalyser en underklass av kvalitativa analyser som främst fokuserar på att extrahera mening från text. Dessa undersöker exempelvis texters ton (Tetlock, 2007) eller vilka ämnen som återkommer i text (Frazier, Ingram & Tennyson, 1984). Inom textanalys utgör läsbarhet ett eget område. Den skiljer sig från de andra metoderna genom att läsbarheten, istället för att försöka extrahera mening från texter, eftersträvar att mäta förmågan hos läsaren att uttyda det avsedda meddelandet i texten (Loughran & McDonald, 2016).

#### 2.1.1 Definitionen på läsbarhet

Vad som faktiskt menas med begreppet *läsbarhet* har det visat sig vara svårt att ge en exakt definition på. Läsbarhetsbegreppet och dess olika mått utvecklades från början inom språkvetenskapen för att gradera språket i läroböcker, försäkringskontrakt och instruktioner i militära tillämpningar. Inom dessa områden har den valda definitionen av läsbarhet varit avgörande för vilket mått på läsbarhet som har använts (Tekfi, 1987). Enligt Loughran och McDonald (2014) har det föredragna måttet på läsbarhet inom finans- och redovisningslitteraturen på ett liknande sätt varit beroende av hur författaren har valt att definiera begreppet i den specifika studien. Vi identifierar tre huvudsakliga perspektiv på hur läsbarhet har definierats inom läsbarhetslitteraturen.

Det första perspektivet summeras väl av Klares (1963, s.15) definition på läsbarhet: ”... *enkelheten av förståelse eller begripande på grund av skrivstil...*” [egen översättning]. Enligt DuBay (2004) tenderar detta perspektiv att fokusera på innehåll, koherens och organisering av text. Mått som har tagits fram utifrån detta perspektiv, som exempelvis Fog Index, fokuserar huvudsakligen på meningars längd samt polysyllabiska ord<sup>3</sup>. Enligt Loughran och McDonald (2014) fungerar måtten väl för att gradera texter där dessa två faktorer är de särskiljande egenskaperna. Som exempel, om målet är att jämföra läsbarheten av Pippi Långstrump mot Doktor Glas, är troligen meningars längd och ords komplexitet användbara faktorer att mäta läsbarheten med.

Författare inom det andra perspektivet breddar begreppet i syfte att lägga vikt på den avsedda målgruppen när läsbarhet avgörs. Exempelvis definierar skaparen av SMOG Index, McLaughlin (1968, s.193), läsbarhet som: ”... *graden till vilken en given kategori av människor finner ett visst läsmaterial tilltalande och övergripande.*” [egen översättning]. Enligt detta perspektiv är därmed bakgrundkunskapen som antas hos läsaren viktigare än att försöka få en text att passa en grad av läsbarhet som definieras av en matematisk formel.

---

<sup>3</sup> Ord med många stavelser.

Det sista och bredaste perspektivet exemplifieras väl med ett citat från Dave och Chall (1948): *”I den vidaste bemärkelsen, är läsbarhet den totala summan av (inklusive interaktionerna mellan) alla beståndsdelar i ett specifikt tryckt material vilket påverkar en grupp läsares framgång att förstå den.”* [egen översättning] (Citerad i Tekfi, 1987, s.262). Enligt Loughran och McDonald (2014) är denna definition på läsbarhet lämpligast att utgå ifrån inom redovisningslitteraturen. De anser att finansiella upplysningar, jämfört med annan text, inte lika enkelt går att skilja från varandra, med exempelvis Fog Index, på grund av att de flesta finansiella rapporter använder polysyllabiska ord och likartade skrivstilar. Utifrån detta utvecklar Loughran och McDonald följande definition av läsbarhet (2014, s.1649): *“Förmågan hos individuella investerare och analytiker att assimilera värderrelevant information från en finansiell upplysning”* [egen översättning].

Det är Loughrans och McDonalds (2014) definition på läsbarhet som vi väljer att utgå från i denna studie. Deras definition stämmer väl överens med många andra studier inom området (Li, 2008; Miller, 2010; Lawrence, 2013; Ertugrul, Lei, Qiu & Wan, 2017; Bonsall, Leone, Miller & Rennekamp, 2017). Vidare öppnar detta bredare perspektiv upp för användningen av alternativa mått på läsbarhet, som exempelvis totalt antal ord i en text (Lawrence, 2013) och filstorlek av ett dokument (Ertugrul et al, 2017).

### 2.1.2 Traditionella läsbarhetsmått

Som ovan nämnt har ett flertal läsbarhetsmått tagits fram i olika försök att finna ett mått som kvantitativt kan fånga och gradera texters läsbarhet. Många av de mer välkända måtten på läsbarhet som idag används inom finans- och redovisningslitteraturen togs fram av språkforskare under den senare hälften av 1900-talet för syften som är vitt skilda från granskningen av finansiella upplysningar. Fog Index, SMOG Readability Formula, FORCAST Formula, Dale-Chall Formula, Flesch Reading Ease Score och Flesch-Kincaid Index är bara några av de mer välkända måtten (DuBay, 2004; Loughran & McDonald, 2016). Av dessa är Fog Index, Flesch Reading Ease Score och Flesch-Kincaid Index de mest använda läsbarhetsmåtten av studier inom finans- och redovisningslitteraturen (Bonsall et al, 2017; De Franco, Hope, Vyas & Zhou, 2015; Moreno & Casasola, 2016). Gemensamt för alla tre måtten är att de baseras på någon kombination av det genomsnittliga antalet stavelser i ett ord och det genomsnittliga antalet ord per mening. Enligt Leheavy, Li och Merkle (2011) finns det tre huvudsakliga anledningar till populariteten av Fog Index och Flesch-måtten. För det första är måtten objektiva då de inte baseras på analytikerundersökningar eller åsikter. För det andra erbjuder måtten enkla och välkända formler att mäta läsbarhet med. Slutligen gör måtten det möjligt att studera egenskaperna av finansiella upplysningar bland en stor och mångfaldig grupp företag.

### 2.1.3 Kritik mot traditionella mått

De traditionella läsbarhetsmåtten har dock, enligt du Toit (2017), kritiserats inom språkvetenskapen av ett flertal skäl. För det första fokuserar dessa mått huvudsakligen på längden på ord och meningar. Därav ignoreras viktiga egenskaper av texter såsom ordföljd, stil, format, grafiska element och andra eventuella attribut som inte fångas upp av längden på

ord och meningar. För det andra är måtten oförändrade sedan de skapades i mitten av 1900-talet. Därför tar de inte hänsyn till naturliga förändringar i det engelska språket som har skett sedan deras skapelse. Slutligen har måtten från början sitt ursprung i att utvärdera lässvårigheten av barnböcker och är därför inte särskilt lämpade för vuxet läsningsmaterial.

Mer specifikt inom finans- och redovisningslitteraturen framför Loughran och McDonald (2014) bevis på att mått som Fog Index och Flesch-Kincaid Index kan vara olämpliga redskap att mäta läsbarheten av finansiella rapporter med. De finner att variabeln *genomsnittligt antal ord per mening* ger rimliga empiriska korrelationer med andra läsbarhetsmått. Dock är mätningen av meningars längd i finansiella dokument väsentligt mindre exakt än det är vid mätning av meningars längd i traditionella texter. Vidare finner Loughran och McDonald att *svåra ord* respektive *genomsnittligt antal stavelser per ord* är dåligt specificerade mått gällande finansiella dokument. Baserat på 66 707 observationer av amerikanska årsredovisningar under åren 1994–2011 visar de att alla polysyllabiska ord som är mest vanligt förekommande i årsredovisningen troligen är kända av den typiske investeraren eller analytikern.

På grund av kritiken mot de traditionella läsbarhetsmåten använder många studier inom finans- och redovisningslitteraturen ett antal alternativa läsbarhetsmått, parallellt med de traditionella måtten (Li, 2008; Miller, 2010; Lawrence, 2013; Loughran & McDonald, 2014; Bonsall et al, 2017).

#### 2.1.4 Dokumentlängd - Ett alternativt läsbarhetsmått

*Totalt antal ord*, även kallat *dokumentlängd*, är ett vanligt förekommande alternativt läsbarhetsmått som ofta används i kombination med de traditionella läsbarhetsmåten (Li, 2008; Miller, 2010; Lawrence, 2013; Loughran & McDonald, 2014; Bonsall et al, 2017). Många av dessa studier finner en stark korrelation mellan det totala antalet ord i en text och de traditionella läsbarhetsmåten. Detta tyder på att typerna av mått i viss mån mäter samma sak. Loughran och McDonald (2014) resonerar att mer koncist skrivna dokument är mer troliga att läsas och därmed är det mer troligt att informationen i årsredovisningar kommer att återspeglas i aktiepriser och analytikers prognoser. Enligt författarna är styrkan av *dokumentlängd* att det enklare måttet till stor del undviker de ovan beskrivna mätfel som kan uppstå för traditionella läsbarhetsmått när algoritmer granskar finansiella upplysningar. Själva finner Loughran & McDonald (2014) att längden av text bättre kan förklara aktieprisens volatilitet än de traditionella måtten och att det därmed är ett bättre läsbarhetsmått för finansiella dokument. Enligt Li (2008) finns det dock en svaghet i att använda längden på text som ett mått på läsbarhet jämfört med traditionella mått; Jämfört med traditionella mått är det mer troligt att måttet är korrelerat med mängden faktisk information som lämnas.



## 2.1.5 Tidigare forskning

Många tidigare studier visar på att det finns ett samband mellan redovisade siffror och läsbarheten av upplysningarna i företags årsredovisningar. Li (2008) undersöker relationen mellan läsbarheten av årsredovisningar och företags prestanda samt *earnings persistence*<sup>4</sup>. Han finner att företag med lägre redovisat resultat tenderar att ha årsredovisningar som är mer svårlästa (högt Fog Index och högt totalt antal ord). Bloomfield (2008) noterar att detta kan bero på att dåligt presterande företag behöver mer text och längre meningar för att kunna förklara situationen för deras investerare. Vidare finner Li (2008) även att företagen med mer lättlästa årsredovisningar även har högre *earnings persistence*. I helhet resonerar Li att de två resultaten kan bero på att företagsledare opportunistiskt utformar årsredovisningar för att gömma negativ information från investerarna. Lo et al (2016) bygger vidare på Lis (2008) upptäckter. De undersöker ett specifikt avsnitt i 10-k rapporter,<sup>5</sup> ”*Management Discussion and Analysis*”, och finner att det är mer troligt att företag har manipulerat sina resultat för att slå föregående års resultat, givet att avsnittets text är mer svårläst. Enligt Lo et al (2016) motbevisar detta Bloomfields (2008) förklaring att goda nyheter av sin natur är enklare att kommunicera och visar på att företagsledningens fördunkling av information bidrar till att göra upplysningar mer komplexa.

Vidare ger den existerande litteraturen bevis på att investerare påverkas av årsredovisningars läsbarhet. Exempelvis finner Lawrence (2013) att investerare är mer troliga att investera i företag med årsredovisningar som är kortare och har lägre Fog Index. Liknande finner Miller (2010) att mer komplicerade årsredovisningar (längre och med högre Fog Index) är associerade med mindre handel överlag, och att denna relation uppstår på grund av att små investerare är mindre villiga att köpa dessa företags aktier. Vidare tyder kontrollerade experiment på att läsbarheten av ett finansiellt dokument även är avgörande för hur starkt en investerare reagerar på informationen i texten (Rennekamp, 2012; Tan et al, 2015). Exempelvis finner Tan et al (2015) att när prestandamått är inkonsekventa mellan kvartalsrapporter leder högre läsbarhet, i motsats till lägre läsbarhet, av positiva (negativa) prestandatrender till fler (färre) gynnsamma bedömningar av företagets prestanda från investerare. De finner även att högre läsbarhet, i fallet av inkonsekventa prestandamått, förbättrar investerares förståelse för företagets aktuella prestanda, vilket i sin tur påverkar deras bedömningar om företagets framtida prestanda.

Även långgivare och företagens lånekostnader verkar påverkas av årsredovisningars läsbarhet. Ertugrul et al (2017) undersöker påverkan på företagen från bland annat läsbarheten av deras årsredovisningar. De finner att företag med större filstorlek på årsredovisningen har strängare låneavtal och större framtida aktiekursrisk. Resultatet tyder på att läsbarheten av finansiella upplysningar är relaterat till företagsledningens undanhållande av information. Likartat finner Bonsall och Miller (2017) att mindre läsbara finansiella upplysningar är associerade med sämre kreditbetyg, större skiljaktigheter mellan kreditutvärderingsbyråer och högre lånekostnader.

---

<sup>4</sup> Earnings persistence definieras av Bloomfield (2008) som kontinuiteten och hållbarheten av resultatet från period till period. Högre earnings persistence innebär förmågan att behålla nuvarande resultat över perioder.

<sup>5</sup> En 10-K Rapport är en standardiserad årsredovisning som SEC kräver att alla noterade företag på den amerikanska aktiemarknaden måste ge ut.

Läsbarheten av finansiella dokument har också kopplats till analytikens bevakning och spridningen av deras prognoser. I en undersökning finner Lehavy et al (2011) att företag med mindre läsbara årsredovisningar (högre Fog Index) har fler analytiker som följer aktien, större spridning mellan analytikernas prognoser och lägre precision av dessa prognoser. Författarna argumenterar att ju mer bearbetningskostnaderna ökar för årsredovisningar med mindre läsbar text, desto fler analytiker behövs för att bevaka en aktie i syfte att uppfylla investerarnas krav på information.

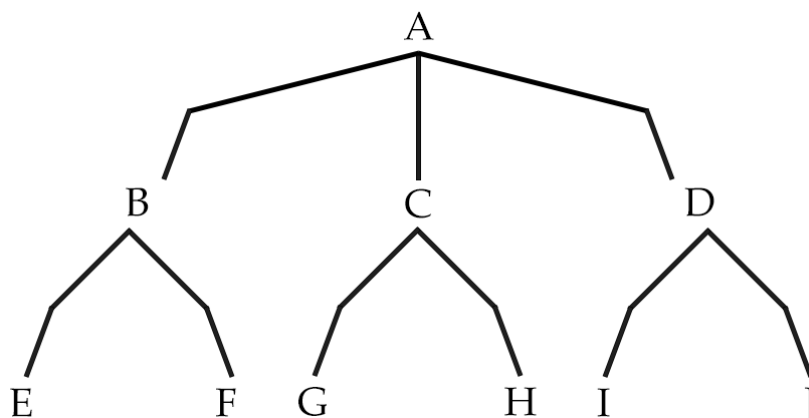
Slutligen har läsbarhet av årsredovisningar och andra upplysningar jämförts mellan länder med och utan engelska som majoritetsspråk. Lundholm et al (2014) jämför årsredovisningar och pressmeddelanden skrivna av utländska företag noterade på den amerikanska aktiemarknaden mot inhemska företag. I studien visar det sig att utländska företag skapar mer läsbara finansiella dokument och att läsbarheten av texterna ökar ju längre bort företagen kommer geografiskt från USA. Läsbarheten ökar även ju mer redovisningsstandarder och investerarskyddslagar i det utländska företagens land skiljer sig från USAs. Lundholm et al (2014) argumenterar att detta beror på att utländska företag har ett större incitament att skapa mer läsbara dokument för att motarbeta partiskhet till hemlandet och locka amerikanska investerare, ju mer deras hemland skiljer sig från USA. Inspirerade av bland annat Lundholm et al (2014) undersöker Brochet et al (2016) hur språkbarriärer påverkar kapitalmarknadens reaktioner på upplysningar som lämnas i konferenssamtal. De antar och finner att samtalen av företag från länder där språket skiljer sig mer från engelska är mer troliga att innehålla engelska som är *mindre enkel* (non-plain English) med fler oriktiga uttryck. Brochet et al (2016) finner att dessa två faktorer påverkar kapitalmarknaden. Kapitalmarknadens reaktioner på *mindre enkel engelska* och mer oriktiga uttryck är mer negativ när företaget är lokaliserat i ett icke engelsktalande land.

## 2.2 Historisk Lingvistik

Språk är en av mänsklighetens viktigaste kommunikationsmedel och används i allt från politiska debatter till upplysningar som lämnas i årsredovisningar av företag. Samtidigt befinner sig alla världens språk i ständig utveckling. Under historiens gång har de splittrats och tagit nya former (Hickey, 2010, s.53). Läran om hur språken har förändrats kallas för *historisk lingvistik* (Beekes, 1995, s.2). Denna vetenskap strävar efter att spåra förändringarna i språken och utforska hur de har uppstått. Sedan tidigt är det inom den historiska lingvistikens känt att det finns vissa specifika likheter mellan olika språk, vilket tyder på att de är relaterade med varandra och har ett gemensamt ursprung. Den gren av den historiska lingvistikens som studerar dessa förhållanden mellan språk heter *komparativ lingvistik* (Beekes, 1995, s.3). För att urskilja systematiska likheter mellan två eller flera språk används den *komparativa metoden*. Detta är den systematiska jämförelsen av former i studerade språk, vilket tillåter en *komparativ rekonstruktion* som approximerar språkens gemensamma ursprungsspråk. Denna återskapade approximation av ursprungsspråket benämns *proto-språk* och språken som härstammar från denna gemensamma stamfader sägs vara *genetiskt relaterade* (Beekes, 1995, s.3; Fortson, 2010, s.3).

## 2.2.1 Trädmodellen

Detta genetiska synsätt används inom *Trädmodellen*. Modellen drar slutsatsen att om språk är relaterade eller besläktade mer varandra så innebär detta att de en gång i tiden var ett och samma språk (Anttila, 1989, s.300). Denna idé avspeglas i modellens komponenter av moder-/faderspråk och dotterspråk, vilket i sin tur avspeglar analogin om språks födelse och död (Hickey, 2010, s.53).



Figur 1 - Trädmodellen

Trädmodellen är allmänt adopterad inom språkvetenskapen och tillsammans med den komparativa metoden är den vanligen använd för att fastställa språkfamiljer. Språkfamiljer är grupper av språk vars släktskap till varandra har identifierats. Modellen bortser från likheter mellan språken som kunde uppstå slumpmässigt och fokus läggs på deras gemensamma egenskaper. Ju fler likheter som finns språken emellan inom en grupp, desto starkare slutsatser kan dras om språkens släktskap. Där två eller fler språk möts representeras av en nod i trädmodellen, vilken motsvarar deras gemensamma moderspråk (Anttila, 1989, s.302; Fortson, 2010, s.4). I modellen kopplas fler och fler språk på detta sätt ihop med noder tills släktskapen mellan språken får utseendet av ett träd, därav modellens namn.

Det är viktigt att nämna att trädmodellen inte tar hänsyn till *lånande* (se avsnitt 2.2.2). Lånande påverkar inte språks genetiska status. Språkens splittringar från varandra är alltid slutliga och vidare inverkar inte lånande på släkttredens koppling språkfamiljer emellan (Hickey, 2010, s.53).

## 2.2.2 Lånande

Lånande är en process genom vilket ett språk inför delar av ett annat språk, huvudsakligen ord, i sitt eget system. Enligt Hoffer (2002) sker denna process alltid i någon grad när två kulturer kommer i kontakt med varandra. Ett flertal faktorer har identifierats som påverkar mängden och takten av lånande. Exempelvis hade invasionen av normander en signifikant betydelse för lånande av ord från franskan till engelskan under medeltiden (Hock & Joseph, 2009). Hoffer (2002) ger ett mycket senare exempel: industrialiseringen har lett till skapandet av kommunikationskanalerna TV och Radio. Detta i kombination med globaliseringen av marknader har lett till att produkter fört termer och namn från skilda språk världen över.

## 2.2.3 De europeiska språkfamiljerna

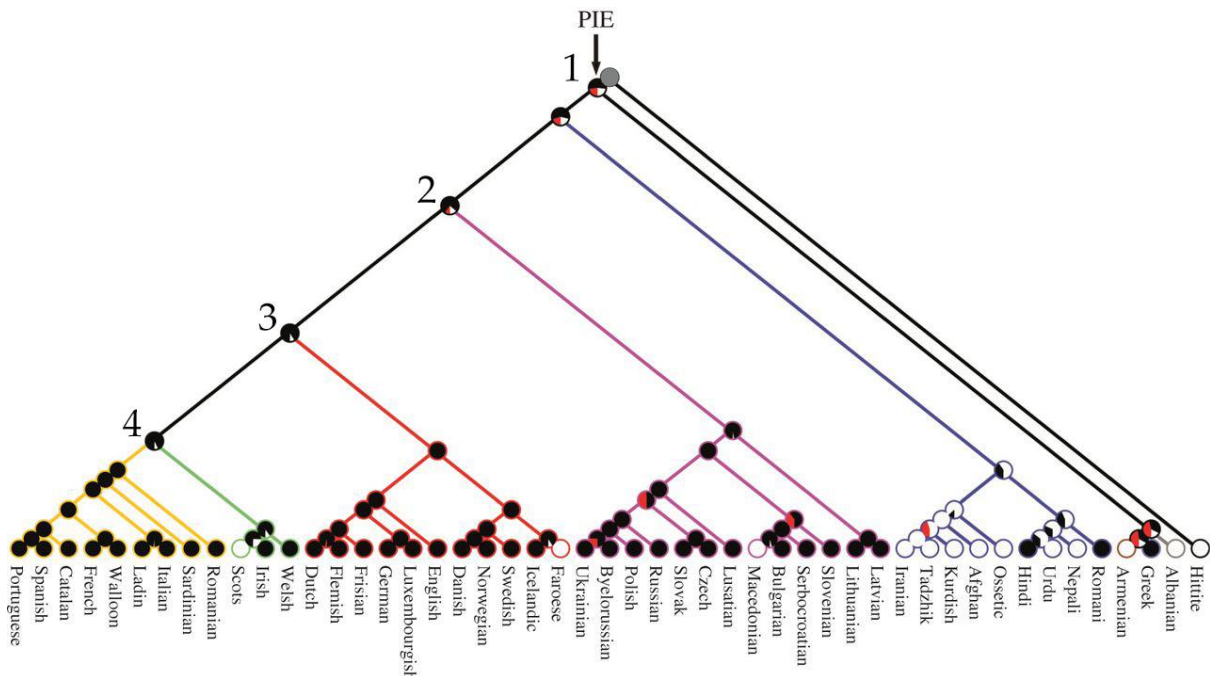
### 2.2.3.1 Indo-Europeiska

Trots att språk troligen har existerat lika länge som det har funnits människor var det först på 1800-talet den *komparativa lingvistik* utvecklades (Beekes, 1995, s.11). Startskottet för läran var 1786, när Sir William Jones inspirerade omvärlden med förslaget att vissa indiska och europeiska språk hade gemensamma egenskaper, vilket tydde på att dessa språk var besläktade (Hock & Joseph, 2009, s.34). Idén om den Indo-Europeiska språkfamiljen fick sedan fäste vid upptäckten av Sanskrit. Detta språk möjliggjorde jämförelser mellan de indiska och europeiska språken (Beekes, 1995, s.13). Idag är den Indo-Europeiska språkfamiljen, som dominerar den europeiska kontinenten, den mest väl undersökta av alla kända språkfamiljer (Anttila, 1989, s.300). Inom Indo-Europeiskan har tio huvudsakliga undergrupper identifierats. De sex inhemska undergrupper som talas av en signifikant andel av den europeiska befolkningen är följande (Fortson, 2010; Hock & Joseph, 2009):

1. De *Italiska språken*, som huvudsakligen utgörs av de romanska språken och latin. De romanska språken delas upp i öst (rumänska) och italo-väst. De Italo-västromanska språken består av de västromanska (franska, spanska och portugisiska) och de italo-dalmatiska (italienska).
2. De *keltiska språken*, vars enda levande undergrupp är den insulära keltiskan. Insulär keltiska består av irländska, walesiska, korniska och det skotska språket.
3. De *germanska språken*, vars levande undergrupper delas upp i nord och väst. De nordgermanska språken representeras av svenska, danska, norska och isländska. De västgermanska språken representeras i sin tur av tyska, nederländska, luxemburgiska och engelska.
4. De *balto-slaviska språken*, som i sin tur består av de baltiska språken och de slaviska. De baltiska språken är lettiska och litauiska. Mer komplicerat är de slaviska språken. Dessa delas upp i öst, syd och väst. Bland de östslaviska språken finns vitryska, ryska och ukrainska. De västslaviska består i sin tur huvudsakligen av tjeckiska, slovakiska och polska. Slutligen delas de sydslaviska språken upp i öst (bulgariska och makedonska) och väst (bosniska, kroatiska, serbiska, slovenska och montenegriska).
5. *Grekiska*, vilket utgör sin egen undergrupp i den Indo-Europeiska språkfamiljen.
6. *Albanska*, vilket likt grekiska utgör en egen undergrupp av Indo-Europeiska.

Hur släktskapen ser ut inom dessa undergrupper råder det en generell konsensus om. Dock är det fortfarande hett omdebatterat hur släktskapen i sin tur ser ut mellan de tio undergrupperna och det råder inte konsensus om något specifikt släkträd. Av denna anledning framställs traditionellt den Indo-Europeiska familjen som ett buskliknande träd från vilket tio huvudgrenar växer från en och samma rot (Beekes, 1995, s.17; Fortson, 2010, s.11). Idag finns det två huvudsakliga konkurrerande teorier om hur de tio undergrupperna är besläktade med varandra och varifrån de härrör. Den första skolan argumenterar för ett ursprung från stäppen norr om det svarta havet (Da Silva & Tehrani, 2016; Hamp, 2013) och den senare argumenterar att Indo-Europeiskan härrör från Anatolien (Bouckaert et al, 2012).

Ett möjligt Indo-Europeiskt släkträd presenteras av Da Silva och Tehrani (2016). I deras studie utgår de ifrån trädmodellen när de med hjälp av folksägner, befolkningshistorier och geografiska avstånd undersöker hur de Indo-Europeiska språken är besläktade. Med hjälp av statistisk analys skapar Da Silva och Tehrani (2016) ett släkträd med ursprung norr om svarta havet (se Figur 2). Av de språk som berörs i vår studie bryts grekiskan ut från de resterande språken vid den första splittringen (1). Detta följs av splittringen mellan de baltoslaviska språken och resterande språk (2). Därefter splittras de resterande språken upp i den germanska respektive den italo-keltiska undergruppen (3). Den senare undergruppen splittras slutligen upp i de italiska de och keltiska språken (4).



Figur 2 - Det Indo-Europeiska släkträdet (da Silva & Tehrani, 2016)

Denna rekonstruktion av den Indo-Europeiska familjen är mycket lik de träd som presenteras i tidigare studier som stödjer ett ursprung norr om Svarta havet (Haak, Lazaridis, Patterson, Rohland, Mallik, Llamas, ..., Reich, 2014; Hamp, 2013). Till och med trädet i studien av Bouckaert et al (2012) som framför bevis på ett Anatoliskt ursprung är förvånansvärt likt da Silvas och Tehranis (2016). De enda stora skillnaderna är när och hur grekiskan kom att splittras från de övriga språken samt var roten av trädet, proto-Indo-Europeiska (PIE), börjar geografiskt.

### 2.2.3.2 Finsk-Ugriska och Afro-Asiatiska

Den andra språkfamiljen som har majoritet i tre europeiska länder är Finsk-Ugriskan. Finsk-Ugriska består av två huvudsakliga undergrupper: ugriska språk, till vilket ungerska tillhör; och finska språk, till vilket finska och estländska tillhör. Jämfört med den Indo-Europeiska språkfamiljen är de Finsk-Ugriska språken unika i att de bland annat saknar könsbaserade genus och att prepositioner skrivs som ändelser på pronomen (Anttila, 1989, s.305).

Slutligen finns det en sista språkfamilj som talas av en majoritet i ett europeiskt land, Afro-Asiatiskan. Denna språkfamilj representeras av maltesiska, vilket tillhör den arabiska undergruppen. Enligt Hickey (2010, s.182) är maltesiskan dock speciell då den är ett *blandat språk*. Ett blandat språk är resultatet av extensivt lånande från andra språk (se avsnitt 2.2.2). I maltesiskans fall utgör den ursprungliga arabiska basen endast en tredjedel av ordförrådet, italienska och sicilianska utgör runt hälften och slutligen utgör engelskan cirka tjugo procent av ordförrådet (Hickey, 2010, s.183).

Ett flertal olika släktskap mellan de Indo-Europeiska, Finsk-Ugriska, Afro-Asiatiska och andra språkfamiljerna har föreslagits, men enligt Ruhlen (1994, s.18) har inga av dessa bevisats. Alla världens språkfamiljer antas därför i praktiken vara obesläktade.

## 2.3 Hypotesutveckling

Varje år väljer allt fler företag från icke engelsktalande länder att översätta sina årsredovisningar till engelska, ofta för att få bättre tillgång till den internationella kapitalmarknaden (Jeanjean et al, 2015). I och med detta har det uppkommit frågor inom redovisningsforskningen om kvaliteten av dessa översättningar (t.ex. Courtis & Hassan, 2002; Campbell, Beck & Shrivess, 2005). Likaså har man inom språkforskningen frågat sig om, och hur, en persons modersmål inverkar på dennes engelska. På detta område finner Mauranen (1993) ett flertal skillnader mellan finnar och engelsmäns skrivstil. Bland annat använder finnar färre ord och andra grammatiska böjelser än den genomsnittlige engelsmannen. Liknande finner Valero-Garces (1996) skilda preferenser mellan engelska och spanska författare av akademiska artiklar vad det gäller skrivstil, där spanjorernas texter vanligen innehåller fler ord än engelsmännens.

Tillbaka till finans- och redovisningslitteraturen, finns det även här bevis för att språket i ett land påverkar de frivilliga och obligatoriska upplysningar som företag presenterar för allmänheten. Lundholm et al (2014) jämför årsredovisningar och pressmeddelanden skrivna av utländska företag noterade på den amerikanska aktiemarknaden mot inhemska amerikanska företag. I studien visar det sig att texter från utländska företag överlag är mer läsbara än texterna från engelsktalande länder, trots att författarna kontrollerar för standarder, lagar och geografiskt avstånd. Courtis och Hassan (2002) gör liknande upptäckter. När de undersöker skillnader i läsbarheten av årsredovisningar översatta från kinesiska till engelska, respektive malaysiska till engelska, finner författarna att de malaysiska årsredovisningarna är mer läsbara än de kinesiska.

Givet att modersmålet hos en person har en inverkan på hur denne uttrycker sig på engelska (Mauranen, 1993; Valero-Garces, 1996) och att det finns indikationer för skillnader på läsbarheten av upplysningar beroende på språket i det upprättande företaget (Courtis & Hassan, 2002; Lundholm et al, 2014) resonerar vi följande: Det bör finnas en skillnad på läsbarheten mellan företags engelskspråkiga årsredovisningar som beror på majoritetsspråket i ett land. Vidare, givet att språken i Europa till olika grad är besläktade med likartade språk (se avsnitt 2.2.3), bör denna skillnad även finnas mellan språkgrupper:

**Hypotes 1:** Det finns en skillnad på läsbarheten av den engelska årsredovisningen mellan olika språkgrupper.

Utifrån släktskapen mellan de europeiska språken som presenterades i Avsnitt 2.2.3 är det engelska språket närmare besläktat och mer likt tyska än franska. Franska är i sin tur mer likt engelska än exempelvis polska. Det är ytterst få studier inom finans- och redovisningslitteraturen som utnyttjar det faktum att språk till olika grad är besläktade med varandra. Brochet et al (2016) är hittills de första, och de enda, som utgår ifrån detta avstånd mellan språk. De undersöker språkbarriärers påverkan på kapitalmarknadens reaktioner på upplysningar som lämnas under konferenssamtal. I studien antar de, och finner, att chefer vars modersmål är mer avlägset besläktat med engelskan talar engelska som är *mindre enkel* (non-plain English) och innehåller fler felaktiga uttryck. Detta tyder på att språk som är mindre besläktade med engelska kan vara en barriär som leder till att årsredovisningar blir mindre läsbara. Bevis på ett motsatt samband framför dock Lundholm et al (2014) när de finner att finansiella rapporter från engelsktalande länder har lägre läsbarhet än de från icke-engelsktalande länder. Det finns därmed indikationer, i båda riktningar, på ett samband mellan ett språks släktskap till engelska och läsbarheten av upplysningarna i årsredovisningar:

**Hypotes 2:** Det finns ett samband mellan språkens släktskap till engelska och läsbarheten av den engelska årsredovisningen.

## 3. Data och Variabler

### 3.1 Informationsmedium

Vi fokuserar på de upplysningar som lämnas i IFRS-koncerners årsredovisningar under den vanligt förekommande noten - *Redovisningsprinciper*.<sup>6</sup> Enligt IAS 1, punkt 10, måste fullständiga finansiella rapporter innehålla noter som sammanfattar väsentliga redovisningsprinciper och annan förklarande information. Dessa noter skall bland annat innehålla information om grunden för upprättandet av de finansiella rapporterna samt de särskilda redovisningsprinciper som används (FAR, 2017). Trots att det är upp till företagen själva att avgöra hur de vill strukturera innehållet i noterna, inkluderas nästan alltid noten *Redovisningsprinciper*. Innehållet i noten är också mycket jämförbar företagens årsredovisningar emellan: Vanligt förekommande innehåll är bland annat grunden för rapporternas upprättande, principer för koncernredovisning, omräkning av utländsk valuta, grunden för intäktsredovisning och hanteringen av nya redovisningsprinciper.

Både Li (2008) samt Loughran och McDonald (2016) belyser att traditionella läsbarhetsmått, som Fog Index, har visats påverkas av komplexiteten i det underliggande ämnet som presenteras. Vidare är en svaghet med måttet Totalt antal ord i en text enligt Li (2008) att måttet är korrelerat med mängden faktisk information som lämnas. Överensstämmelsen mellan företag av innehållet i noten *Redovisningsprinciper* innebär att sannolikheten att dessa faktorer påverkar läsbarhetsmåttens begränsas. Av detta skäl är denna not speciellt lämpad för en jämförelse av årsredovisningars läsbarhet mellan IFRS-följande företag.

Det finns dock en nackdel med att utgå från noten *Redovisningsprinciper*. Enligt Lieberman och Asaba (2006) tenderar företag i samma bransch att imitera varandra. Därför finns det en risk att en text som är standardiserad mellan företag vad det gäller innehåll, även är det beträffande skrivstil. Det innebär att läsbarheten av *Redovisningsprinciper* kan vara likartad mellan årsredovisningar trots att årsredovisningarna i helhet skiljer sig åt. För vår studie innebär det en ökad chans att vi inte finner ett samband mellan språkgrupper och läsbarhet och därför inte förkastar nollhypoteserna trots att ett samband faktiskt finns. Det vill säga, det finns en ökad risk för typ-2 fel (Field, 2018, s.82).

### 3.2 Urval och Insamling

Urvalet i denna studie består av alla företag inom energisektorn som är noterade i Europa och har gett ut årsredovisningar för åren 2014–2016. Flertalet tidigare studier har berört problematiken i att skilja komplexiteten av företagets verksamheter från komplexiteten i dokumenten (Li, 2008; Lundholm et al, 2014; Loughran & McDonald, 2014). I en recension av tidigare litteratur på området konstaterar Loughran och McDonald (2016, s. 1198):

---

<sup>6</sup> “Kärt barn har många namn” stämmer väl för noten *Redovisningsprinciper*. Bland de årsredovisningar som vi samlade in har noten bland annat benämnts på engelska som: Accounting Principles, Accounting Policies, General Principles, General Policies, Main Principles, samt Significant Applied Accounting Principles.



“Forskare står inför problemet att separera verksamheten från dokumentet. Dessa problem är sammanflätade eftersom dokumentet försöker beskriva företagets ekonomiska verklighet.” [egen översättning].

Det finns därmed en risk att läsbarhetsmåten inte bara avspeglar dokumentens läsbarhet, utan även komplexiteten i företagens verksamheter. I ett försök att lindra detta validitetsproblem väljer vi att begränsa urvalet till endast energisektorn. Urvalet görs på databasen Thomson Reuters Datastream med ICB-koder som sökkriterium. Industry Classification Benchmark (ICB) är ett system av koder för branschklassificering av företag. Systemet består av fyra siffror, vilka betecknar industri (0XXX-9XXX), supersektor (X5XX-95XX), sektor (XX3X-XX7X), och undersektor (XXX3-XXX8). Av dessa koder utgår vi ifrån två supersektorer: *olja och gas* (0500) samt *Allmännyttigheter* (7500). Därefter gallrar vi undersektorerna för att exkludera rena tjänsteföretag. Detta resulterar i nio unika ICB-koder som används i sökningen. Inom olja och gas motsvarar ICB-koderna *utforskning och produktion* (533), *integrerad olja och gas* (537), *oljeutrustning och tjänster* (573), *pipelines* (577) och *utrustning för förnybar energi* (587). Inom Allmännyttigheter motsvarar koderna *konventionell elektricitet* (7535), *distribution av gas* (7573), *“multi-utilities”* (7575) och *vatten* (7577). Detta resulterar i ett urval av 314 företag och 942 observationer under räkenskapsåren 2014–2016. Från detta exkluderar vi observationer med företag som inte följer IFRS, inte längre är noterade eller saknar data om totala tillgångar i Datastream. Detta resulterar i **639** observationer.

För analysen av noten *Redovisningsprinciper* har vi följande strategi: Vi söker efter alla årsredovisningar på respektive företags hemsida. Om årsredovisningen finns tillgänglig och är på engelska laddar vi ned den. Från varje årsredovisning kopieras sedan noten *Redovisningsprinciper* för hand och utan styckesindelning till textfiler med programvaran Office Word. Enligt Laksmana, Tietz och Yang (2012) är läsbarhetsmått designade för text i menings-format. Mjukvaror som mäter läsbarhet kan därav få problem när texter gör avsteg från detta format. Därför utelämnar vi tabeller, diagram, punktlistor och rubriker när noten kopieras. Kan inte en årsredovisning kopieras på grund av skrivskydd eller dess format utelämnas denna. Likaså utelämnas årsredovisningar som helt saknar noten *Redovisningsprinciper* eller motsvarande noter. Denna process resulterar i att **453** observationer finns kvar.

På grund av skillnader i utformningen av noter mellan olika företag och år, är inte innehållet av noten *Redovisningsprinciper* alltid densamma mellan årsredovisningar. Vi finner att företagen i vissa fall inkluderar nästan alla upplysningar som inte hänvisas till av exempelvis resultaträkningen i denna not. I andra fall väljer företagen att dela upp samma upplysningar i flertalet skilda noter. Av reliabilitetsskäl skapar vi därför en lista med upplysningar som vi alltid kopierar respektive exkluderar från rapporterna oavsett hur noterna är utformade. På detta sätt förhindras att läsbarhetsmåten påverkas av skillnader på årsredovisningarnas notindelningar. Vi inkluderar alltid upplysningar om: grunder för upprättandet av redovisningen; grunder för konsolidering; nya standarder och tolkningar; viktiga uppskattningar, antaganden

och bedömningar; utländsk valuta; samt redovisningspolicyer.<sup>7</sup> Motsatsvis exkluderar vi alltid upplysningar om: finansiella risker; segmentredovisning; samt generell företagsinformation.<sup>8</sup> Förhoppningen är att på detta sätt minska skillnader mellan oss två författare i vad vi kopierar från årsredovisningarna och därmed förhindra en låg interbedömarreliabilitet.

Det finansiella data och aktiemarknadsdata som används i studien hämtas från databasen WRDS Compustat. Då datat är i skilda valutor räknas dessa om till Euro med respektive växelkurs för 2013-12-31. En fast växelkurs används för att förhindra att studien påverkas av eventuella valutaeffekter. Datat förs sedan samman med läsbarhetsdatat från textfilerna. På grund av att Compustat saknar data för 11 av företagen, resulterar detta i att **422** observationer finns kvar.

Slutligen väljer vi att exkludera ett maltesiskt företag med tre observationer för räkenskapsåren 2014–2016 från urvalet. Detta sker av två anledningar. För det första skulle inkluderingen av Malta innebära att endast ett företag representerar en hel språkgrupp (se Avsnitt 3.4.2). För det andra är maltesiska svårklassificerad i den statistiska analysen, då den står ut från övriga språk i urvalet. Den är ett blandat språk med starka influenser från både italienska och engelska. Samtidigt är maltesiska ensamt i Europa att tillhöra språkfamiljen Afro-Asiatiska (se Avsnitt 2.2.3). Sammanfattningsvis resulterar därför studiens slutliga urval i **419** observationer.

***Tabell 1 - Urval & bortfall***

<b>År</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>Totalt</b>
<b>Antal företag</b>	314	314	314	942
<b>Ej IFRS, avnoterad eller DS data saknas</b>	97	101	105	303
	217	213	209	639
<b>Utan rapport</b>	47	55	55	157
	170	158	154	482
<b>Skrivskyddad eller utan not</b>	19	5	5	29
	151	153	149	453
<b>CS data saknas</b>	11	10	10	31
	140	143	139	422
<b>Maltesiska företag</b>	1	1	1	3
<b>Slutligt urval</b>	<b>139</b>	<b>142</b>	<b>138</b>	<b>419</b>

<sup>7</sup> Vanliga engelska rubriker för dessa upplysningar är: Basis for Preparation; Basis for Consolidation; New Standards and Interpretations; Critical Accounting Judgements, Estimates and Assumptions; Currency Translation; samt Significant Accounting Policies.

<sup>8</sup> Vanliga engelska rubriker för dessa upplysningar är: Financial Risk Management; Segment Information; samt General Information.

### 3.3 Statistiska analysmetoder

**Hypotes 1** testas genom jämförelse av genomsnittlig läsbarhet av årsredovisningar mellan språkgrupper, och **Hypotes 2** testas genom granskning av sambandet mellan språks relation till engelska och läsbarhet. För detta krävs två skilda statistiska modeller för respektive hypotes. Båda tester genomförs med hjälp av STATA.

#### 3.3.1 ANCOVA (Hypotes 1)

För att testa **Hypotes 1**, att det finns en skillnad mellan olika språkgrupper vad det gäller läsbarheten av årsredovisningar på engelska, använder vi kovariansanalys (ANCOVA). ANCOVA är en linjär modell som jämför medelvärdet av en beroende variabel mellan två eller fler självständiga grupper. Modellen kontrollerar för att andra variabler än gruppindelningen kan påverka värdet av den beroende variabeln. Detta gör ANCOVA genom att inkludera dessa andra variabler i modellen som *kovariater* (Field, 2018, s.575). ANCOVA-modellen som används i studien är den följande:

$$\begin{aligned} READABILITY_{XJ} = & \mu + T_X + \beta_0(SIZE_{XJ} - \overline{SIZE}) + \beta_1(MTB_{XJ} - \overline{MTB}) \\ & + \beta_2(ROA_{XJ} - \overline{ROA}) + \beta_3(LEVERAGE_{XJ} - \overline{LEVERAGE}) \\ & + \beta_4(YEAR15_{XJ} - \overline{YEAR15}) + \beta_5(YEAR16_{XJ} - \overline{YEAR16}) + \varepsilon_{XJ} \end{aligned}$$

I denna ekvation är den beroende variabeln  $READABILITY_{XJ}$  läsbarheten av observation  $J$  i språkgrupp  $X$ . Variabeln baseras på en av de tre läsbarhetsmått som används i studien (se Avsnitt 3.4.1). Av de oberoende variablerna är  $\mu$  medelvärdet på läsbarheten av alla språkgrupper och  $T_X$  är medelvärdet på läsbarheten av alla observationer i språkgrupp  $X$ . Resterande variabler är kovariater som inkluderas baserat på tidigare litteratur (se Avsnitt 3.4.3). I modellen är  $SIZE_{XJ}$  observation  $J$  av kovariaten  $SIZE$  i språkgrupp  $X$ . Liknande är  $\overline{SIZE}$  medelvärdet på kovariaten  $SIZE$  för alla språkgrupper. Samma resonemang gäller för *samtliga* kovariater i modellen. Vi jämför läsbarheten mellan sju språkgrupper (se Avsnitt 3.4.2).<sup>9</sup>

#### 3.3.2 Post-hoc tester (Hypotes 1)

En nackdel med ANCOVA är dock att testet inte kan avgöra vilka grupper som ligger bakom signifikansen av modellen. Det vill säga, testet kan bara svara på om medelvärdet skiljer sig åt mellan någon av språkgrupperna. Den kan inte avgöra mellan *vilka* språkgrupper som

---

<sup>9</sup> Matematiskt är vår ANCOVA-modell densamma som en linjär regression (Field, 2018, s.525) med följande ekvation:  $READABILITY_X = \beta_0 + \beta_1GERMANIC + \beta_2ITALIC + \beta_3BALTIC + \beta_4SLAVIC + \beta_5GREEK + \beta_6FINNO\_UG + \beta_7SIZE + \beta_8MTB + \beta_9ROA + \beta_{10}LEVERAGE + \beta_{11}YEAR15 + \beta_{12}YEAR16 + \varepsilon$  Där variablerna GERMANIC till och med FINNO\_UGRIC är dummyvariabler som identifierar vilken språkgrupp ett företag tillhör. Engelska är kontrollgruppen som resterande språk jämförs mot.  $\beta_0$  är medelvärdet av READABILITY hos den engelsktalande kontrollgruppen, justerat för kovariaternas påverkan. Liknande är  $\beta_1$ -6 skillnaden i medelvärdet av READABILITY jämfört med kontrollgruppen.

skillnaderna finns. (Field, 2018, s.537). Exempelvis svarar därmed testet inte på om det finns en skillnad på läsbarheten mellan specifikt italiska och slaviska företag.

En lösning på detta problem är att parvis jämföra de olika gruppernas medelvärden tills alla möjliga kombinationer är testade. Detta tillvägagångssätt, med flera tester efter den första analysen, kallas för Post-hoc tester. En nackdel med att parvist jämföra varje grupp är dock att risken för typ-1 fel ökar för testet i helhet med varje enskild jämförelse. Typ-1 fel är risken att finna en signifikant skillnad mellan grupperna, trots att det inte finns någon. Vanligen betecknas denna risk med symbolen  $\alpha$  (Field, 2018, s.550). Av detta skäl försöker de flesta post-hoc tester på olika sätt kontrollera för denna stegring av risken för typ-1 fel. Enligt Field (2018, s.84) är dock nackdelen med dessa åtgärder att de samtidigt höjer risken för typ-2 fel. Sannolikheten för typ-2 fel är i detta fall risken att vi inte finner en skillnad på läsbarheten mellan språkgrupperna, trots att denna skillnad finns.

Vi använder Sidák proceduren för post-hoc testerna i denna studie. Sidák proceduren används när flera statistiska tester genomförs samtidigt och kontrollerar för den ökade risken för typ-1 fel. Den sätter signifikansnivån för varje enskild jämförelse ( $P_{sig}$ ) enligt följande:

$$P_{sig} = 1 - 1(1 - \alpha)^{1/n}.$$

Där  $n$  är antalet jämförelser. Exempelvis, med en signifikansnivå på 5 % för alla jämförelser tillsammans ( $\alpha = 0,05$ ) och med 10 jämförelser ( $n = 10$ ) blir den justerade signifikansnivån för varje enskild jämförelse cirka 0,5116 %. Fördelen med Sidák proceduren är att den inte höjer risken för typ-2 fel lika mycket som många andra korrigerande procedurer (Field, 2018, s.587). Denna fördel är även anledningen till att proceduren används av oss framför andra alternativ. Givet att när vi jämför läsbarheten av årsredovisningar med den standardiserade noten *Redovisningsprinciper*, finns det redan en ökad risk för typ-2 fel (se Avsnitt 3.1).

### 3.3.3 Regressionsanalys (Hypotes 2)

För att sedan testa **Hypotes 2**, att det finns ett samband mellan språkens släktskap till det engelska språket och engelska årsredovisningars läsbarhet, används följande linjära regression:

$$READABILITY = \beta_0 + \beta_1 DISTANCE + \beta_2 SIZE + \beta_3 MTB + \beta_4 ROA \\ + \beta_5 LEVERAGE + \beta_6 YEAR15 + \beta_7 YEAR16 + \varepsilon$$

där den beroende variabeln *READABILITY* är en av de tre läsbarhetsmått som används i studien (se Avsnitt 3.4.1) och *DISTANCE* är den oberoende variabeln som representerar språkens avstånd från engelska (se Avsnitt 3.4.2). Resterande variabler är kontrollvariabler som baseras på tidigare litteratur (se Avsnitt 3.4.3).

### 3.3.4 Heteroskedasticitet och robust regression (Hypotes 2)

Vid testet av **Hypotes 2** resulterar datat i residualer vars varians bryter mot regressionsanalysens antagande om homoskedasticitet. Av denna anledning använder vi

robusta regressioner. Robust regression är en form av regressionsanalys som är utformad specifikt för att påverkas mindre av att datat bryter mot den klassiska regressionsanalysens antaganden. En fördel av robust regressionsanalys framför den klassiska vid problem med heteroskedasticitet är att den minskar risken för typ-1 fel. Ett resultat av detta är att robusta regressioner rapporterar standardavvikelser och p-värden vilka skiljer sig från de som rapporteras av vanliga regressioner (Field, 2018, s.425). Standardavvikelsena kallas i detta fall för *robusta standardavvikelser*.<sup>10</sup>

### 3.3.5 Signifikansnivå

Likt många tidigare studier (t.ex. Miller, 2010; Laksmana et al, 2012) utgår vi från en lägsta signifikansnivå på 5 %. En signifikansnivå på 5 % innebär en motsvarande sannolikhet att noll-hypotesen vid respektive test förkastas trots att den är sann (Field, 2018, s.86). P-värdena som rapporteras baseras på två-svansade tester, då studiens hypoteser inte är i specifika riktningar.

## 3.4 Variabler

### 3.4.1 Beroende variabler - Läsbarhet

Vi använder tre variabler för att mäta läsbarheten på noten *Redovisningsprinciper*. Den första är **Fog Index (FOG)**. Fog Index, skapad av Robert Gunning, är en av de enklare formlerna för att mäta läsbarhet. Givet att en text är välformad och logisk fångar den textens komplexitet som en funktion av antalet *stavelser per ord* och *ord per mening* (DuBay, 2004). Indexet indikerar antalet utbildningsår i det amerikanska skolsystemet som behövs för att kunna förstå en text vid den första läsningen (Loughran & McDonald, 2016). Exempelvis innebär ett index på 14 att fjorton år av utbildning behövs för att förstå en text. Fog Index beräknas enligt följande funktion:

$$FOG = 0,4 \times (\text{genomsnittligt antal ord per mening} + \text{procent svåra ord})$$

Varav *procent svåra ord* definieras som ord med tre stavelser eller mer. Li (2008) beskriver relationen mellan Fog Index och läsbarhet enligt följande:  $FOG \geq 18$  betyder att texten är oläsbar; 14–18 att den är svårläst; 12–14 anses idealiskt; 10–12 är acceptabelt; och 8–10 beskrivs som barnsligt.

Den andra variabeln som vi använder för att mäta läsbarheten med är **Flesch-Kincaid Index (FLESCH\_K)**. Flesch-Kincaid Index, även kallad Flesch-Kincaid Formula och Flesch Grade Level, togs från början fram av den amerikanska flottan och är baserad på Flesch Reading Ease Score. Likt Fog Index använder Flesch-Kincaid Index endast två element: *genomsnittligt antal ord per mening* och *genomsnittligt antal stavelser per ord*. Måttet är även likt Fog Index i att det indikerar antalet amerikanska skolår som behövs för att kunna

---

<sup>10</sup> För att genomföra de robusta regressionerna i STATA använder vi följande kommando: `regress READABILITY DISTANCE SIZE MTB ROA LEVERAGE i.YEAR, vce(robust)`.

läsa en text (Laksmana et al, 2012). Exempelvis innebär ett index på 8,0 att ett dokument är förståeligt för den genomsnittlige åttondeklassaren. Funktionen för måttet är den följande:

$$FLESCH\_K = (11,8 \times \text{genomsnittligt antal stavelser per ord}) \times (0,39 \times \text{genomsnittligt antal ord per mening}) - 15,59$$

Den slutliga variabeln som vi använder för att mäta läsbarhet med är **Totalt antal ord** (*LENGTH*). Som vi har angett ovan innebär längre texter att informationen i dem blir svårare att bearbeta och ta till sig (Loughran & McDonald, 2014). Därmed är längre texter mindre läsbara. Vi hänvisar läsaren till Avsnitt 2.1.4 för en djupare beskrivning av måttet samt dess styrkor och svagheter. Vi väljer att definiera längden på noten *Redovisningsprinciper* som:

$$LENGTH = \ln(\text{Totalt antal ord})$$

Den naturliga logaritmen av *Totalt antal ord* används, istället för totalt antal ord för sig själv, på grund av att antalet ord är skevt fördelat mellan företagen och för att motverka extrema värden. Just denna version av måttet har använts av flertalet studier för att mäta läsbarheten av hela årsredovisningar och specifika avsnitt i dem (Li, 2008; Miller, 2010; Lundholm et al, 2014; Loughran & McDonald, 2014; Lawrence, 2013).

#### 3.4.1.1 Readable.io och dess validitet

Vi använder molntjänsten Readable.io för att analysera alla de bearbetade textfilerna och beräkna deras Fog Index, Flesch-Kincaid Index och totala antal ord. Detta verktyg har bland annat använts av Cash och Tsai (2018), vilka undersökte relationen mellan läsbarheten av kreditkortskontrakt och de finansiella avgifter som kreditföretagen tar av kunden.

Loughran och McDonald (2016) noterar att mjukvaror som utvärderar texters läsbarhet inte alltid ger helt pålitliga resultat. Läsbarhetsmjukvaror nyttjar algoritmer och ordlistor för att avgöra bland annat hur många stavelser det är i ett ord och hur lång en mening faktiskt är. Dessa ordlistor och algoritmer är dock inte utan brister. Ett exempel som Loughran och McDonald (2016) ger är Fathom-paketet i Perl, vilket är en av de mest vanligt använda mjukvarorna för att mäta läsbarhet. Författarna skapar en lista på över 40 000 ord med ett föridentifierat antal stavelser. De finner att Fathom-paketet endast kan identifiera antalet stavelser tillförlitligt i 75 % av fallen. Av validitetsskäl är det därför viktigt att kontrollera att mjukvaran faktiskt mäter läsbarhetsvariabeln i fråga.

För att granska validiteten av Readable.io:s förmåga att beräkna Fog och Flesch-Kincaid Index jämför vi därför resultaten från tjänsten mot att manuellt beräkna måtten. Vi väljer slumpmässigt ut två paragrafer från tio av våra bearbetade dokument och låter Readable.io granska dem. Samtidigt beräknar vi för hand antalet ord per mening och antalet stavelser per ord i samma texter. Resultaten av detta går att finna i Bilaga 1. Gällande Fog Index är skillnaden mellan de två metoderna lite under fem procent. För Flesch-Kincaid Index är gapet större. Här finner vi en skillnad på just över åtta procent. I båda fallen är det Readable.io som i genomsnitt ger lägre värden på måtten. Detta beror på att tjänsten räknar färre antal ord per mening än vad vi gör manuellt. Readable.io verkar därmed mäta den faktiska läsbarheten av studiens texter, dock med en tendens att till viss del övervärdera deras läsbarhet.

## 3.4.2 Oberoende variabler - Språk

### 3.4.2.1 Oberoende variabel - *SPRÅKGRUPP* (Hypotes 1)

I testet av **Hypotes 1** används en kategorisk variabel för att särskilja de sju grupper vars medelvärden jämförs. Vi utgår ifrån undergrupperna inom den Indo-Europeiska språkfamiljen samt den Finsk-Ugriska språkfamiljen (se Avsnitt 2.2.3). Baserat på detta delas språken in enligt följande: engelska språket (*ENGLISH*), germanska språk (*GERMANIC*), italiska språk (*ITALIC*), baltiska språk (*BALTIC*), slaviska språk (*SLAVIC*), grekiska språket (*GREEK*) och Finsk-Ugriska språk (*FINNO-UGRIC*). Att ett företag tilldelas en viss kategori innebär att företaget har sin hemvist i ett land med ett majoritetsspråk som tillhör denna språkgrupp. Hemvisten baseras på landet där företaget har sitt huvudkontor enligt Compustat.

### 3.4.2.2 Oberoende variabel - *DISTANCE* (Hypotes 2)

För att testa **Hypotes 2** använder vi i regressionen en diskret variabel som mäter hur nära besläktat ett lands majoritetsspråk är till engelskan, *DISTANCE*. *DISTANCE* bygger på de släktskap inom den Indo-Europeiska språkfamiljen som da Silva och Tehrani (2016) presenterar i sin version av det Indo-Europeiska språkträdet (se Avsnitt 2.2.3). Vår variabel tar ett värde mellan 0 och 6, där ett mer avlägset släktskap med det engelska språket innebär ett högre värde på variabeln. Detta innebär att företag från länder där engelska är majoritetsspråket får värdet 0, resterande språk i den västgermanska undergruppen får värdet 1 och språken i den nordgermanska undergruppen får värdet 2. Vidare får språk i den Italiska undergruppen värdet 3, språk tillhörande den balto-slaviska undergruppen får värdet 4 och Grekiska får värdet 5. Slutligen, givet att finska, estniska och ungerska tillhör en egen språkfamilj, ges dessa språk värdet 6. Nedanstående Tabell 2 visar vilket värde ett företag får beroende på i vilket land det har sin hemvist.

**Tabell 2 - Gradering av släktskap**

Länder	Undergrupp / Språk	Tilldelat Värde
Storbritannien, Irland	Engelska	0
Belgien, Luxemburg, Nederländerna, Schweiz, Tyskland, Österrike	Västgermanska språk	1
Danmark, Island, Norge, Sverige	Nordgermanska språk	2
Frankrike, Italien, Portugal, Rumänien, Spanien	Italiska språk	3
Bulgarien, Kroatien, Lettland, Litauen, Makedonien, Polen, Serbien, Slovakien, Slovenien, Tjeckien	Balto-slaviska språk	4
Cypern, Grekland	Grekiska	5
Estland, Finland, Ungern,	Finsk-ugriska språk	6

### 3.4.3 Kontrollvariabler och Kovariater

Det är dock inte bara språket som kan vara avgörande för läsbarheten av årsredovisningar. I tidigare studier finner bland annat Li (2008) och Lundholm et al (2014) ett flertal faktorer som också har betydelse för läsbarheten. Det är därför viktigt att kontrollera för dessa element när vi söker efter interaktioner mellan språk och läsbarheten av årsredovisningar. Av denna anledning använder studien följande kontrollvariabler/kovariater:

- **SIZE: Storlek** fångar många aspekter av ett företags operativa miljö och affärsmiljö. Exempelvis har storlek använts som en proxy för ett företags politiska kostnader (Li, 2008; Laksmana et al, 2012; Moreno & Casasola, 2016). Därför inkluderar vi variabeln *SIZE* och definierar den likt Laksmana et al (2012) som logaritmen av totala tillgångar vid räkenskapsårets slut. I förväg förväntar vi oss att större företag kommer att ha längre och mer komplicerade (svårlästa) årsredovisningar.
- **MTB:** Hur ett företag är prissatt i förhållande till dess bokförda värde (**Market-to-Book**) är enligt Lundholm et al (2014) avgörande för bland annat dess investeringsmöjligheter och tillväxtpotential. Exempelvis kan tillväxtföretag (med högre MTB) ha mer komplicerade och osäkra affärsmodeller, och därför mer svårlästa årsredovisningar. Därför inkluderas variabeln *MTB*. *MTB* definieras som marknadsvärdet av det egna kapitalet plus det bokförda värdet på totala skulder dividerat med bokfört värde av totala tillgångar. Alla tre beståndsdelar är vid räkenskapsårets slut.
- **ROA:** Enligt Li (2008) tenderar företag med lägre lönsamhet att ha årsredovisningar som är mindre läsbara (höga Fog Index). Han resonerar att detta kan bero på att företagsledare opportunistiskt försöker gömma negativ information. Av denna anledning inkluderar vi, likt Lundholm et al (2014), variabeln **räntabilitet på totala tillgångar**. *ROA* beräknas som årets resultat dividerat med årets ingående totala tillgångar.
- **LEVERAGE:** Studier har visat att hur ett företag är finansierat har en inverkan på läsbarheten i många av de upplysningar som lämnas (t.ex. Ertugrul et al, 2017; Bonsall & Miller, 2017). Därför inkluderas en variabel som kontrollerar för förhållandet mellan totala skulder och totala tillgångar. *LEVERAGE* definieras som totala skulder dividerat med totala tillgångar vid räkenskapsårets slut.
- **YEAR:** Slutligen, för att ta hänsyn till skillnader mellan åren, inkluderas två dummyvariabler som designerar redovisningsåret för varje observation. Variablerna *YEAR15* respektive *YEAR16* tar värdet 1 om observationens räkenskapsår slutar 2015 respektive 2016. Tar båda variabler värdet 0 innebär detta att observationen istället tillhör räkenskapsåret 2014.



## 4. Resultat och analys

### 4.1 Beskrivande statistik

I detta avsnitt beskriver vi studiens insamlade data. Nedanför i Tabell 3 går det att se antalet observationer och medelvärden av språk- och läsbarhetsvariabler fördelat på de sju språkgrupperna.<sup>11</sup> Mer detaljerad information för varje enskild språkgrupp går att finna i Bilaga 2. För information per land hänvisar vi läsaren till Bilaga 3. Bland språkgrupperna utgör germanska företag den till antalet största gruppen. Engelska är dock också ett germanskt språk. Inkluderar vi även dessa observationer tillsammans med den germanska gruppen, utgör grupperna tillsammans 56,32% av alla observationer i studien. Därmed är en majoritet av observationerna i vår studie germanska företag. En liknande slutsats går att dra av variabeln *DISTANCE*. Med ett medelvärde på 2,19 och en median på 2 (se Bilaga 2) är det genomsnittliga företaget i vårt urval ett nordgermanskt företag. Detta formar värdet på de summerande statistikor som anger totalen för alla språkgrupper.

**Tabell 3 - Beskrivande statistik**

<b>Språkgrupp</b>	Engelska	Germanska	Italiska	Baltiska	Slaviska	Grekiska	Finsk-ugriska	Totalt
Observationer (n)	92	144	98	12	44	17	12	419
Andel (%)	21,96%	34,37%	23,39%	2,86%	10,50%	4,06%	2,86%	100%
<b>Variabler</b>								
<i>FOG</i>	19,62	18,70	20,38	20,31	20,92	19,20	19,08	19,60
<i>FLESCHE_K</i>	16,27	15,30	17,04	17,12	17,24	15,81	15,62	16,21
<i>LENGTH</i>	8,55	8,71	9,12	8,89	9,18	8,76	8,94	8,83
<i>WORDS</i>	5627	6434	9684	8016	10286	6813	8124	7531
<i>DISTANCE</i>	0	1,69	3	4	4	5	6	2,19

Variablerna är angivna i deras medelvärden.

Medelvärdet på *FOG* (Fog Index) för hela urvalet är 19,60. Den språkgrupp med lägst *FOG* är den germanska (18,70) och den med högst är den slaviska (20,92). Med ett *FOG* på 19,62 ligger de engelsktalande företagen nära medelvärdet för hela urvalet. Sammanfattningsvis anger *FOG* att den genomsnittliga noten *Redovisningsprinciper* skulle vara oläsbar enligt Lis (2008) skala. Medelvärdet på *FLESCHE\_K* (Flesch-Kincaid Index) är 16,21 för hela urvalet. I vår studie anger därmed *FLESCHE\_K* att företagen i genomsnitt skriver tre skolår mer läsbart än vad Fog anger. Även här skriver germanska företag mest läsbart (15,30) och slaviska företag minst (17,24). Liknande *FOG* ligger engelsktalande företag nära medelvärdet för hela urvalet (16,27). Slutligen är den genomsnittliga noten *Redovisningsprinciper* 7531 ord lång. Mest kortfattade är de engelsktalande företagen (5627 ord) och mest omfattande är de slaviska (10 286 ord).

<sup>11</sup> OBS! Värden i detta avsnitt (4.1) och i Bilaga 2–3 ska inte blandas ihop med de justerade medelvärden och standardavvikelser som presenteras i Avsnitt 4.2. Värdena i Avsnitt 4.2 har justerats för kovariaternas påverkan.

Inkluderat i Bilaga 4 finns en korrelationsmatris av alla variabler som finns med i studien. Vid en parvis jämförelse är *FOG* och *FLESCHE\_K* mycket starkt korrelerade (0,974). Det tredje läsbarhetsmåttet, *LENGTH*, är medelstarkt korrelerad med både *FOG* (0,455) och *FLESCHE\_K* (0,485). Samma sak gäller för totalt antal ord jämfört med *FOG* (0,485) och *FLESCHE\_K* (0,451). Likt tidigare studier kan denna korrelation mellan de två traditionella läsbarhetsmått och totalt antal ord indikera att de i viss mån mäter samma sak. Slutligen är den oberoende variabeln *DISTANCE* svagt positivt korrelerad med både *FOG* (0,224), *FLESCHE\_K* (0,202) och *LENGTH* (0,384). Detta är en preliminär indikation på att språkgrupper som är mindre besläktade med engelska skriver mindre läsbart.

## 4.2 Hypotes 1 - Finns det en skillnad på läsbarheten mellan språkgrupper?

I detta avsnitt undersöker vi om det finns någon skillnad på läsbarheten av noten *Redovisningsprinciper* mellan språkgrupper. Tabell 4 nedan visar resultatet av de tre kovariansanalyserna med *FOG*, *FLESCHE\_K* respektive *LENGTH* som läsbarhetsmått och *SIZE*, *LEVERAGE*, *MTB*, *ROA* och *YEAR15-16* som kovariater. I alla tre fallen är modellen statistiskt signifikant på 1-procentsnivån. Därmed förutsäger vår ANCOVA-modell läsbarheten på noten *Redovisningsprinciper* bättre än det naiva antagandet att nästa observerade årsredovisning kommer att vara lika läsbart som genomsnittet av alla tidigare årsredovisningar. Vidare visar  $R^2$  att modellen i de tre fallen förklarar 36–40 % av variansen på läsbarhetsmått.

De tre analyserna fortsätter även att likna varandra när det kommer till variablerna. Den oberoende variabeln *SPRÅKGRUPP* är statistiskt signifikant på 1-procentsnivån för samtliga läsbarhetsmått. Detta innebär att språkgruppen som ett företags hemland tillhör har betydelse för läsbarheten på noten *Redovisningsprinciper*. Alternativt formulerat finns det en skillnad på läsbarheten av noten *Redovisningsprinciper* mellan olika språkgrupper. Resultatet stödjer därför **Hypotes 1**. Det finns en skillnad på läsbarheten av den engelska årsredovisningen mellan olika språkgrupper. Frågan återstår dock mellan *vilka* språkgrupper som skillnaderna på läsbarhet finns.

**Tabell 4 - ANCOVA**

	Läsbarhet som <i>FOG</i>		Läsbarhet som <i>FLESCH_K</i>		Läsbarhet som <i>LENGTH</i>	
	(df; df_r)	P-värde	(df; df_r)	P-värde	(df; df_r)	P-värde
F	(12; 406)	18,98	(12; 406)	22,36	(12; 406)	21,77
Modellens p-värde		0,000 **		0,000 **		0,000 **
R <sup>2</sup>		0,3594		0,3979		0,3915
Justerat R <sup>2</sup>		0,3405		0,3801		0,3735
Observationer (n)		419		419		419
<b>Variabler</b>	<b>F</b> <b>(df; df_r)</b>	<b>P-värde</b>	<b>F</b> <b>(df; df_r)</b>	<b>P-värde</b>	<b>F</b> <b>(df; df_r)</b>	<b>P-värde</b>
<i>SPRÅKGRUPP</i>	27,34 (6; 406)	0,000 **	31,09 (1; 406)	0,000 **	21,47 (1; 406)	0,000 **
<i>SIZE</i>	10,26 (1; 406)	0,002 **	13,45 (1; 406)	0,000 **	49,48 (1; 406)	0,000 **
<i>MTB</i>	1,6 (1; 406)	0,207	0,91 (1; 406)	0,339	0,31 (1; 406)	0,581
<i>ROA</i>	1,5 (1; 406)	0,221	1,97 (1; 406)	0,161	2,45 (1; 406)	0,119
<i>LEVERAGE</i>	2,96 (1; 406)	0,086	5,25 (1; 406)	0,022 *	9,74 (1; 406)	0,002 **
<i>YEAR 15-16</i>	0,24 (2; 406)	0,789	0,36 (2; 406)	0,696	1,83 (2; 406)	0,162

\*, \*\* Signifikant på 5-procentsnivån respektive 1-procentsnivån.

Talen i parenteserna anger antalet frihetsgrader som nyttjas och antalet residuella frihetsgrader.

Variabler: SPRÅKGRUPP är en kategorisk variabel som anger vilken språkgrupp ett företag tillhör.

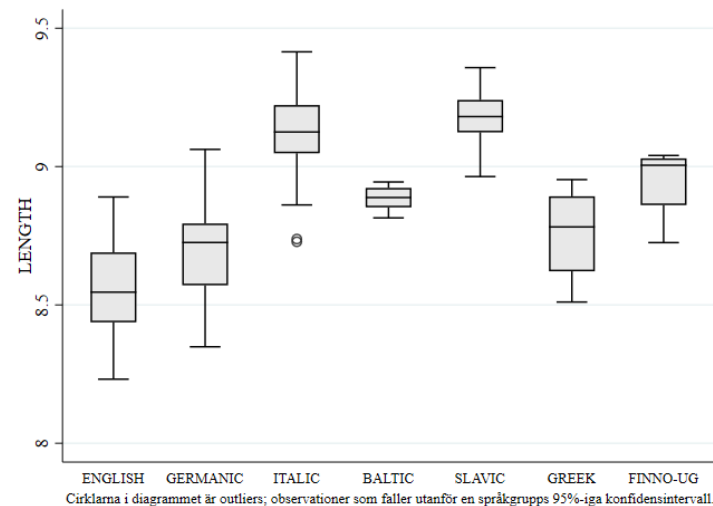
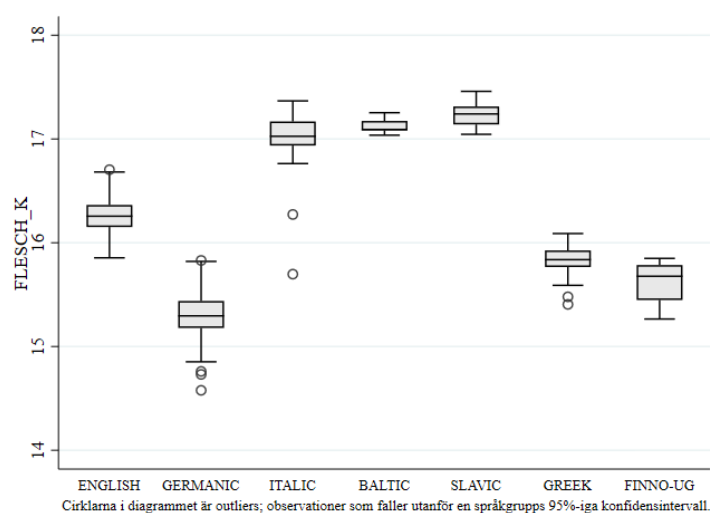
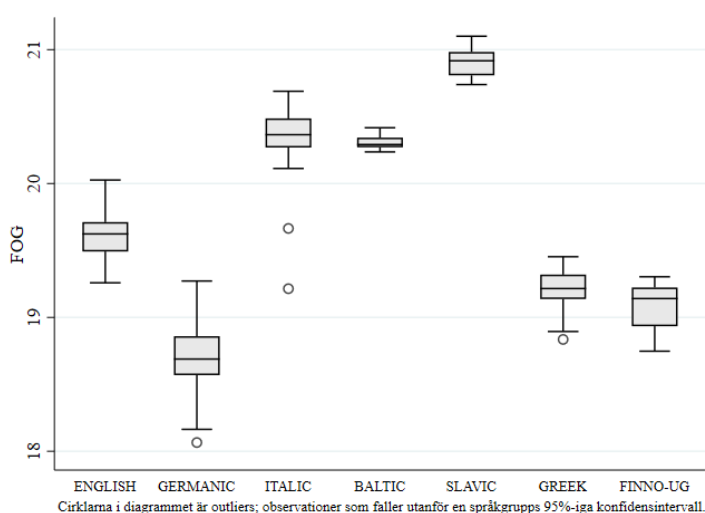
SIZE är den naturliga logaritmen av totala tillgångar. MTB är skillnaden mellan företagets marknadsvärde och dess bokförda värde. ROA är räntabilitet på totala tillgångar. LEVERAGE är totala skulder delat på totala tillgångar. YEAR är dummyvariabler som identifierar observationernas räkenskapsår.

#### 4.2.1 Post-hoc tester

I detta avsnitt undersöker vi mellan vilka språkgrupper som skillnaderna i läsbarhet finns. Tabell 5 visar medelvärden, standardavvikelser och konfidensintervallen av läsbarhetsmått för studiens sju språkgrupper, efter att måtten har justerats för kovariaternas påverkan. Det är dessa värden som jämförs i post-hoc testerna. Vidare är Figur 3a till 3c låddiagram som visar distributionen av läsbarhetsmått per språkgrupp. Även dessa är efter justering för kovariaternas påverkan.

**Tabell 5 - Justerade medelvärden**

Språkgrupp	Läsbarhet som FOG				Läsbarhet som FLESCH_K				Läsbarhet som LENGTH			
	Medel	Std.av.	95% konf.int.		Medel	Std.av.	95% konf.int.		Medel	Std.av.	95% konf.int.	
ENGLISH	19,63	0,117	19,40	19,86	16,27	0,106	16,07	16,48	8,57	0,038	8,50	8,65
GERMANIC	18,73	0,099	18,53	18,92	15,34	0,090	15,16	15,52	8,73	0,032	8,67	8,80
ITALIC	20,31	0,118	20,08	20,55	16,98	0,107	16,77	17,19	9,03	0,038	8,95	9,10
BALTIC	20,41	0,324	19,78	21,05	17,21	0,293	16,64	17,79	9,01	0,104	8,81	9,22
SLAVIC	20,88	0,172	20,54	21,22	17,19	0,156	16,89	17,50	9,20	0,055	9,09	9,31
GREEK	19,31	0,271	18,78	19,85	15,93	0,245	15,45	16,41	8,80	0,087	8,62	8,97
FINNO-UGRIC	19,03	0,322	18,40	19,66	15,56	0,291	14,99	16,13	8,93	0,103	8,72	9,13



I låddiagrammen visar strecken i varje låda medianen på läsbarhetsmättet för en språkgrupp. Lådans ramar visar i sin tur kvartil två och tre, inom vilka 50 % av alla observationer faller. Slutligen visar svansarna det 95-procentiga konfidensintervallet. På detta sätt visualiserar diagrammen skillnader på läsbarheten mellan språkgrupperna.

Tabellerna 6a-c visar resultatet av post-hoc testerna där språkgrupperna parvis har jämförts. Dessa matriser visar för respektive läsbarhetsmått skillnaden på medelvärdet mellan individuella språkgrupper och om denna skillnad är statistiskt signifikant.

Både med *FOG* och *FLESCH\_K* mätt hamnar den engelsktalande gruppen (*ENGLISH*) i mitten av alla språkgrupper. Enligt de två måtten skiljer sig *ENGLISH* signifikant från tre språkgrupper: *GERMANIC*, vilken skriver nästan ett poäng (skolår) mer läsbart än *ENGLISH*; och *ITALIC* samt *SLAVIC*, vilka skriver 0,5–1 poäng (skolår) mindre läsbart än *ENGLISH*. Bland de språkgrupper som vid jämförelserna inte skiljer sig signifikant från *ENGLISH* är *FOG* och *FLESCH\_K* fortfarande högre (mindre läsbart) för *BALTIC* samt lägre (mer läsbart) för *GREEK* och *FINNO-UGRIC*. När läsbarhet istället mäts med *LENGTH* skriver *ENGLISH* mest kortfattat och läsbart av alla språkgrupper. Här skiljer sig *ENGLISH* signifikant från alla språkgrupper utom *GREEK*.

Totalt av 21 parvisa jämförelser finner vi 10 signifikanta skillnader mellan språkgrupperna mätt med *FOG*, 12 mätt med *FLESCH\_K* och 8 mätt med *LENGTH*. Att de parvisa jämförelserna resulterar i signifikanta skillnader i endast 38–57 %<sup>12</sup> av fallen innebär att resultaten från de tidigare kovariansanalyserna (Avsnitt 4.2) bör tolkas med viss försiktighet. Språkgruppen som ett företags hemland tillhör har betydelse för läsbarheten. Dock indikerar post-hoc testerna att det i runt hälften av fallen inte går att säga om företag tillhörande en viss språkgrupp skriver mer eller mindre läsbart än företag från en annan språkgrupp.

**Tabell 6a - Jämförelse av FOG mellan språkgrupper**

Rad Medel - Kolumn Medel	ENGLISH	GERMANIC	ITALIC	BALTIC	SLAVIC	GREEK
GERMANIC	-0,905 ** (0,000)					
ITALIC	0,681 ** (0,001)	1,586 ** (0,000)				
BALTIC	0,779 (0,392)	1,684 ** (0,000)	0,098 (1,000)			
SLAVIC	1,249 ** (0,000)	2,154 ** (0,000)	0,569 (0,127)	0,471 (0,990)		
GREEK	-0,321 (0,999)	0,584 (0,608)	-1,002 * (0,017)	-1,099 (0,176)	-1,570 ** (0,000)	
FINNO-UG	-0,602 (0,823)	0,303 (1,000)	-1,283 ** (0,004)	-1,381 (0,053)	-1,851 ** (0,000)	-0,281 (1,000)

\*, \*\* Signifikant på 5-procentsnivån respektive 1-procentsnivån.

Ovan är skillnaden på medelvärdet av *FOG* mellan språkgrupperna efter att det har justerats för kovariaternas påverkan. I parenteserna är p-värden efter korrigering med Sidák-proceduren. Sidák-proceduren förklaras i avsnitt 3.3.2.

<sup>12</sup> 10/21=47,62 %; 12/21=57,14 %; 8/21=38,10 %

**Tabell 6b - Jämförelse av Flesch\_K mellan språkgrupper**

Rad Medel - Kolumn Medel	ENGLISH	GERMANIC	ITALIC	BALTIC	SLAVIC	GREEK
GERMANIC	-0,935 ** (0,000)					
ITALIC	0,702 ** (0,000)	1,638 ** (0,000)				
BALTIC	0,937 (0,054)	1,872 ** (0,000)	0,234 (1,000)			
SLAVIC	0,918 ** (0,000)	1,853 ** (0,000)	0,215 (0,998)	-0,019 (1,000)		
GREEK	-0,342 (0,991)	0,594 (0,393)	-1,044 ** (0,002)	-1,278 * (0,017)	-1,259 ** (0,000)	
FINNO-UG	-0,716 (0,363)	0,220 (1,000)	-1,418 ** (0,000)	-1,652 ** (0,002)	-1,633 ** (0,000)	-0,374 (1,000)

\*, \*\* Signifikant på 5-procentsnivån respektive 1-procentsnivån.

Ovan är skillnaden på medelvärdet av FLESCH\_K mellan språkgrupperna efter att det har justerats för kovariaternas påverkan. I parenteserna är p-värden efter korrigerig med Sidák-proceduren. Sidák-proceduren förklaras i avsnitt 3.3.2.

**Tabell 6c - Jämförelse av LENGTH mellan språkgrupper**

Rad Medel - Kolumn Medel	ENGLISH	GERMANIC	ITALIC	BALTIC	SLAVIC	GREEK
GERMANIC	0,161 * (0,031)					
ITALIC	0,454 ** (0,000)	0,293 ** (0,000)				
BALTIC	0,439 ** (0,002)	0,278 (0,211)	-0,015 (1,000)			
SLAVIC	0,629 ** (0,000)	0,468 ** (0,000)	0,175 (0,174)	0,190 (0,902)		
GREEK	0,223 (0,337)	0,062 (1,000)	-0,231 (0,282)	-0,216 (0,913)	-0,406 ** (0,002)	
FINNO-UG	0,353 * (0,029)	0,193 (0,818)	-0,100 (1,000)	-0,085 (1,000)	-0,275 (0,321)	0,131 (1,000)

\*, \*\* Signifikant på 5-procentsnivån respektive 1-procentsnivån.

Ovan är skillnaden på medelvärdet av LENGTH mellan språkgrupperna efter att det har justerats för kovariaternas påverkan. I parenteserna är p-värden efter korrigerig med Sidák-proceduren. Sidák-proceduren förklaras i avsnitt 3.3.2.

### 4.3 Hypotes 2 - Finns det ett samband mellan släktskap till engelska och läsbarhet?

I detta avsnitt undersöker vi om det finns ett samband mellan hur nära besläktade språkgrupper är till engelska (*DISTANCE*) och hur läsbart företagen skriver (*READABILITY*). Tabell 7 visar resultaten av de tre robusta regressionerna med läsbarhetsmåten *FOG*, *FLESCH\_K* respektive *LENGTH* som beroende variabler. Det är dessa resultat som diskuteras i detta avsnitt, men resultatet av motsvarande icke-robusta regressioner går att finna i Bilaga 5. Alla tre modeller statistiskt signifikanta på 1-procentsnivån. När läsbarheten mäts med *LENGTH* förklarar modellen enligt  $R^2$  33 % av variansen på läsbarhetsmättet. När

den istället mäts med *FOG* och *FLESCHE\_K* förklarar modellen endast 14–15% av variansen på läsbarhetsmått. De senare är låga, men i linje med tidigare studier (t.ex. Li, 2008; Lundholm et al, 2014). Möjliga förklaringar till varför  $R^2$  är så lågt diskuteras i Avsnitt 5.

Koefficienten för *DISTANCE* är signifikant positiv på 1-procentsnivån för alla tre läsbarhetsmått. Kom ihåg att *DISTANCE* stiger ju mer avlägset besläktat ett språk är till engelskan enligt trädmodellen (se Avsnitt 2.2.3 och 3.4.2). Därför innebär de positiva koefficienterna att ett avsteg från det engelska språket (+1 på *DISTANCE*) leder till att läsbarheten minskar när *FOG* och *FLESCHE\_K* ökar med 0,17–0,14 poäng. Då *LENGTH* är logaritmerad innebär dess koefficient istället att ett avsteg från det engelska språket leder till en ökning på 10 % av det totala antalet ord som används. Därmed skriver företag från länder med språk som är mer avlägset besläktade med engelskan mindre läsbart. Dessa företag använder både längre meningar och ord med fler stavelser (*FOG* och *FLESCHE\_K*), samt nyttjar totalt fler ord (*LENGTH*) för att skriva noten *Redovisningsprinciper*. Resultatet stödjer därför **Hypotes 2**. Det finns ett samband mellan språkens släktskap till engelska och läsbarheten av den engelska årsredovisningen.

**Tabell 7 - Robusta regressioner**

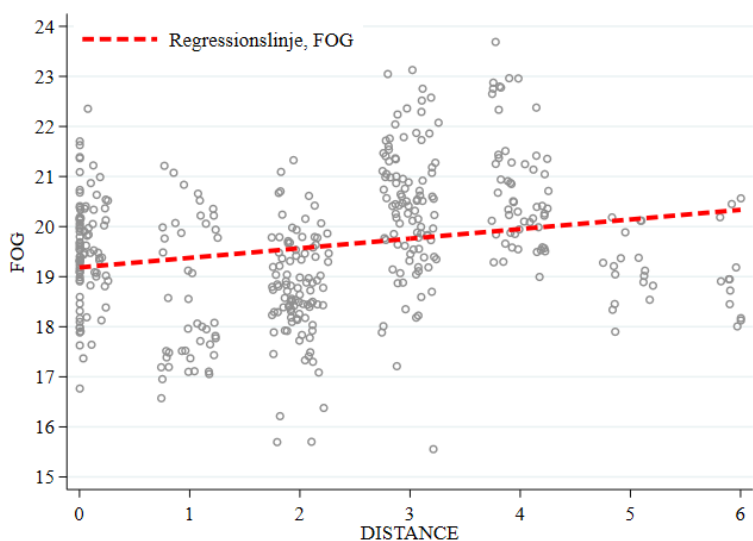
Variabler	Läsbarhet som <i>FOG</i>		Läsbarhet som <i>FLESCHE_K</i>		Läsbarhet som <i>LENGTH</i>	
	Koeff. (P-värde)	Robust Std.av.	Koeff. (P-värde)	Robust Std.av.	Koeff. (P-värde)	Robust Std.av.
<i>DISTANCE</i>	0,170 ** (0,000)	0,040	0,140 ** (0,000)	0,036	0,102 ** (0,000)	0,012
<i>SIZE</i>	0,165 ** (0,000)	0,034	0,168 ** (0,000)	0,032	0,082 ** (0,000)	0,009
<i>MTB</i>	-0,059 ** (0,001)	0,017	-0,070 ** (0,000)	0,017	-0,002 (0,784)	0,008
<i>ROA</i>	-0,049 (0,931)	0,560	-0,085 (0,874)	0,535	-0,232 (0,222)	0,189
<i>LEVERAGE</i>	-0,823 ** (0,005)	0,290	-0,837 ** (0,004)	0,291	0,177 ** (0,007)	0,065
<i>YEAR15</i>	-0,043 (0,781)	0,153	-0,064 (0,653)	0,142	-0,052 (0,258)	0,046
<i>YEAR16</i>	0,042 (0,781)	0,152	0,014 (0,918)	0,140	0,030 (0,467)	0,041
Konstant	18,534 ** (0,000)	0,294	15,219 ** (0,000)	0,271	7,890 ** (0,000)	0,087
Observationer (n)	419		419		419	
P-värde	0,000		0,000		0,000	
$R^2$	0,1389		0,1511		0,3250	
Justerat $R^2$	0,1242		0,1366		0,3135	

\*, \*\* Signifikant på 5-procentsnivån respektive 1-procentsnivån.

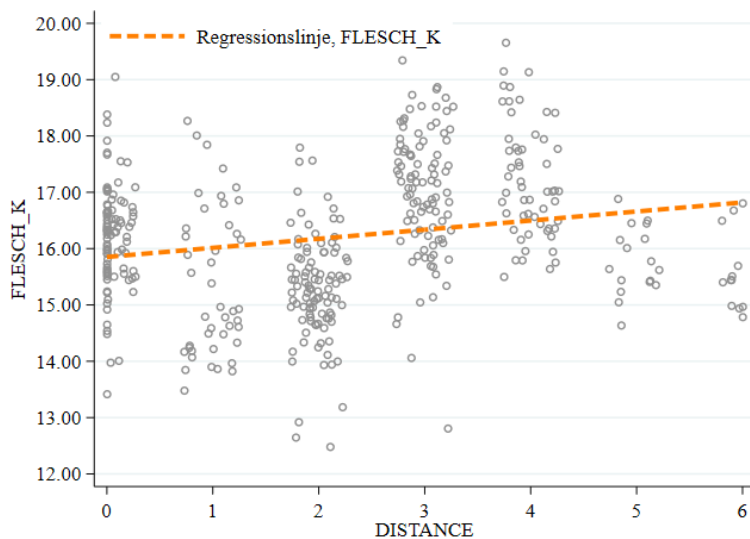
Variabler: *DISTANCE* är en variabel som tar värdet 0 till 6 beroende på släktskapet till engelska. *SIZE* är den naturliga logaritmen av totala tillgångar. *MTB* är skillnaden mellan företagets marknadsvärde och dess bokförda värde. *ROA* är räntabilitet på totala tillgångar. *LEVERAGE* är totala skulder delat på totala tillgångar. *YEAR* är dummy-variabler som identifierar observationernas räkenskapsår.

Vi resonerar att det kan finnas två möjliga förklaringar till detta resultat. En förklaring kan vara att de större skillnaderna mellan företagets egna språk och engelskan gör det svårare för företaget att uttrycka sig skriftligen. Företaget använder därför längre meningar och totalt fler ord för att uttrycka samma information som engelsktalande företag kan uttrycka mer kortfattat. En annan förklaring är att sättet som företaget skriver i sitt egna språk påverkar hur de skriver på engelska. Ju mer avlägset språken är besläktade med engelska desto större blir skillnaderna. Detta resulterar i en engelsk text som är mindre läsbar.

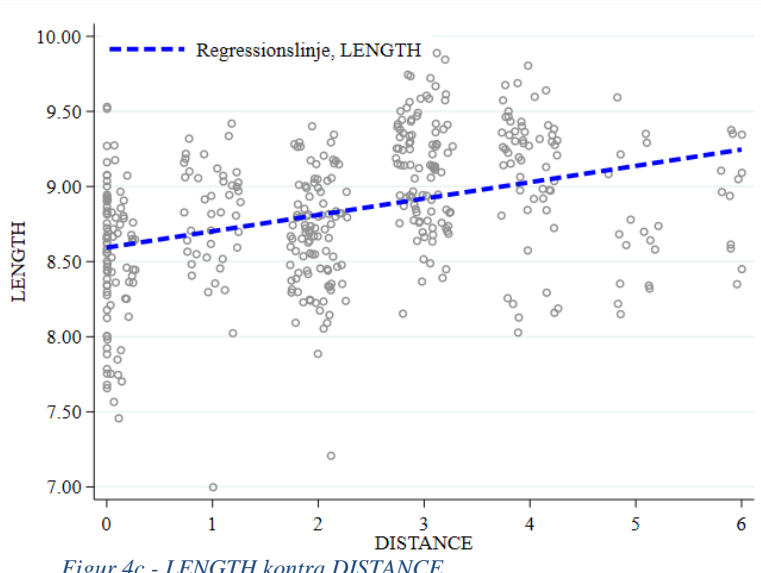
Nedan presenterar Figur 4a till 4c modellens regressionslinjer för respektive läsbarhetsmått över scatterplots av alla observationer. Utifrån dessa figurer kan det ifrågasättas om sambandet *DISTANCE-READABILITY* faktiskt är linjärt. En djupare diskussion kring detta förs i Avsnitt 5.



Figur 4a - FOG kontra DISTANCE



Figur 4b - FLESCH\_K kontra DISTANCE



Figur 4c - LENGTH kontra DISTANCE



### 4.3.1 Kontrollvariablerna

Koefficienten för kontrollvariabeln *SIZE* är positiv och signifikant på 1-procentsnivån för alla läsbarhetsmått. Detta är i linje med våra förväntningar om att större företag kommer att ha längre och mer komplicerade (svårlästa) årsredovisningar. Även Koefficienten för *MTB* är signifikant på 1-procentsnivån när läsbarheten mäts med *FOG* och *FLESCH\_K*. Att *MTB* är negativt innebär att företag som är högre värderade på aktiemarknaden i förhållande till bokfört värde skriver mer läsbart. Detta kan tolkas som att våra resultat motsäger Lundholm et al (2014) om att tillväxtföretag med högre *MTB* även har mindre läsbara årsredovisningar. Vår studie består dock av noterade energiföretag. Tillväxtföretag utgör därför troligtvis en mycket liten del av detta urval. Därmed behöver inte våra resultat gå emot Lundholm et al (2014). Sambandet mellan läsbarhet och *MTB* är dock mer komplicerat än vad de föreslår. Slutligen är koefficienten för *LEVERAGE* signifikant på 1-procentsnivån över alla läsbarhetsmått. Koefficienten är dock negativ för *FOG* och *FLESCH\_K* och positiv för *LENGTH*. Att *FOG* och *FLESCH\_K* är negativa samstämmer med Bonsall och Miller (2017), och kan förklaras med att mer skuldsatta företag har ett större incitament att hålla nere lånekostnaderna. Därför skriver de mer läsbart. Att *LENGTH* är positivt stämmer i sin tur överens med Ertugrul et al:s (2017) upptäckt att företag med strängare låneavtal måste lämna mer upplysningar och skapar därför längre årsredovisningar. Troligen har företagen med högre belåningsgrad i vårt urval även strängare låneavtal.

## 4.4 Företagsspecifika fasta effekter

Av alla 419 observationer för 2014–2016 finns det 120 företag som förekommer under alla tre åren. Vi resonerar att hur läsbart ett företag skriver en årsredovisning kan påverkas av hur läsbart företaget skrev sin årsredovisning det föregående året. Urvalets observationer är därför inte oberoende mellan åren, trots inkluderandet av dummyvariablerna *YEAR*. På grund av att vi inte inkluderar kontrollvariabler för dessa företagsspecifika effekter kan generaliserbarheten av studiens resultat begränsas. Risker finns därför att resultaten endast kan generaliseras till de företag som ingår i denna studie. För att säkerställa att vi inte drar slutsatser som inte kan generaliseras bortom studiens urval gör vi följande: Vi tar bort alla observationer utom de för 2014 och genomför samma kovariansanalyser och regressioner som i Avsnitt 4.2 och 4.3. Tabeller från detta finns i Bilaga 6. Sammanfattningsvis är resultaten av dessa analyser mycket lika de i avsnitten ovan. De påverkar därför inte studiens slutsatser.

## 5. Diskussion

I detta avsnitt berör vi delar av resultatet från Avsnitt 4 och diskuterar möjliga slutsatser som kan dras från dessa.

Både modellen för kovariansanalyserna och modellen för regressionsanalyserna har låga förklaringsgrader, med  $R^2$  som understiger 50 %.<sup>13</sup> De två modellerna förklarar därför inte ens hälften av all varians på läsbarhetsmåten. Taget tillsammans indikerar detta att det finns faktorer som påverkar läsbarheten av noten *Redovisningsprinciper*, men som inte är inkluderade som variabler i modellerna. Detta är förståeligt. Något så komplext som läsbarheten av texter kan omöjligen förklaras med endast studiens sex oberoende variabler. Ytterligare faktorer, såsom landets utbildningssystem samt lagar och reglering, spelar säkerligen viktiga roller i att förklara årsredovisningars läsbarhet.

En faktor som möjligen saknas i denna studie är kulturell distans mellan de olika språken, det vill säga lånande. Enligt lingvistikern sker lånande när kulturer kommer i kontakt och inför delar av varandras språk i sina egna system. Trädmodellen utgår dock endast från genetiska relationer mellan språk och tar inte hänsyn till detta lånande. När vi genomför regressionerna antar variabeln *DISTANCE*, utifrån trädmodellen, därför endast ett linjärt samband mellan läsbarhet och släktskap till engelskan. I kovariansanalysen görs istället inget antagande om något linjärt samband mellan språkgruppernas släktskap.<sup>14</sup> Att förklaringsgraden ( $R^2$ ) av kovariansanalyserna är högre än för regressionerna med motsvarande läsbarhetsmått kan därför indikera att lånande är avgörande för läsbarheten av noten *Redovisningsprinciper* och att kovariansanalyserna fångar detta lånande. Framtida studier som undersöker läsbarheten av engelska årsredovisningar och andra finansiella upplysningar bör därför lämpligen inkludera både en variabel som mäter släktskap till engelska (*DISTANCE*) och en variabel som mäter lånande.

De låga förklaringsgraderna av regressionerna jämfört med kovariansanalyserna kan även ha förklaringen att sambandet mellan läsbarhet och släktskap till engelska inte är linjärt. Som kan ses i Tabell 7 och Figurerna 4a till 4c är lutningen av modellen svag under alla tre läsbarhetsmått. Samtidigt ser sambandet mellan *DISTANCE* och respektive läsbarhetsmått inte ut att vara linjärt utifrån punktdiagrammen. Detta kan vara ett tecken på att de skillnader som uppstår mellan engelska och andra språk när de blir mer avlägsna från varandra inte endast behöver innebära att årsredovisningar blir mindre läsbara. Som exempel på detta finns Mauranens (1993) och Valero-Garces (1996) bevis att finnar använder färre ord och spanjorer fler ord än engelsmän när de skriver motsvarande texter. Detta resonemang ges även stöd av vår studie, när både de grekisktalande och Finsk-Ugriska språkgrupperna skriver mer läsbart än exempelvis de italiska och slaviska vid parvisa jämförelser (se Tabell 6a-c). Sambandet mellan hur nära besläktat ett språk är till engelska och hur läsbart ett företag skriver kan därför vara mer komplext än vad antagandet om linjäritet fångar.

---

<sup>13</sup> ANCOVA:  $R^2 = 36-40\%$ ; Regression:  $R^2 = 14-33\%$ .

<sup>14</sup> Kom ihåg att en ANCOVA är det samma som en regression med dummies för språkgrupperna.

## 6. Slutsatser

Syftet av denna studie är att undersöka om majoritetsspråket i ett land inom Europa är avgörande för läsbarheten av företagens engelska finansiella rapporter. I linje med detta testade vi två alternativhypoteser med de tre läsbarhetsmåten Fog Index, Flesch-Kincaid Index och totalt antal ord:

**Hypotes 1** var att det finns en skillnad på läsbarheten av den engelska årsredovisningen mellan olika språkgrupper. I samstämmighet med detta finner vi att det finns en skillnad mellan språkgrupper på läsbarheten av noten *Redovisningsprinciper*, sett till samtliga tre läsbarhetsmått. Detta indikerar att majoritetsspråket i ett land är avgörande för hur läsbara företagens finansiella upplysningar på engelska är. Skillnaden på läsbarhet finner vi dock bara mellan vissa språkgrupper. 21 parvisa jämförelser mellan alla sju språkgrupper visar att det endast finns skillnader på läsbarheten vid 10 jämförelser för Fog Index, 12 för Flesch-Kincaid Index, 8 för totalt antal ord. Utifrån vår studie går det därför inte alltid att konstatera att företag tillhörande en viss språkgrupp skriver mer eller mindre läsbart än företag från en annan språkgrupp.

**Hypotes 2** var att det finns ett samband mellan språkens släktskap till engelska och läsbarheten av den engelska årsredovisningen. I överensstämmelse med detta finner vi ett samband för alla tre läsbarhetsmåten. Ju mer avlägset besläktat ett språk är till engelska, desto mindre läsbart skriver företagen noten *Redovisningsprinciper*. De nyttjar längre meningar och ord med mer stavelser, samt använder fler ord i texten totalt. Detta kan förklaras med att företagen får det svårare att uttrycka sig på engelska eller att deras egna språk i en större grad påverkar hur de skriver på engelska. Släktskap till det engelska språket är därför en variabel som framtida läsbarhetsstudier in finans- och redovisningslitteraturen bör ta hänsyn till när deras urval inkluderar icke-engelsktalande företag.

Slutligen finner vi att förklaringsgraden konsekvent är högre hos modellen som undersöker skillnaden på läsbarhet mellan språkgrupper än vad den är hos modellen som undersöker sambandet mellan läsbarhet och släktskap till engelska. Utifrån trädmodellen utgår den senare modellen endast från genetiska släktskap och struntar i lånande. Den tidigare modellen gör istället inget antagande utifrån släktskap och fångar därför även effekterna av lånande. Detta indikerar att lånande av ord och begrepp mellan språk kan vara en avgörande faktor för läsbarheten av årsredovisningar. Därför bör framtida läsbarhetsstudier som inkluderar icke-engelsktalande företag lämpligen även inkludera variabler som mäter lånandet mellan språk.

## 7. Förslag till fortsatt forskning

Denna studie bortsåg från lånande vid utformandet. Vi finner dock indikationer på att lånande kan vara en förklarande faktor i hur läsbart företag från icke-engelsktalande länder skriver. En studie som lyckas kvantifiera lånande mellan språk och undersöka dess påverkan på läsbarhet hade därför varit passande. På detta sätt kan förståelsen av rollen hemlandets språk spelar för läsbarheten av finansiella upplysningar på engelska utökas.

Vidare hade en studie som undersöker skillnader i syntax mellan olika språk varit passande. Läsbarhetsmättet Totalt antal ord utgår endast från att längre text är mindre läsbar. Dessutom baseras Fog Index och Flesch-Kincaid Index på antagandet att texten är logiskt utformad när den graderas. Därmed fångar ingen av de tre måtten skillnader i läsbarhet som beror på skillnader i syntax. En studie i stil med den här som inkluderar syntax hade därför tjänat att bekräfta eller bestrida upptäckterna i denna studie.

## 8. Källförteckning

### Böcker

Anttila, R. (1989). *Historical and comparative linguistics* (2. rev. ed., Current issues in linguistic theory, 6). Amsterdam: John Benjamins Pub.

Beekes, R. (1995). *Comparative Indo-European linguistics: An introduction*. Philadelphia; Amsterdam: Benjamins.

DuBay, W. H. (2004) *The Principles of Readability*. Impact Information, Costa Mesa, CA.

Far. (2017). *IFRS-volymer 2017* (Fjortonde upplagan ed.). Stockholm: Far akademi.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5.th ed.). Los Angeles; London: Sage.

Fortson, B. (2010). *Indo-European language and culture: An Introduction* (2.nd ed., Blackwell textbooks in linguistics, 19). Chichester: Wiley-Blackwell.

Hickey, R. (2010). *The Handbook of Language Contact*. Blackwell Publishing Ltd, United Kingdom.

Hock, H. & Joseph, B. (2009). *Language history, language change, and language relationship: An introduction to historical and comparative linguistics* (2nd rev. ed., Trends in linguistics. Studies and monographs; 218). New York, NY: Mouton de Gruyter.

Klare, G.R. (1963) *The measurement of readability*. Ames, Iowa: Iowa State University Press.

Marton, J. (2013). *Redovisning: Förståelse, teori och principer* (1. uppl. ed.). Lund: Studentlitteratur.

Marton, J., Lundqvist, P. & Pettersson, A.K. (2016). *IFRS - i teori och praktik, 4:e upplagan*. Sanoma Utbildning, Stockholm.

Ruhlen, M. (1994). *On the origin of languages: Studies in linguistic taxonomy*. Stanford, CA: Stanford University Press.

### Vetenskapliga tidskrifter

Bloomfield, R. (2008). Discussion of 'Annual report readability, current earnings, and earnings persistence'. *Journal of Accounting and Economics*, 45(2-3), 248-252.

Bonsall, S. & Miller, B. (2017). The impact of narrative disclosure readability on bond ratings and the cost of debt. *Review of Accounting Studies*, 22(2), 608-643.

- Bonsall, S., Leone, A. & Miller, B. Rennekamp, K. (2017). A plain English measure of financial reporting readability. *Journal of Accounting and Economics*, 63(2-3), 329-357.
- Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S.J., Alekseyenko, A.V., Drummond, A. J., . . . & Atkinson, Q.D. (2012). Mapping the origins and expansion of the Indo-European language family. *Science*, 337(6097), 957-960.
- Brochet, F., Naranjo, P. & Yu, G. (2016). The Capital Market Consequences of Language Barriers in the Conference Calls of Non-U.S. Firms. *The Accounting Review*, 91(4), 1023-1049.
- Campbell, D., Beck, C. & Shrikes, P. (2005). A note on comparative language interrogation for content analysis: The example of English vs. German. *The British Accounting Review*, 37(3), 339-350.
- Cash, A. & Tsai, H. (2018). Readability of the credit card agreements and financial charges. *Finance Research Letters*, 24, 145-150.
- Courtis, J. & Hassan, S. (2002). Reading Ease of Bilingual Annual Reports. *Journal of Business Communication*, 39(4), 394-413.
- Da Silva, S. & Tehrani, J. (2016). Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales. *Royal Society open science*, 3(1), 150645.
- De Franco, G., Hope, O., Vyas, D., & Zhou, Y. (2015). Analyst Report Readability. *Contemporary Accounting Research*, 32(1), 76-104.
- Du Toit, E. (2017). The readability of integrated reports. *Meditari Accountancy Research*, 25(4), 629-653.
- Ertugrul, M., Lei, J., Qiu, J., & Wan, C. (2017). Annual Report Readability, Tone Ambiguity, and the Cost of Borrowing. 52(2), 811-836.
- Frazier, K., Ingram, R. & Tennyson, B. (1984) A Methodology for the Analysis of Narrative Accounting Disclosures, *Journal of Accounting Research* 22, 318–31.
- Haak, W. Lazaridis, I. Patterson, N. Rohland, N. Mallick, S. Llamas, B., . . . & Reich, D. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*, 522(7555), 207-11.
- Hamp, E (2013) The Expansion of the Indo-European Languages: An Indo-Europeanist's Evolving View. *SINO-PLATONIC PAPERS*. 239 (August 2013).
- Hoffer, B. (2002). Language borrowing and language diffusion: An overview. *Intercultural communication studies*, 11(4), 1-37.
- Jeanjean, T. Lesage, C. Stolowy, H. (2010). Why do you speak English (in your annual report)? *The International Journal of Accounting*, 45(2), 200-223.

- Jeanjean, T. Stolowy, H. Erkens, M. & Yohn, T.L. (2014). International evidence on the impact of adopting English as an external reporting language. *Journal of International Business Studies*, 46(2), 180-205.
- Laksmana, I. Tietz, W. & Yang, Y.W. (2012). Compensation discussion and analysis (CD&A): Readability and management obfuscation. *Journal of Accounting and Public Policy*, 31(2), 185-203.
- Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of Accounting and Economics*, 56(1), 130-147.
- Lehavy, R. Li, F. Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *Accounting Review*, 86(3), 1087-1115.
- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2-3), 221-247.
- Lieberman, M. & Asaba, S. (2006). Why do firms imitate each other? *Academy of Management Review*, 31(2), 366-385.
- Lo, k. Ramos, F. & Rogo, R. (2017). Earnings management and annual report readability. *Journal of Accounting and Economics*, 63(1), 1-25.
- Loughran, T. & Mcdonald, B. (2014). Measuring Readability in Financial Disclosures. *Journal of Finance*, 69(4), 1643-1671.
- Loughran, T., & Mcdonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, 54(4), 1187-1230.
- Lundholm, R.J., Rogo, R., & Zhang, J.L. (2014). Restoring the tower of Babel: How foreign firms communicate with U.S. investors. *Accounting Review*, 89(4), 1453-1486
- Mauranen, A. (1993). Contrastive ESP Rhetoric: Metatext in Finnish-English Economics Texts, *English for Specific Purposes*, 12,
- McLaughlin, G.H. (1968). Proposals for British readability measures. *Third international reading symposium eds.* Brown and Downing. London: Cassell.
- Miller, B. (2010). The effects of reporting complexity on small and large investor trading. *The Accounting Review: A Journal of the American Accounting Association*, 85(6), 2107-2143.
- Moreno, A. & Casasola, A. (2016). A Readability Evolution of Narratives in Annual Reports. *Journal of Business and Technical Communication*, 30(2), 202-235.
- Nickerson, C. (2005). English as a lingua franca in international business contexts. *English for Specific Purposes*, 24(4), 367-380.

Rennekamp, K. (2012). Processing Fluency and Investors' Reactions to Disclosure Readability. *Journal of Accounting Research*, 50(5), 1319-1354.

Tan, H., Wang, E. & Zhou, B. (2015). How does readability influence investors' judgments? Consistency of benchmark performance matters. 90(1), 371-393.

Tekfi, C. (1987). Readability formulas: An overview, *Journal of Documentation*, 43 257–269.

Tetlock, P.C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market, *Journal of Finance* 62, 1139–68.

Valero-Garcés, C. (1996). Contrastive ESP rhetoric: Metatext in Spanish-English economics texts. *English for Specific Purposes*, 15(4), 279-294.

## Övriga källor

Deloitte. (u.å.). Disclosure initiative — Overview. Hämtad 2018-03-27, från <https://www.iasplus.com/en/projects/major/disclosure-initiative-overview>

IFRS Foundation. (2017a). Better Communication in Financial Reporting – Making disclosure more meaningful. Hämtad 2018-03-28, från <https://www.ifrs.org/-/media/project/disclosure-initiative/better-communication-making-disclosures-more-meaningful.pdf>

IFRS Foundation. (2017b). Snapshot: Disclosure Initiative - Principles of Disclosure. Hämtad 2018-03-28, från <http://www.ifrs.org/-/media/project/disclosure-initiative/disclosure-initiative-principles-of-disclosure/discussion-paper/education-materials/principles-of-disclosure-snapshot-2017.pdf/>

SEC (1998) Plain English Handbook – How to create clear SEC disclosure documents. The Office of Investor Education and Assistance, Washington DC. Hämtad 2018-03-25, från <https://www.sec.gov/pdf/handbook.pdf>

US Securities and Exchange Commission (SEC). (2016). Plain Writing Initiative. Hämtad 2018-03-28, från <https://www.sec.gov/plainwriting.shtml>



## Bilaga 1 – Validitetstest av Readable.io

### *Beräkning av läsbarhet, slumpmässigt urval*

	Fog Index		Flesch-Kincaid Index		Totalt antal ord	
	Readable.io	Manuellt	Readable.io	Manuellt	Readable.io	Manuellt
<b>Text 1</b>	19,60	19,50	17,10	17,04	173	172
<b>Text 2</b>	17,90	19,10	14,85	16,84	403	403
<b>Text 3</b>	23,60	23,73	19,34	19,43	262	262
<b>Text 4</b>	13,90	14,12	11,08	11,72	268	267
<b>Text 5</b>	15,70	17,21	13,08	15,41	351	343
<b>Text 6</b>	10,90	14,56	10,30	16,53	361	360
<b>Text 7</b>	20,00	21,11	15,75	16,05	212	209
<b>Text 8</b>	16,60	18,09	12,90	14,85	246	238
<b>Text 9</b>	18,20	18,20	14,00	14,00	492	492
<b>Text 10</b>	27,50	27,50	22,53	22,53	240	240
<b>Genomsnitt</b>	<b>18,39</b>	<b>19,31</b>	<b>15,09</b>	<b>16,44</b>	<b>300,8</b>	<b>298,6</b>

	Ord per mening		Stavelser per ord		Fog Index
	Readable.io	Manuellt	Readable.io	Manuellt	
<b>Text 1</b>	24,71	24,57	1,95	1,95	<b>Re.io / Man.</b> -4,77%
<b>Text 2</b>	23,71	28,79	1,80	1,80	<b>Flesch-Kincaid</b>
<b>Text 3</b>	32,75	32,75	1,88	1,89	<b>Re.io / Man.</b> -8,19%
<b>Text 4</b>	13,40	14,83	1,82	1,82	<b>Totalt antal ord</b>
<b>Text 5</b>	16,71	21,44	1,88	1,92	<b>Re.io / Man.</b> 0,74%
<b>Text 6</b>	10,31	25,71	1,85	1,87	<b>Ord per mening</b>
<b>Text 7</b>	23,56	23,22	1,88	1,91	<b>Re.io / Man.</b> -11,21%
<b>Text 8</b>	17,57	19,83	1,83	1,92	<b>Stavelser per ord</b>
<b>Text 9</b>	22,36	22,36	1,77	1,77	<b>Re.io / Man.</b> -1,07%
<b>Text 10</b>	40,00	40,00	1,91	1,91	
<b>Genomsnitt</b>	<b>22,51</b>	<b>25,35</b>	<b>1,86</b>	<b>1,88</b>	

Re.io innebär Readable.io och Man innebär manuellt.

## Bilaga 2 – Beskrivande statistik per språkgrupp

Variabel	Engelsktalande företag					Germanska företag				
	n	Medel	Median	Kvartil 1	Kvartil 3	n	Medel	Median	Kvartil 1	Kvartil 3
Läsbarhetsmått:										
<i>FOG</i>	92	19,62	19,70	18,95	20,30	144	18,70	18,60	17,90	19,50
<i>FLESCH_K</i>	92	16,27	16,37	15,59	16,75	144	15,30	15,19	14,69	15,83
<i>LENGTH</i>	92	8,55	8,58	8,33	8,83	144	8,71	8,73	8,47	8,96
<i>WORDS</i>	92	5627	5328	4145	6816	144	6434	6162	4752	7774
Språkvariabler:										
<i>DISTANCE</i>	92	0	0	0	0	144	1,69	2	1	2
Kontrollvariabler:										
<i>SIZE</i>	92	7,52	7,47	6,39	8,82	144	6,97	7,35	4,83	8,24
<i>MTB</i>	92	0,99	0,94	0,67	1,26	144	2,49	1,16	0,96	2,87
<i>ROA</i>	92	1,03%	3,04%	-2,36%	6,10%	144	-1,24%	2,36%	-2,65%	5,12%
<i>LEVERAGE</i>	92	0,54	0,54	0,29	0,77	144	0,62	0,62	0,47	0,77
<i>YEAR</i>	92	2015	2015	2014	2016	144	2015	2015	2014	2014

Läsbarhetsmått: *FOG* = 0,4 x (genomsnittligt antal ord per mening + procent svåra ord); *FLESCH\_K* = (11,8 x genomsnittligt antal stavelser per ord) x (0,39 x genomsnittligt antal ord per mening) – 15,59; *LENGTH* = ln(*WORDS*); *WORDS* är totalt antal ord.

Språkvariabel: *DISTANCE* beskriver släktskapet till engelska och tar ett värde på 0 till 6.

Kontrollvariabler: *SIZE* är den naturliga logaritmen av totala tillgångar. *MTB* är skillnaden mellan företagets marknadsvärde och dess bokförda värde. *ROA* är räntabilitet på totala tillgångar. *LEVERAGE* är totala skulder delat på totala tillgångar. *YEAR* motsvarar observationernas räkenskapsår.

Variabel	Italiska företag					Baltiska företag				
	n	Medel	Median	Kvartil 1	Kvartil 3	n	Medel	Median	Kvartil 1	Kvartil 3
Läsbarhetsmått:										
<i>FOG</i>	98	20,38	20,30	19,70	21,20	12	20,31	20,40	19,95	20,75
<i>FLESCH_K</i>	98	17,04	17,16	16,40	17,84	12	17,12	17,06	16,74	17,59
<i>LENGTH</i>	98	9,12	9,16	8,86	9,34	12	8,89	8,84	8,58	9,27
<i>WORDS</i>	98	9684	9466	7068	11 388	12	8016	6894	5368	10 682
Språkvariabler:										
<i>DISTANCE</i>	98	3	3	3	3	12	4	4	4	4
Kontrollvariabler:										
<i>SIZE</i>	98	9,05	8,95	7,80	10,34	12	6,42	6,48	6,13	6,69
<i>MTB</i>	98	1,06	1,00	0,91	1,20	12	0,89	0,90	0,70	1,07
<i>ROA</i>	98	4,51%	4,73%	2,60%	6,77%	12	3,86%	4,28%	2,48%	5,05%
<i>LEVERAGE</i>	98	0,66	0,68	0,55	0,75	12	0,45	0,52	0,31	0,57
<i>YEAR</i>	98	2015	2015	2014	2016	12	2015	2015	2014	2016

Läsbarhetsmått: *FOG* = 0,4 x (genomsnittligt antal ord per mening + procent svåra ord); *FLESCH\_K* = (11,8 x genomsnittligt antal stavelser per ord) x (0,39 x genomsnittligt antal ord per mening) – 15,59; *LENGTH* = ln(*WORDS*); *WORDS* är totalt antal ord.

Språkvariabel: *DISTANCE* beskriver släktskapet till engelska och tar ett värde på 0 till 6.

Kontrollvariabler: *SIZE* är den naturliga logaritmen av totala tillgångar. *MTB* är skillnaden mellan företagets marknadsvärde och dess bokförda värde. *ROA* är räntabilitet på totala tillgångar. *LEVERAGE* är totala skulder delat på totala tillgångar. *YEAR* motsvarar observationernas räkenskapsår.

Variabel	Slaviska företag					Grekisktalande företag				
	n	Medel	Median	Kvartil 1	Kvartil 3	n	Medel	Median	Kvartil 1	Kvartil 3
Läsbarhetsmått:										
<i>FOG</i>	44	20,92	20,70	19,95	21,65	17	19,20	19,40	18,80	19,50
<i>FLESCH_K</i>	44	17,24	17,00	16,51	18,12	17	15,81	15,86	15,54	16,11
<i>LENGTH</i>	44	9,18	9,25	9,11	9,43	17	8,76	8,65	8,50	9,12
<i>WORDS</i>	44	10 286	10 391	9061	12 476	17	6813	5732	4924	9171
Språkvariabler:										
<i>DISTANCE</i>	44	4	4	4	4	17	5	5	5	5
Kontrollvariabler:										
<i>SIZE</i>	44	7,99	7,89	7,19	8,82	17	7,15	7,58	5,26	8,88
<i>MTB</i>	44	0,94	0,87	0,73	1,01	17	1,26	0,93	0,79	1,22
<i>ROA</i>	44	5,70%	5,87%	3,76%	7,87%	17	6,08%	6,29%	1,96%	10,93%
<i>LEVERAGE</i>	44	0,45	0,48	0,37	0,55	17	0,64	0,67	0,59	0,72
<i>YEAR</i>	44	2015	2015	2014	2016	17	2015	2015	2014	2016

Läsbarhetsmått: *FOG* = 0,4 x (genomsnittligt antal ord per mening + procent svåra ord); *FLESCH\_K* = (11,8 x genomsnittligt antal stavelser per ord) x (0,39 x genomsnittligt antal ord per mening) – 15,59; *LENGTH* = ln(*WORDS*); *WORDS* är totalt antal ord.

Språkvariabel: *DISTANCE* beskriver släktskapet till engelska och tar ett värde på 0 till 6.

Kontrollvariabler: *SIZE* är den naturliga logaritmen av totalt tillgångar. *MTB* är skillnaden mellan företagets marknadsvärde och dess bokförda värde. *ROA* är räntabilitet på totalt kapital. *LEVERAGE* är totala skulder delat på totala tillgångar. *YEAR* motsvarar observationernas räkenskapsår.

Variabel	Finsk-ugriska företag					Totalt				
	n	Medel	Median	Kvartil 1	Kvartil 3	n	Medel	Median	Kvartil 1	Kvartil 3
Läsbarhetsmått:										
<i>FOG</i>	12	19,08	18,75	18,35	19,55	419	19,60	19,60	18,70	20,40
<i>FLESCH_K</i>	12	15,62	15,27	15,10	15,97	419	16,21	16,26	15,28	17,01
<i>LENGTH</i>	12	8,94	8,97	8,65	9,17	419	8,83	8,84	8,53	9,17
<i>WORDS</i>	12	8124	7876	5737	9595	419	7531	6880	5081	9632
Språkvariabler:										
<i>DISTANCE</i>	12	6	6	6	6	419	2,19	2	1	3
Kontrollvariabler:										
<i>SIZE</i>	12	8,44	9,20	7,07	9,81	419	7,72	7,81	6,48	8,99
<i>MTB</i>	12	1,34	1,31	0,95	1,80	419	1,53	1,01	0,86	1,36
<i>ROA</i>	12	7,31%	6,76%	2,42%	12,10%	419	2,02%	3,82%	0,13%	6,74%
<i>LEVERAGE</i>	12	0,52	0,54	0,49	0,58	419	0,59	0,59	0,43	0,74
<i>YEAR</i>	12	2015	2015	2014	2016	419	2015	2015	2014	2016

Läsbarhetsmått: *FOG* = 0,4 x (genomsnittligt antal ord per mening + procent svåra ord); *FLESCH\_K* = (11,8 x genomsnittligt antal stavelser per ord) x (0,39 x genomsnittligt antal ord per mening) – 15,59; *LENGTH* = ln(*WORDS*); *WORDS* är totalt antal ord.

Språkvariabel: *DISTANCE* beskriver släktskapet till engelska och tar ett värde på 0 till 6.

Kontrollvariabler: *SIZE* är den naturliga logaritmen av totala tillgångar. *MTB* är skillnaden mellan företagets marknadsvärde och dess bokförda värde. *ROA* är räntabilitet på totala tillgångar. *LEVERAGE* är totala skulder delat på totala tillgångar. *YEAR* motsvarar observationernas räkenskapsår.

## Bilaga 3 – Beskrivande statistik per land

### Panel A: Läsbarhetsvariabler

<b>Land</b>	<b>ISO-kod</b>	<b>n</b>	<b>FOG</b>	<b>FLESCH_K</b>	<b>LENGTH</b>	<b>WORDS</b>
Belgien	BEL	5	20,14	16,77	8,70	6118
Cypern	CYP	6	18,67	15,42	8,50	5000
Danmark	DNK	5	18,54	15,42	8,39	4447
Estland	EST	3	18,77	15,32	8,48	4825
Finland	FIN	6	18,50	15,18	8,99	8132
Frankrike	FRA	20	20,93	17,44	9,07	9358
Grekland	GRC	11	19,49	16,03	8,90	7802
Irland	IRL	3	19,80	16,78	8,52	5021
Italien	ITA	35	20,56	17,23	9,20	10 489
Kroatien	HRV	6	21,05	17,89	8,84	8474
Lettland	LVA	3	19,83	16,65	8,26	3894
Litauen	LTU	9	20,47	17,28	9,09	9391
Luxemburg	LUX	3	19,70	16,59	9,13	9244
Nederländerna	NLD	9	20,37	17,00	8,92	7699
Norge	NOR	90	18,76	15,24	8,69	6327
Polen	POL	26	21,03	17,36	9,19	10 351
Portugal	PRT	6	21,25	17,62	9,37	11 868
Rumänien	ROU	14	20,25	16,81	8,83	6885
Slovakien	SVK	3	20,27	16,39	9,30	10 949
Slovenien	SVN	3	19,67	15,85	9,38	11 809
Spanien	ESP	23	19,47	16,41	9,16	9876
Storbritannien	GBR	89	19,61	16,25	8,55	5647
Sverige	SWE	5	18,74	15,43	8,53	5385
Tjeckien	CZE	6	21,22	17,23	9,26	10 722
Tyskland	DEU	19	17,49	14,39	8,73	6862
Ungern	HUN	3	20,57	16,81	9,32	11 404
Österrike	AUT	8	17,90	14,63	8,70	6241
n-viktat genomsnitt	–	15,52	19,60	16,21	8,83	7531

## Panel B: Språk- och Kontrollvariabler

<b>Land</b>	<b>ISO-kod</b>	<b>n</b>	<b><i>DISTANCE</i></b>	<b><i>SIZE</i></b>	<b><i>MTB</i></b>	<b><i>ROA</i></b>	<b><i>LEVERAGE</i></b>
Belgien	BEL	5	1	8,45	1,17	3,96%	0,67
Cypern	CYP	6	5	5,84	1,93	2,90%	0,76
Danmark	DNK	5	2	4,36	2,62	-4,01%	0,84
Estland	EST	3	6	5,34	1,87	12,38%	0,58
Finland	FIN	6	6	9,42	1,31	6,44%	0,48
Frankrike	FRA	20	3	10,21	0,97	1,99%	0,74
Grekland	GRC	11	5	7,86	0,90	7,82%	0,58
Irland	IRL	3	0	4,34	0,62	-2,53%	0,13
Italien	ITA	35	3	8,93	1,04	4,48%	0,67
Kroatien	HRV	6	4	7,07	1,60	3,17%	0,27
Lettland	LVA	3	4	6,64	0,69	6,01%	0,20
Litauen	LTU	9	4	6,34	0,96	3,14%	0,53
Luxemburg	LUX	3	1	8,68	3,69	9,91%	0,32
Nederländerna	NLD	9	1	9,81	0,94	3,93%	0,60
Norge	NOR	90	2	6,34	2,83	-3,08%	0,59
Polen	POL	26	4	8,11	0,79	5,71%	0,47
Portugal	PRT	6	3	8,96	1,14	4,69%	0,63
Rumänien	ROU	14	3	6,49	0,93	6,67%	0,48
Slovakien	SVK	3	4	7,71	0,80	7,04%	0,35
Slovenien	SVN	3	4	7,32	1,03	6,50%	0,64
Spanien	ESP	23	3	9,80	1,22	5,39%	0,69
Storbritannien	GBR	89	0	7,63	1,00	1,15%	0,55
Sverige	SWE	5	2	6,84	6,51	-5,73%	1,07
Tjeckien	CZE	6	4	8,83	0,94	7,13%	0,48
Tyskland	DEU	19	1	8,32	1,14	0,59%	0,69
Ungern	HUN	3	6	9,56	0,88	3,98%	0,54
Österrike	AUT	8	1	7,90	1,42	6,49%	0,50
n-viktat genomsnitt	–	15,52	2,19	7,72	1,53	2,02%	0,59

## Bilaga 4 – Parvis korrelation

### *Korrelationsmatris*

	<i>FOG</i>	<i>FLESCH_K</i>	<i>LENGTH</i>	<i>WORDS</i>	<i>DISTANCE</i>	<i>SIZE</i>	<i>ROA</i>	<i>LEVER</i>	<i>MTB</i>	<i>YEAR</i>
<i>FOG</i>	1,000									
<i>FLESCH_K</i>	0,974	1,000								
<i>LENGTH</i>	0,455	0,415	1,000							
<i>WORDS</i>	0,485	0,451	0,955	1,000						
<i>DISTANCE</i>	0,224	0,202	0,386	0,384	1,000					
<i>SIZE</i>	0,257	0,280	0,430	0,458	0,089	1,000				
<i>ROA</i>	0,102	0,103	0,123	0,122	0,163	0,290	1,000			
<i>LEVERAGE</i>	-0,098	-0,107	0,183	0,150	-0,007	0,231	0,056	1,000		
<i>MTB</i>	-0,142	-0,170	-0,087	-0,107	-0,021	-0,206	-0,082	0,008	1,000	
<i>YEAR</i>	0,016	0,006	0,032	0,031	0,030	-0,004	-0,004	-0,038	0,081	1,000



## Bilaga 5 – Icke-robusta regressioner

### *Icke-robusta regressioner*

Variabler	Läsbarhet som <i>FOG</i>		Läsbarhet som <i>FLESCH_K</i>		Läsbarhet som <i>LENGTH</i>	
	<i>Koeff.</i> (P-värde)	Std.av.	<i>Koeff.</i> (P-värde)	Std.av.	<i>Koeff.</i> (P-värde)	Std.av.
<i>DISTANCE</i>	0,170 ** (0,000)	0,040	0,140 ** (0,000)	0,037	0,102 ** (0,000)	0,012
<i>SIZE</i>	0,165 ** (0,000)	0,032	0,168 ** (0,000)	0,030	0,082 ** (0,000)	0,009
<i>MTB</i>	-0,059 (0,071)	0,032	-0,070 * (0,021)	0,030	-0,002 (0,825)	0,009
<i>ROA</i>	-0,049 (0,935)	0,599	-0,085 (0,879)	0,555	-0,232 (0,186)	0,175
<i>LEVERAGE</i>	-0,823 ** (0,001)	0,251	-0,837 ** (0,000)	0,233	0,177 * (0,016)	0,073
<i>YEAR15</i>	-0,043 (0,780)	0,152	-0,064 (0,650)	0,141	-0,052 (0,242)	0,044
<i>YEAR16</i>	0,042 (0,785)	0,154	0,014 (0,920)	0,143	0,030 (0,502)	0,045
Konstant	18,534 ** (0,000)	0,296	15,219 ** (0,000)	0,274	7,890 ** (0,000)	0,086
Observationer (n)	419		419		419	
P-värde	0,000		0,000		0,000	
R <sup>2</sup>	0,1389		0,1511		0,3250	
Justerat R <sup>2</sup>	0,1242		0,1366		0,3135	

\*, \*\* Signifikant på 5-procentsnivån respektive 1-procentsnivån.

Variabler: *DISTANCE* är en variabel som tar värdet 0 till 6 beroende på släktskapet till engelska. *SIZE* är den naturliga logaritmen av totala tillgångar. *MTB* är skillnaden mellan företagets marknadsvärde och dess bokförda värde. *ROA* är räntabilitet på totala tillgångar. *LEVERAGE* är totala skulder delat på totala tillgångar. *YEAR* är dummy-variabler som identifierar observationernas räkenskapsår.

## Bilaga 6 – H1 och H2, utan 2015-2016

### Tabell 2014 - ANCOVA (Hypotes 1)

	Läsbarhet som <i>FOG</i>		Läsbarhet som <i>FLESCH_K</i>		Läsbarhet som <i>LENGTH</i>	
	(10; 128)	7,85	(10; 128)	9,22	(10; 128)	11,78
Modellens p-värde		0,000 **		0,000 **		0,000 **
R <sup>2</sup>		0,3802		0,4186		0,4792
Justerat R <sup>2</sup>		0,3317		0,3732		0,4385
Observationer (n)		139		139		139
<b>Variabler</b>	<b>F</b> <b>(df; df_r)</b>	<b>P-värde</b>	<b>F</b> <b>(df; df_r)</b>	<b>P-värde</b>	<b>F</b> <b>(df; df_r)</b>	<b>P-värde</b>
<i>SPRÅKGRUPP</i>	9,88 (6; 128)	0,000 *	10,83 (6; 128)	0,000 **	9,39 (6; 128)	0,000 **
<i>SIZE</i>	2,27 (1; 128)	0,134	3,88 (1; 128)	0,051	21,55 (1; 128)	0,000 **
<i>MTB</i>	0,37 (1; 128)	0,544	0,09 (1; 128)	0,771	0,52 (1; 128)	0,474
<i>ROA</i>	0,22 (1; 128)	0,637	0,32 (1; 128)	0,574	0 (1; 128)	0,995
<i>LEVERAGE</i>	0,43 (1; 128)	0,515	1,07 (1; 128)	0,303	7,64 (1; 128)	0,007 **

\*, \*\* Signifikant på 5-procentsnivån respektive 1-procentsnivån.

Talen i parenteserna anger antalet frihetsgrader som nyttjas och antalet residuella frihetsgrader.

Variabler: *SPRÅKGRUPP* är en kategorisk variabel som anger vilken språkgrupp ett företag tillhör. *SIZE* är den naturliga logaritmen av totala tillgångar. *MTB* är skillnaden mellan företagets marknadsvärde och dess bokförda värde. *ROA* är räntabilitet på totala tillgångar. *LEVERAGE* är totala skulder delat på totala tillgångar. *YEAR* utelämnas pga endast observationer från 2014 inkluderas.

### Tabell 2014 - Justerade medelvärden

Språkgrupp	Läsbarhet som <i>FOG</i>				Läsbarhet som <i>FLESCH_K</i>				Läsbarhet som <i>LENGTH</i>			
	Medel	Std.av.	95% konf.int.		Medel	Std.av.	95% konf.int.		Medel	Std.av.	95% konf.int.	
ENGLISH	19,62	0,199	19,23	20,01	16,27	0,180	15,91	16,62	8,60	0,058	8,49	8,72
GERMANIC	18,72	0,168	18,38	19,05	15,36	0,152	15,06	15,66	8,73	0,049	8,64	8,83
ITALIC	20,27	0,212	19,85	20,69	16,94	0,192	16,56	17,32	9,00	0,062	8,88	9,13
BALTIC	20,42	0,562	19,30	21,53	17,37	0,508	16,37	18,38	9,08	0,163	8,76	9,40
SLAVIC	21,08	0,304	20,48	21,68	17,38	0,275	16,83	17,92	9,28	0,088	9,11	9,46
GREEK	19,45	0,500	18,47	20,44	16,13	0,452	15,24	17,03	8,78	0,145	8,49	9,06
FINNO-UGRIC	19,22	0,554	18,12	20,32	15,71	0,501	14,72	16,70	8,91	0,161	8,59	9,23

**Tabell 2014 - Jämförelse av FOG mellan språkgrupper**

<b>Rad Medel - Kolumn Medel</b>	ENGLISH	GERMANIC	ITALIC	BALTIC	SLAVIC	GREEK
GERMANIC	-0,904 ** (0,000)					
ITALIC	0,681 ** (0,001)	1,584 ** (0,000)				
BALTIC	0,780 (0,388)	1,683 ** (0,000)	0,099 (1,000)			
SLAVIC	1,249 ** (0,000)	2,153 ** (0,000)	0,568 (0,127)	0,469 (0,990)		
GREEK	-0,321 (0,999)	0,583 (0,612)	-1,002 (0,016)	-1,101 (0,174)	-1,570 ** (0,000)	
FINNO-UG	-0,602 (0,823)	0,302 (1,000)	-1,283 ** (0,004)	-1,382 (0,053)	-1,851 ** (0,000)	-0,281 (1,000)

\*, \*\* Signifikant på 5-procentsnivån respektive 1-procentsnivån.

Ovan är skillnaden på medelvärdet av FOG mellan språkgrupperna efter att det har justerats för kovariaternas påverkan. I paranteserna är p-värden efter korrigering med Sidák-proceduren. Sidák-proceduren förklaras i avsnitt 3.3.2.

**Tabell 2014 - Jämförelse av Flesch\_K mellan språkgrupper**

<b>Rad Medel - Kolumn Medel</b>	ENGLISH	GERMANIC	ITALIC	BALTIC	SLAVIC	GREEK
GERMANIC	-0,934 ** (0,000)					
ITALIC	0,702 ** (0,000)	1,636 ** (0,000)				
BALTIC	0,938 (0,053)	1,872 ** (0,000)	0,236 (1,000)			
SLAVIC	0,917 ** (0,000)	1,851 ** (0,000)	0,215 (0,998)	-0,021 (1,000)		
GREEK	-0,342 (0,991)	0,592 (0,398)	-1,044 ** (0,002)	-1,280 * (0,017)	-1,259 ** (0,000)	
FINNO-UG	-0,716 (0,363)	0,218 (1,000)	-1,418 ** (0,000)	-1,653 ** (0,002)	-1,633 ** (0,000)	-0,374 (1,000)

\*, \*\* Signifikant på 5-procentsnivån respektive 1-procentsnivån.

Ovan är skillnaden på medelvärdet av FLESCH\_K mellan språkgrupperna efter att det har justerats för kovariaternas påverkan. I paranteserna är p-värden efter korrigering med Sidák-proceduren. Sidák-proceduren förklaras i avsnitt 3.3.2.

## Tabell 2014 - Jämförelse av LENGTH mellan språkgrupper

Rad Medel - Kolumn Medel	ENGLISH	GERMANIC	ITALIC	BALTIC	SLAVIC	GREEK
GERMANIC	0,162 * (0,029)					
ITALIC	0,453 ** (0,000)	0,292 ** (0,000)				
BALTIC	0,440 ** (0,002)	0,278 (0,216)	-0,014 (1,000)			
SLAVIC	0,628 ** (0,000)	0,467 ** (0,000)	0,175 (0,178)	0,189 (0,909)		
GREEK	0,222 (0,344)	0,060 (1,000)	-0,231 (0,285)	-0,217 (0,911)	-0,406 ** (0,002)	
FINNO-UG	0,354 * (0,030)	0,192 (0,827)	-0,100 (1,000)	-0,086 (1,000)	-0,275 (0,327)	0,131 (1,000)

\*, \*\* Signifikant på 5-procentsnivån respektive 1-procentsnivån.

Ovan är skillnaden på medelvärdet av LENGTH mellan språkgrupperna efter att det har justerats för kovariaternas påverkan. I paranteserna är p-värden efter korrigering med Sidák-proceduren. Sidák-proceduren förklaras i avsnitt 3.3.2.

## Tabell 2014 - Robusta regressioner (Hypotes 2)

Variabler	Läsbarhet som <i>FOG</i>		Läsbarhet som <i>FLESCH_K</i>		Läsbarhet som <i>LENGTH</i>	
	<i>Koeff.</i> (P-värde)	<i>Robust</i> Std.av.	<i>Koeff.</i> (P-värde)	<i>Robust</i> Std.av.	<i>Koeff.</i> (P-värde)	<i>Robust</i> Std.av.
<i>DISTANCE</i>	0,208 ** (0,003)	0,069	0,174 ** (0,007)	0,063	0,101 ** (0,000)	0,020
<i>SIZE</i>	0,159 ** (0,004)	0,054	0,170 ** (0,001)	0,049	0,085 ** (0,000)	0,015
<i>MTB</i>	-0,09 ** (0,009)	0,033	-0,106 ** (0,001)	0,031	0,000 (0,995)	0,021
<i>ROA</i>	0,080 (0,910)	0,708	0,052 (0,939)	0,678	0,042 (0,797)	0,162
<i>LEVERAGE</i>	-0,846 * (0,018)	0,353	-0,869 ** (0,010)	0,331	0,234 (0,067)	0,127
Konstant	18,547 ** (0,000)	0,496	15,199 ** (0,000)	0,443	7,820 ** (0,000)	0,147
Observationer (n)	139		139		139	
P-värde	0,000		0,000		0,000	
R <sup>2</sup>	0,1532		0,1720		0,3912	
Justerat R <sup>2</sup>	0,1214		0,1409		0,3683	

\*, \*\* Signifikant på 5-procentsnivån respektive 1-procentsnivån.

Variabler: *DISTANCE* är en variabel som tar värdet 0 till 6 beroende på släktskapet till engelska. *SIZE* är den naturliga logaritmen av totala tillgångar. *MTB* är skillnaden mellan företagets marknadsvärde och dess bokförda värde. *ROA* är räntabilitet på totala tillgångar. *LEVERAGE* är totala skulder delat på totala tillgångar. *YEAR* utelämnas pga endast observationer från 2014 inkluderas.