UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Departamento de Lingüística Aplicada
Departament de Lingüística Aplicada

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Dept. of Applied Linguistics

Analysing factors that affect Automated Speech recognition in the context of global English use by Spanish startups.

Master's Thesis

Master's Degree in Languages and Technology

AUTHOR: Purchall, Edward John

Tutor: Candel Mora, Miguel Ángel

ACADEMIC YEAR: 2022/2023

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Departamento de Lingüística Aplicada
Departament de Lingüística Aplicada

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Dept. of Applied Linguistics

Analysing factors that affect Automated Speech Recognition in the context of global English use by Spanish startups.

Master's Thesis

Master's Degree in Languages and Technology

AUTHOR: Purchall, Edward

TUTOR: Candel Mora, Miguel Ángel

ACADEMIC YEAR: 2021/2022

DECLARACIÓN DE AUTORÍA

Declaro que he redactado el Trabajo de Fin de Máster “Analysis of factors that affect Automated speech recognition in the context of Global English use by Spanish startups” para obtener el título de Máster en Lenguas y Tecnología en el curso académico 2022-2023 de forma autónoma, y con la ayuda de las fuentes consultadas y citadas en la bibliografía (libros, artículos, tesis, etc.).

Además, declaro que he indicado claramente la procedencia de todas las partes tomadas de las fuentes mencionadas.

Firmado: Edward Purchall

Acknowledgements:

I would like to express my gratitude to my tutor for his exceptional support and unwavering commitment throughout the academic year. Through his guidance and encouragement, I was able to develop ideas which served as the foundation for this investigation. I am particularly grateful for his diligence and for the feedback I received during the final stages of the project, which played an important role in the completion of this TFM.

ABSTRACT:

Technological advancements and the recent accelerated adoption of video communication platforms have made Automated Speech Recognition software (ASR) omnipresent in digital working environments. Most notably, international corporations use ASR tools to generate real-time transcriptions during online meetings or events, promoting greater accessibility among multilingual participants. As a global language, non-native speakers increasingly use English as a medium for communication, where intelligibility and mutual understanding take preference over standard native norms. Nevertheless, ASR appears to be trained using monolingual models, thus not accurately representing the 'Global English' paradigm. One area where English use is prevalent is in the Spanish startup ecosystem, as presenting in English is viewed as an effective way to raise international visibility and seek foreign investment. However, in multilingual environments, there are disparities in linguistic competence between attendees, and ASR errors in transcription could affect meaning. Since there appear to be no similar studies of 'Global English' use in the Spanish startup sector, this experiment aims to analyse possible factors that affect ASR in the context of a Demo Day pitch. English speech samples of 10 startup founders were recorded, and a series of experiments were conducted to evaluate accuracy.

Keywords:

Automated Speech Recognition, Speech-to-Text, Global English, Spanish Startups, Startup Pitch.

TITULO:

Análisis de los factores que afectan al reconocimiento del habla automático en el contexto de las startups españolas que utilizan el inglés global.

Resumen:

Los avances tecnológicos y la reciente adopción acelerada de plataformas de video comunicación han hecho que el software de reconocimiento automático del habla (ASR) esté omnipresente en los entornos de trabajo digitales. En particular, las empresas internacionales utilizan herramientas de ASR para generar transcripciones en tiempo real durante reuniones o eventos en línea, promoviendo una mayor accesibilidad entre los participantes multilingües. Como lengua global, los hablantes no nativos utilizan cada vez más el inglés como medio de comunicación, donde la inteligibilidad y el entendimiento mutuo tienen preferencia sobre las normas nativas estándar. Sin embargo, la ASR parece entrenarse con modelos monolingües, por lo que no representa con exactitud el paradigma del "inglés global". Uno de los ámbitos en los que prevalece el uso del inglés es el ecosistema de las startups españolas, ya que presentarse en inglés se considera una forma eficaz de aumentar la visibilidad internacional y buscar inversiones extranjeras. Sin embargo, en entornos multilingües, existen disparidades de competencia lingüística entre los asistentes, y los errores de ASR en la transcripción podrían afectar al significado. Dado que no parece haber estudios similares sobre el uso del "inglés global" en el sector de las startups españolas, este experimento pretende analizar los posibles factores que afectan a la ASR en el contexto de una presentación breve durante una jornada de demostración. Se grabaron muestras de habla inglesa de 10 fundadores de startups y se realizaron una serie de experimentos para evaluar la precisión.

Títol:

Anàlisi dels factors que afecten el reconeixement de la parla automátic en el context de les startups espanyoles que utilitzen l'anglès global.

Valencià:

Els avanços tecnològics i els recents posicionaments accelerats de plataformes de comunicació han donat com a resultat un software de Reconeixement Automàtic de Veu, una omnipresent transformació en els entorns de treball digital. Molt més notable, en les corporacions internacionals fent servir ferramentes ASR per a generar en temps real transcripcions durar les reunions on-line o als esdeveniments, promovent un augment de l´accés mitjançant els participants multilingüístics. Com a llengua global, l´anglès està incrementant l´ús per els parlants no-natius com a mitjà per la comunicació a on la intel·ligibilitat pren preferència damunt les normes natives estàndard. No obstant, per una altra banda, aparentment ASR està entrenat per l´ús de models monolingüístics, no oferint una representació realista de el paradigma 'Global English'. Una àrea a on l´anglès es fa servir de manera preferent, es a les startups espanyoles a on es presenten en anglès per a incrementar la visibilitat internacional i poder trobar inversió estrangera. D´altra banda, en aquests contextos multilingüístics hi ha molta disparitat pel que fa en les habilitats lingüístiques, i les errades en la transcripció del reconeixement de veu pot afectar el significat. Com pareix que no hi ha investigacions similars en aquest sector, aquest estudi te com a objecte avaluar els factors que poden afectar a ASR en el context de una presentación en una jornada de demostracions. Es van gravar mostres de parla anglesa de 10 fundadors de startups i es van fer una sèrie d'experiments per avaluar la precisió.

Contents

1. Introduction

The increasing use of digital tools and the importance of English as a global language highlight the need for ASR systems to cater to multilingual users (Seone & Suarez-Gomez, 2016). This entails acknowledging the diversity and fluidity of English usage and the sociolinguistic and sociocultural dimensions of English users in the contemporary globalised world (Rose & Galloway, 2019, as cited in Lee & Jeon, 2023). Given the proliferation of non-native English speaker interactions in the international business arena and the increased use of ASR tools, it is argued that these systems should be attuned to the global context of English, reflecting the reality of its users in an international environment where English serves as the business lingua franca.

Previous studies highlight that ASR systems are trained on monolingual models and do not accurately represent the 'Global English' paradigm. The global nature and multilingual character of the Spanish startup sector highlights the importance for entrepreneurs to communicate in English, where proficiency in language helps to facilitate growth strategies such as scalability and international expansion. While previous studies have focused on ASR bias and English varieties, there appears to be little data regarding ASR efficacy within the international startup community.

Moreover, few studies have focused on the specific needs of Spanish founders presenting in English during a startup pitch. A pitch allows startups to demonstrate credibility through an engaging narrative that confers a sense of legitimacy upon a new venture, reducing doubt about funding it (Lounsbury & Glynn, 2001; Fairbairn et al., 2022). Given the potential to receive investment, these interactions are high stakes; thus, clarity is critical in transmitting confidence. However, in an environment where pitches are increasingly delivered online, it raises questions regarding the extent to which ASR tools accurately express intelligible outputs. As English is becoming more important in the global startup scene and Automated Speech Recognition (ASR) plays a more prominent role in online communication, this study aims to investigate the effectiveness and potential biases of ASR tools in this setting.

Research questions:

This TFM aims to explore how variation in the oral output of non-native speakers of English, specifically Spanish startup founders, impacts automated speech recognition transcription accuracy. Using an industry-leading state-of-the-art ASR tool, the study aims to analyse extemporaneous speech with reference to a Global English context where intelligibility and mutual understanding are prioritised over native language standards. Through a series of experiments, it will address the following questions:

RQ1. To what extent are ASR systems tolerant of variations in the oral output of non-native English speakers, particularly Spanish startup founders?

RQ2. How do the errors in ASR transcription correlate with sociolinguistic factors such as age, gender, and level of English proficiency among Spanish startup founders?

RQ3. Regarding intelligibility, to what extent are variations in output intelligible for human comprehension but not ASR transcription?

RQ4. Are there any common linguistic patterns used by Spanish speakers of English that cause inaccuracies with ASR transcriptions? Should certain elements be given more focus or even avoided?

2. Theoretical Framework

The prevalence of English as a global language is closely linked to the internationalisation of trade and culture, which peaked around the turn of the century (Jenkins, 2015). Today, the ubiquity of English is unmatched, and it is 'unassailable' in its supremacy (Crystal, 1997, pp. 61-62, cited in Tsuda, 2008). House (2002) suggests that the proliferation of English can be attributed to various factors, including the dominance of the British Empire, the emergence of the United States as an economic and political powerhouse, the rapid advancement of information technologies, and the increase in international mergers and acquisitions. In addition, it should be acknowledged that in the last two decades, greater accessibility to communication technology and the creation of social media have only helped to expedite interaction and contact between individuals worldwide. Furthermore, Graddol (2019) comments that it is a remarkable milestone in human history that a single language has emerged as a global lingua franca, enabling communication between people speaking different languages. English has become a language commonly used to express human experiences beyond the confines of its original-speaking populations (Mair, 2003). The term 'Global English' is increasingly used to refer to this phenomenon. According to Rose & Galloway (2019), it is a proposed umbrella term that encompasses the linguistic, sociolinguistic, and sociocultural diversity and adaptability of English usage and users in a globalised world. In other words, it considers English use among individuals who share it as their first language within a particular country, as well as by individuals from different countries and language backgrounds (Jenkins, 2006).

However, the current term 'Global English' represents an evolution of perspectives and models surrounding the changes that the English language has undergone. In the context of globalisation, continuing widespread usage has led to 'uncharted divergence' (Graddol, 2019, para. 10) and new varieties due to greater mobility and contact with diverse speech communities and languages (Canagarajah, 2013). Subsequently, these diversities have raised questions regarding the legitimacy of Standard English (SE) in a globalised non-native setting promoting debate regarding the validity of non-native English norms and assumptions that ownership belongs to native speakers (Matsuda, 2003). Accordingly, throughout the periods of post-colonialism (1945 onwards) and postmodern globalisation (1970-1990), scholars have

attempted to categorise non-native speaker varieties through different models (Canagarajah, 2013). However, the relationship between globalisation and English is complex, and the role that the English language has played throughout history is divisive. For example, descriptions referring to the impact it has on the individuals and communities that use it range from marginalising and hegemonic to empowering and promoting social mobility. As such, the existence of this phenomenon has led to both cooperative relationships and conflicts between global and local influences. The widespread nature has also meant that it has also had significant effects on language, ideology, socio-cultural factors, politics, and education (Sharifian, 2009). Due to these reasons, the proposed labelling of such models to classify English speakers is often contested and not without controversy. Furthermore, there seems to be a lack of consensus regarding concrete definitions of said models, as each has its own history, which leads to scholars viewing these independently. Nevertheless, it is important to discuss concepts such as Standard English (SE), as well as the models World Englishes (WE) and English as a Lingua Franca (ELF), as they help to illustrate developments which precipitated the term 'Global English'.

In the following sections, key aspects will be explored. Firstly, a historical overview of standard English will be discussed. Secondly, a summary of the key emergent models that preceded Global English will be provided. Lastly, the concept of Global English, its characteristics, and the implications that it has had in recent years, predominantly in a pedagogical and business context, will be outlined.

2.1. Standard English

In a historical context, processes such as language standardisation and normalisation have occurred for more than two centuries (Milroy & Milroy, 2012). However, it was in the 18th century that the belief in establishing a "standard" form of English became prominent (Hickey, 2012). This concept involved creating a consistent language style throughout the country. Milroy & Milroy (2012) state that the concept of standardisation aligned with the prevailing notion of authoritarianism and prescription, especially in matters related to linguistics. According to Crystal (2005), this period saw the development of standard English through codification, the process of establishing a publicly recognised and fixed language form

(Trudgill, 1992 cited in Trudgill 1999). Many educators, especially language practitioners, subscribed to these ideas. This resulted in a noticeable surge in the creation of English grammar books and beliefs related to standardised forms (Nevalainen & Van Ostade, 2006). Furthermore, during this period, the concept of a standardised language ideology emerged. Irvine and Gal (2000 p.35) define this as 'the ideas with which participants and observers frame their understanding of linguistic varieties and map those understandings onto people, events and activities that are significant to them'. These ideas are significant as the belief that a standard language represents a country led to the promotion of standard English as the language of power and control, thus surpassing all other variations that can be compared to it (Hickey, 2012).

In recent decades, there has been debate and scrutiny around the representation and ownership of Standard English. Paradoxically, the standardised form of the English language was presented as representing the nation, whereas it only reflected the dialect of a small and privileged group (Hickey, 2012). Scholars argue that national languages often reflect an affluent minority and thus present biased and idealised representations of language (Lippi-Green, 1997). Moreover, the preference by influential organisations and institutions to achieve a standardised and consistent way of speaking can sometimes lead to the imposition of a language. However, some scholars argue that it is for this very reason that a standard form takes prevalence. Trudgill (1999) argues that Standard English is a distinct dialect, perceived as prestigious, lacking a specific accent, and not part of any geographical continuum. Prestige in society refers to those with high social status, material wealth, political power, and educational achievements. This group commonly uses English, which then becomes the community's standard. In turn, educated members of the community advocate and recognise Standard English as the preferred form of education and communication (Trudgill, 1999).

In addition to issues surrounding representation and marginalisation, scholars also address the issue of ownership. According to Widdowson (1994), the presence of a "standard" language within a community can result in the exclusion of those who cannot conform to established language norms. Moreover, those who adhere to the conventional standards of

English can use it to exert power over those who are not members of the same language community (Lowenberg, 2000).

According to Cushing (2023), educational institutions tend to prioritise Standard English and associate it with achievement and accuracy. This leads to the creation of an idealised individual who is expected to conform to a set of linguistic tools and a predetermined identity characterised by "good" or standardised English. The growth of English Language Teaching (ELT) worldwide serves as a good example of an industry that is based on similar principles. Milroy & Milroy (2012) assert that standard English is the variation that is typically taught to non-native learners. Standard English is a strict form of language that is primarily used in writing. It was widely used in the English-speaking world during the twentieth century, with only small differences in spelling, vocabulary, phrases, and grammar between different regions (Fisher, 1996). However, as Trudgill (1999) points out, essentially, it does not allow for any variations.

In the context of globalisation, this last point is particularly pertinent. English is widely used by a wide range of different backgrounds and varying world views, resulting in the development of various forms of the language. This divergence emerges as English speakers adapt the language to their needs within their speech communities. Therefore, it raises questions regarding ownership and the extent to which native speakers have the ultimate authority to decide which language realisations are grammatically correct and which are considered standard. Owing to this, debates surrounding whether to uphold standard English or recognise diverse forms of English as legitimate began to arise (Pennycook, 2008).

2.2. Emergent Models and Varieties

English has become a dominant language, surpassing other language for global and local use. Increasingly, research by scholars such as Kachru (1965) has explored the evolution of the English language from a monolithic entity to a pluralistic one (Smith, 1976, as cited in Sadeghpour & D'Angelo, 2022). As a result, this shift has given rise to new perspectives, including World Englishes (WE) and English as a Lingua Franca (ELF). These models emerged in response to the criticisms of standardised English and due to changes in language owing to

increased global use and challenged ideologies rooted in native speakerism (Holliday, 2006). Essentially, it is argued that the main goal of these models is to demonstrate the validity of English varieties and highlight their unique linguistic features, which reflect the identities and cultures of their speakers. Above all, they have the shared aim of promoting respect for multilingualism. The following section will briefly discuss these models.

2.2.1. World Englishes (WE)

The World Englishes model studies the evolution of the English language through contact with other languages and communities. Languages are not static but an ever-changing and complex network of multiple factors. Kachru's model outlines how English has spread globally, leading to divergences in various regions. The emergence of World Englishes challenges the idea of a "standard" English language. Moreover, several varieties could now be considered the ideal model, as discussed by McKay (2002). Kachru (1985) defines the indigenisation of English as adapting to diverse communities' needs and habits through diffusion (D'Angelo, n.d.). Over time, different regions have developed their own unique versions of English, each with its own rules and customs. The term 'World Englishes' covers the various functional and structural differences, diverse social and linguistic settings, creative expressions, and different cultural influences found in both Western and non-Western parts of the world (Kachru, 1992, p. 2). The division of speakers into native and non-native categories was questioned; however, D'Angelo (2012) notes that emerging forms of language used by non-native speakers are now recognised as distinct varieties rather than being dismissed as interlanguages or learner varieties. Non-native speakers who speak differently from native speakers are not speaking broken or deficient English; instead, their language is simply different (Canagarajah, 2013).

2.2.2. Kachru's Three Centric Circles

According to Kachru's model, English speakers can be categorised into three circles: the Inner Circle, Outer Circle, and Expanding Circle (Kachru, 1985). The Inner Circle contains countries such as the USA, UK, and Canada, where English is the primary and often only language spoken and sets the norms that spread to other communities (Canagarajah, 2013; Mariño, 2011). The Outer Circle includes post-colonial nations such as India and Kenya, where English is not the

primary language but is widely used. In these countries, English is influenced by local languages and customs, leading to its evolution. The Expanding Circle covers countries like China and South Korea, which were not colonised by the British and use English primarily for international communication. They are traditionally seen as being dependent on the Inner Circle norms.

However, some critics argue that the Expanding Circle does not fully conform to the norms of the Inner Circle (House, 2002; Jenkins, 2000). Regarding their connections, multilingual individuals often use practical methods to establish open communication and collaborate effectively. This allows them to efficiently handle and resolve any variations that may arise. In addition, WE does not consider the interaction between different groups of people. It focuses only on the differences within the three particular circles. In this situation, these interlocutors must negotiate and co-construct meaning and local norms in situational contexts. Nevertheless, this solution does not address or solve the potential difficulties that can occur due to interactions across circle boundaries.

2.2.3. English as a Lingua Franca

To understand Global English, it is important to highlight the link with English as a Lingua Franca (D'Angelo, 2016). The main goal was to create a more comprehensive model than the previous one used by WE. This was achieved by focusing on the English usage by Expanding Circle speakers, particularly in mixed or international settings, and also considering the indigenous use of English in the Outer Circle. Additionally, the ELF movement suggests that there are principles that can facilitate effective communication among non-native speakers in global settings. Its goal was to establish a global form of English that is shaped by ELF speakers collectively instead of being dictated by native speakers. (Seidlhofer 2004; Jenkins in 2006).

The key features include focusing on multilingual norms in EFL interactions and aspects related to accommodation. For example, speakers in ELF adjust or even simplify their language use to accommodate their interlocutors' linguistic preferences (Seidlhofer, 2004; Jenkins, 2007). Moreover, during ELF interactions, speakers dynamically mix linguistic norms, features, and practices from their native languages (Cogo, 2012). Essentially, participants use their

multilingual skills to communicate in a flexible form of English that enhances understanding and fosters inclusivity by recognising multilingualism (Mauranen, 2012).

Interestingly, unlike traditional lingua franca, ELF emphasises pragmatic strategies necessary for effective intercultural communication. According to the ELF framework, the ideal English speaker model is not a native speaker but a bilingual speaker who is fluent in the language. This speaker still identifies with their national accent and has the ability to effectively communicate with other non-native speakers (Graddol, 2006). Furthermore, ELF communication is seen as a more fluid and changing phenomenon used in 'communities of practice' (Lave & Wenger, 1991, as cited in Baker, Dewy & Jenkins, 2018).

2.2.4. Lingua Franca Core

Many studies have been conducted to identify and prioritise specific phonological characteristics (Jenner, 1989, as cited in Archer, 2023). New developments, such as the Lingua Franca Core (LFC), emerged through these investigations as a way of assisting non-native speakers in producing clear and acceptable speech. The LFC provides a framework of linguistic features that can promote mutual intelligibility and successful communication in ELF settings (Jenkins, 2000). Jenkin's model is particularly useful as it concentrates on the key factors needed for achieving comprehensibility amongst speakers involved in L2 interactions rather than comprehensibility from the viewpoint of L1 comprehension (Archer, 2023). According to Deterding (2013), the ability to produce key pronunciation features from the LFC and approximate the interlocutor's pronunciation promotes intelligibility in ELF contexts. Within the ELF framework, the goal is not to conform to a native English speaker model but to become a fluent bilingual speaker who maintains their national identity in terms of accent.

Additionally, speakers should possess the necessary skills to effectively communicate and understand other non-native speakers (Graddol, 2006; Cogo & House, 2018). Interestingly, there are a number of studies which appear to support the viewpoint that intelligibility in ELF situations is not necessarily linked to speaking with a native speaker accent (Kiczkowiak, 2019).

The LFC outlines four essential areas to eliminate errors in ELF communication (Walker, 2010).

1. Individual consonant sounds
2. Consonant clusters
3. Vowels
4. Nuclear stress placement.

As outlined, English as a Lingua Franca (ELF) prioritises intelligibility for non-native speakers. Here are the key points regarding its phonetics (Walker, 2010):

Individual Consonant Sounds:

- Non-native speakers often replace unfamiliar English consonants with those from their language, leading to confusion.
- Key English consonants for ELF include /p, t, k, l, and r/.
- Aspiration, especially for /p, t, k/, is essential.
- The sound /t/ is recommended to be pronounced as in British English.
- While clear /l/ pronunciation is considered ideal, dark /l/ substitution occurs in some ELF scenarios. The /r/ sound has variations, but some, like [ʁ], could hinder intelligibility.

Consonant Clusters:

- These are groups of consonants within words that can pose challenges for non-native speakers.
- Strategies to tackle these challenges include adding a short vowel or deleting a consonant. The former is preferred as it maintains intelligibility.

Vowel Sounds:

- Pronunciation varies across English accents.

- The LFC emphasises the length of vowels over exact quality.
- Vowel length is crucial, especially in differentiating words.

Nuclear Stress Placement:

- Nuclear stress refers to the emphasis on a syllable within a word group.
- Correct stress placement is key to conveying accurate information in ELF. Inadequate pausing or misplaced stress can cause misunderstandings.

2.3. Global English Paradigm

David Crystal estimated that the number of people using English was approximately 377 million. During this period, approximately 235 million people spoke English as a second language (Crystal, 2005). Crystal also predicted that by 2050, the number of non-native English speakers will double, surpassing the number of English L1 native speakers, significantly shifting the hegemony of English language use (Crystal, 2005). In fact, according to Lowenberg (2000, p.67), non-native English speakers outnumbered native speakers by the year 2000. He also stated that native speakers only made up a small fraction, about one-fifth or less, of all English users worldwide. Crystal later revised these statistics, acknowledging that approximately 1.5 billion people speak English fluently, with 400 million native speakers. In other words, for every native speaker, there are now three or four non-native speakers, a ratio that will progressively increase with time (Graddol, 1999).

Although estimates vary, as of 2023, there are estimated to be between 1.5 and 2 billion English speakers worldwide (The Economist, 2019; British Council, 2023; Statista, 2023). Some argue that there are around 400 million native English speakers, while over 1 billion speak English as a second language. Additionally, it should be noted that less than half of the total English users reside in core English-speaking or inner-circle countries (The Economist, 2023). The number of non-native English speakers is rising steadily. In the EU, around 40% of the population, or approximately 180 million people, speak English, more than the total population of Britain, Canada, Australia, and New Zealand combined. India has an estimated

60 million to 200 million English speakers, making it the second-largest Anglophone nation in the world (Ibid).

The data seems to support the notion that non-native speakers outnumber native speakers, and non-native to non-native interactions are more common than native to non-native ones (Lowenberg, 2000, p.67). As a global language, English is utilised in diverse fields, including science, business, education, and the internet. As such, it can be suggested that a user is much more likely to use English to communicate professionally with a non-native speaker than native speakers. Clearly, the linguistic landscape is changing, and previous investigations on World Englishes and English as a lingua franca have helped promote an ideological shift that tolerates English varieties and views English as belonging to the world rather than any one country (English Effect, 2013).

Global English recognises the hybrid and changing nature of communication and highlights the pluricentricity of language. As previously mentioned, it is an umbrella term informed by some of the changing viewpoints and emergent paradigms, namely, World Englishes (WE) and English as a Lingua Franca (ELF). While not completely aligned in all aspects, these three terms recognise that English is now the global lingua franca used mainly by 'non-native speakers' of the language (Kiczkowiak, 2020). According to Galloway and Rose (2019), these models share a common ideology. Studies that focus on Global English increasingly point out the diverse nature of English and its evolving sociolinguistic landscape worldwide. To most individuals, English serves as a second language in addition to their mother tongue and, in turn, impacts on English. Moreover, English is no longer limited to native speakers or used only to communicate with them. Nowadays, it is used worldwide and has been adapted by its speakers in various ways (Galloway and Rose, 2019).

2.3.1 Global English Paradigm Impact

Increased global mobility has led to significant changes in the sociolinguistic landscape of English. As a result, there is growing research interest in promoting diversity and plurality of Englishes, especially in pedagogical and business contexts (Solmaz, 2023). The following section provides some examples to illustrate this influence.

2.3.1.1. GELT

Global English Language Teaching (GELT) focuses on teaching English as a common language, considering its global variations and usage (Galloway & Rose, 2015). It acknowledges the changes that English undergoes with cultural contexts beyond traditional English-speaking countries such as the UK, USA, or Australia (Galloway & Rose, 2015). The GELT teaching method prioritises enhancing fluency, accuracy, and comprehension when speaking while equipping students with effective learning strategies. Learners need to be proficient in various communication contexts, particularly when English is used as a lingua franca (Vettorel, 2018). Adapting and communicating clearly is the key measure of success, rather than solely adhering to traditional English conventions (Rose & Galloway, 2019). Lastly, GELT acknowledges language as a dynamic entity and emphasises integrating and blending languages in real-life scenarios. Finally, there is a focus on providing learners with a sense of ownership (Matsuda, 2003; Widdowson, 1994, cited in Prabjandee & Savski, 2022).

GELT addresses the dichotomy between native and non-native speakers by focusing on all English users as the targeted interlocutors rather than a traditional native speaker model. In the field of English Language Teaching (ELT), the Global English Language Teaching (GELT) approach strives to enable non-native speakers to use English in a natural way in global settings. This approach emphasises the importance of comprehending and adapting one's communication style to different cultures rather than adhering to a single form of English. GELT helps students develop a global mindset, enhancing their communication skills and moving beyond the confines of traditional native-speaker standards Galloway and Rose (2018).

2.3.1.2. Common European Framework of Reference for Languages (CEFR)

The Common European Framework of Reference (CEFR) is a crucial standard in language education policy. It was developed in Europe to encourage mobility in line with EU integration policies (Prabjandee & Savski, 2022). Established in 2001, it provides guidelines for language learning, teaching materials, and assessing learning outcomes. Its influence is widespread, and it was created as a standardised approach to language learning and assessment across

the continent (Figueras, 2012). Specifically, it has had the intention of acting as a “curriculum guidelines, examinations, textbooks, etc. across Europe” (Council of Europe, 2001, p.1). Moreover, it states that “it describes in a comprehensive way what language learners have to learn to do in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively” (Council of Europe, 2001, p.1). The CEFR also defines descriptors of language proficiency as well as the global scale and self-assessment checklists (Little, 2006).

However, recent updates to the CEFR reveal a changing view of the native speaker's role in language learning and assessment. One of the most influential components of the CEFR is arguably the scales (Alderson, 2007; Deygers et al., 2018). However, these descriptors and scales explicitly referred to the native speaker, such as "understanding conversation between native speakers" or "understanding a native speaker interlocutor" (Appendix C in the CEFR 2001, Council of Europe, 2001). Consequently, McNamara and Shohamy (2016) state that these descriptors imply that non-native speakers only find themselves in situations where communication occurs with native speakers.

In response to changing global language dynamics, the 2018 CEFR Companion Volume (CEFR-CV) was updated to address the shortcomings related to linguistic and cultural diversity. Specifically, 16 descriptors at B and C levels were rephrased to exclude the term "native speaker." In addition, the terminology has shifted from "non-standard accent or dialect" in order to take into account variations or divergence. For example, the B2 descriptor has been changed from "Can understand in detail what is said to him/her in the standard language" to include "[...]or a familiar variety". Whereas the C2 descriptor "non-standard accent or dialect" has been changed to "less familiar variety" (Council of Europe, 2020, p.257). This change indicates a shift towards inclusivity, prioritising understanding over adherence to a native speaker standard. It also acknowledges that advanced speakers may have accents (North, 2021, p.12).

Therefore, the impact of Global English is apparent in these changes, highlighting a world where English is becoming the common language. The move away from 'native-like' norms to emphasising efficient communication underscores this impact. The changes made to the CEFR

reflect that rather than adhering strictly to a native speaker standard, the focus is now on intelligibility and effective communication, which is in line with the values of Global English.

2.3.1.3. Global English Writing Guide

While not pertaining to the paradigm of GE as described by Rose and Galloway (2015), some other guidelines have been suggested, specifically in writing. These will be briefly outlined below:

According to Kohl (2008), Global English is written English that has been optimised for a global reader by following rules that promote clarity. "The Global English Style Guide: Writing Clear, Translatable Documentation for a Global Market" emphasises that Global English is universally applicable and can benefit native and non-native English speakers. Using Global English, communicators can avoid confusion caused by ambiguity, uncustomary non-technical terms, and unfamiliar sentence structures for non-native speakers.

2.3.1.4. Key Features:

- Avoidance of idioms and colloquialisms since they may not be understandable to non-native English speakers.
- Avoidance of using phrasal verbs as they may confuse non-native speakers with their multiple meanings.
- Simple syntax where sentences follow a Subject-Verb-Object structure.
- Active voice promotes simple syntax.
- Simple or compound sentences instead of complex structures.
- Directness over the use of hedging for clarity.
- Avoidance of abbreviations and acronyms.

2.3.1.5. Summary of Global English (GE)

Global English (GE) refers to the different ways and contexts in which English is used globally. The sociolinguistic approach framework highlights the diversity of language, sociolinguistics,

and sociocultural aspects associated with the global use of English (Rose and Galloway 2019). Essentially, GE emphasises the adaptability and flexibility of English as it interacts with various sociocultural settings worldwide.

Linguistic diversity and fluidity in various sociolinguistic environments rather than representing a fixed or standardised language are embraced (Rose & Galloway, 2019). Furthermore, the GE paradigm acknowledges the plurality and diverse variations of the English language, such as WE and ELF. Overall, the goal is to understand language and promote clear communication and mutual intelligibility between speakers of different English dialects (Galloway & Rose, 2018).

GE emphasises the negotiation of meaning over strict adherence to one English variety. This is pertinent in ELF contexts, where mutual understanding is more important than linguistic precision (Fang & Ren, 2018). Finally, the concept of GE prioritises the sense of ownership that ELF speakers have over the language. This means that instead of treating native speaker norms as the ultimate standard, the focus is on valuing each speaker's unique relationship with an adaptation of English (Fang & Ren, 2018).

2.4. Automated Speech Recognition Technology

In the last two decades, there has been a significant increase in the usage of speech-based interactions between humans and machines. This includes applications that enable natural language comprehension (El Hannani et al., 2021) and speech technologies such as automatic speech recognition (ASR), which are used by millions of people worldwide (Koenecke, 2020). One example of an ASR application is virtual assistants, which allow users to control devices and access information through voice commands. ASR has many speech-to-text applications, including subtitling and dictation. In a professional context, especially in online meetings, this technology is particularly useful as it can transcribe speech in real time, generating automated closed captions that improve accessibility. In addition, more and more technological tools are available that allow for the translation of spoken language between individuals who speak different languages (Markl, 2020; Koenecke, 2020).

Also known as machine transcription or Speech-to-Text, ASR uses machine learning techniques to analyse and decode acoustic input and patterns, which are then converted and transcribed into spoken words (Google, n.d.). According to Dong & Li (2015), the main function of ASR technology is to interpret a word sequence from the shape of the speech wave. There are many ways to achieve this, but modern ASR systems primarily utilise advanced machine-learning algorithms (Koenecke, 2020). Over the past few years, the quality of these systems has greatly improved thanks to the advancements in deep learning for speech, natural language processing and the utilisation of large-scale datasets for training (Koenecke, 2020).

Nevertheless, training systems to understand human behaviour continues to be difficult as the audio input that these systems rely on contains granular information that is challenging to interpret (Claval, 2013). As highlighted by Levis and Suvorov (2012), problems include speech variability, such as different voices, accents, styles, contexts, and speech rates. Additionally, recognition units, like words and phrases, syllables, phonemes, and diphthongs, also pose a challenge. A further issue is language complexity. This refers to vocabulary size, complexity, and lexical density. Ambiguity is another aspect which can cause misinterpretations. This can commonly be seen with homophones, word boundaries, and syntactic and semantic ambiguity. A final factor is the environmental conditions, such as background noise or multiple people speaking simultaneously, which present additional obstacles (Levis & Suvorov, 2012).

Several other factors lead to inaccuracies in speech-to-text transcriptions. The issue of variability is especially noteworthy, particularly when it comes to non-native speakers. Many speech systems rely on models of native English speakers, which means that occurrences of non-standard forms can be unrecognisable as they might not align with the training data. There exist numerous studies that support the view that a scarcity of datasets that cater to variabilities results in poorer performance (Van Doremalen et al., (2009); Errattahi et al., 2018). In fact, it is argued by some scholars that variation should be accounted for as it is something which is inherent to a language (Markl, 2022). Moreover, language varieties are not equal across speech communities but seem to be linked closely to the identity of the speaker. However, one problem is that machine learning models typically require more training data to be able to enhance their accuracy which means that they tend to perform

worse for smaller populations within a training dataset (Suresh & Guttag, 2021, cited in Markl, 2022).

The above raises questions regarding the extent to which ASR systems represent the global reality of their non-native users. Therefore, this section will provide an overview of some of the fundamental aspects related to ASR technology, challenges, metrics, and user bias.

2.4.1. Fundamentals of ASR

As stated by Pérez Castillejo (2021), the latest ASR systems use Natural Language Processing (NLP) technology to analyse and convert speech into text that both people and machines can understand. However, due to intellectual property rights, it is not always possible to access the technological specifications of advanced ASR or state-of-the-art systems. This has led to products such as YouTube video captions and Google Voice search being referred to as black-box speech recognition services (Hannani et al. 2021), and some scholars argue for more transparency surrounding industry ASR, especially regarding training corpora (Wassink et al., 2022). Regardless, ASR systems tend to capture audio input from a user through a microphone, analyse it with a pattern, model, or algorithm, and generate an output, which is often in the form of text (Li et al. 2016). While the use case for ASR systems may vary, they typically operate using an acoustic model, a language model, and a decoder (Bouillon et al., 2016, as cited in Anastassiadis Serrat, 2021; Tatman, 2020, as cited in Wassink et al., 2022).

A) Acoustic Model:

The acoustic model (AM) captures audio input as a user speaks into an ASR application. Following this, speech sounds are matched to words, which are subsequently translated into text (Pérez Castillejo, 2021). To enable this process, the information received needs to be parsed into phonemes, which are the smallest perceived linguistic units of sound and the basic building blocks from which words are formed. The ASR system uses statistical probability to analyse the phoneme sequences it detects. Following this, it deduces words that best match those sound strings by referencing the system's pronunciation dictionary (Pérez Castillejo, 2021).

B) Language Model:

The Language Model (LM) is designed to analyse language statistics. It can recognise which series of words are commonly used and predict the frequency of common phrases, n-grams, or words (Tatman, 2020, as cited in Wassink et al., 2022).

C) Decoding:

The decoder's task is to determine the most accurate transcription of the audio. This process involves utilising both the acoustic and language models to generate a transcript.

2.4.2. Challenges of ASR

Promoting the accessibility of ASR tools on a global scale to users is challenging. According to Google (n.d.), their aim is to categorise worldwide information and make it available to everyone. This means ensuring that their products are compatible with multiple languages and highlights the importance of enabling Google Assistant to comprehend human speech. Developing high-quality ASR systems requires extensive amounts of audio and text data, and as the industry is being transformed by advanced neural models that rely on vast amounts of data, the need for this data is becoming increasingly important. The availability of such data is unfortunately limited for many languages (Google, n.d.), and consequently, this can lead to inaccuracies. For example, in some cases, speech sounds may not match perfectly with the specific units of sound found in the phonetic dictionary. This can be seen in traditional methods for training language models, and it has been argued that if variations in one language are not present in another, it can lead to distortion (Li et al. 2020). According to Rodman (1999, as cited in Levis and Suvorov, 2012), speech recognition systems are based on three key factors: speaker dependence, speech continuity, and vocabulary size.

Firstly, there are three categories when it comes to speaker dependence. Speaker-dependent systems require specific training for each speaker to work effectively. Speaker-independent

systems are more versatile as they are trained on various voice samples from different individuals, enabling them to recognise voices they have not been specifically trained on. Adaptive systems start with the versatility of a speaker-independent model but adjust and refine their recognition capabilities over time to better align with a specific user's voice.

Secondly, when it comes to speech recognition, there are different levels of continuity that these systems can handle. Some are only able to recognise isolated words that are spoken distinctly and separately. Others can decipher individual words even if they are pronounced continuously without clear pauses. More advanced systems can seamlessly identify entire sentences, regardless of the pauses (or lack thereof) between words. Additionally, there are systems that are skilled at recognising specific words or phrases within a continuous stream of speech.

Finally, the size of the vocabulary is a crucial factor in speech recognition. Certain systems are customised to recognise only a limited number of words, while others can recognise a wide range of vocabulary (Dong & Li, 2015). In addition, according to Anastassiadis Serrat (2021) ASR systems can struggle with recognising certain words, such as abbreviations, technical terms, and newly emerged language, due to limitations in their pronunciation dictionary (Anastassiadis Serrat, 2021). Based on the above, the accuracy of speech recognition systems is influenced by these three aspects and is susceptible to three types of errors (Rodman, 1999, cited in Levis and Suvorov, 2012):

> "The performance of speech recognition systems depends on each of the three dimensions and is prone to three types of errors: errors in discrete speech recognition, errors in continuous speech recognition, and errors in word spotting. Errors in discrete speech recognition include deletion errors (when a system ignores an utterance due to the speaker's failure to pronounce it loudly enough), insertion errors (when a system perceives noise as a speech unit), substitution errors (when a recogniser identifies an utterance incorrectly."

While there have been notable advancements in ASR technology in recent years, the fundamental errors shared by Rodman (1999, cited in Levis and Suvorov, 2012) are still

relevant in helping to comprehend the basic design principles of ASR and the potential challenges that may arise from the interaction between humans and systems.

Sociolinguistic factors:

Research by Jurafsky and Martin (2020) highlight the relationship between sociolinguistic factors and ASR efficacy (Jurafsky and Martin, 2020, as cited in Anastassiadis Serrat, 2021). Moreover, the following factors are highlighted:

- Wassink et al. (2022) point out that the phonetic variations within a dialect are relatively minor compared to those between distinct languages. Yet, ASR systems often struggle with the multitude of ethnicities and dialects, and neglecting such variations during training can compromise the accuracy of these systems. Moreover, Hinsvark et al. (2021) observed a significant decline in ASR performance when processing accented speech, which is characterised by the unique phonetic and intonation patterns of a particular group of speakers or speech community. This is distinct from a dialect, which may have variations in vocabulary or grammar but is usually mutually intelligible. Significant linguistic variations, such as lexis and syntax, may exist when comparing dialects (Wolfram & Shilling, 2006).

- Coarticulation, as described by Jurafsky & Martin (2023), occurs when phonemes, the units of sound, are inconsistently pronounced due to the influence of adjacent sounds. This inconsistency arises because speakers either anticipate forthcoming sounds or prolong preceding ones, causing alterations in pronunciation. This phenomenon is especially prevalent in continuous speech, where words are articulated in a flow rather than in isolation.

- The rate of speech, commonly quantified in syllables or words per minute, also plays a crucial role. Speaking at a rapid pace can lead to phonetic reductions, such as elisions, incomplete pronunciations, or word groupings (Jurafsky & Martin, 2023).

2.4.3. Word Error Rate (WER)

There are multiple ways to measure the accuracy of speech, but the word error rate (WER) is the most widely accepted standard in the industry. According to Koenecke (2020), WER is a measure of the difference between machine-generated (Hypothesised) and human-generated transcriptions (Reference). WER is calculated by counting the number of errors in the hypothesised transcript and comparing it to the number of correct words in the reference transcript. In simpler terms, WER measures the ratio of wrongly transcribed words to the total number of words, and a lower WER indicates higher accuracy (Google, n.d.; Microsoft, 2023).

WER is formulated as:

$$WER = \frac{S + D + I}{N}$$

The Word Error Rate (WER) metric measures inaccuracies by combining substitution (S), deletion (D), and insertion (I) errors and dividing by the total number of words (N) in the reference transcript, also known as the "ground truth" (Errattahi et al., 2018). Ground truth is a transcription that is considered 100% accurate, typically created by humans, and used to evaluate the system's accuracy (Google, n.d).

2.4.4. Substitution, Deletion and Insertion

Substitution errors (S) occur when a word is replaced with an incorrect transcription (Lai and Markl, 2021; Rodman, 1999, cited in Levis and Suvorov, 2012). Deletions (D) refer to words that are missing in the hypothesised transcript but are present in the ground truth version. Lastly, insertions (I) are words that appear in the hypothesised transcript but were part of the original speech sample. These errors can be the result of speaker volume, background noise, or incorrect identification of speech units. As previously mentioned, speech recognition systems can sometimes misinterpret continuous speech by merging word boundaries. This can result in mistakes where the system confuses single speech units with multiple units or vice-versa.

ASR systems can differ in the type and frequency of errors they produce, with some having a higher occurrence of deletion or substitution errors. It is essential to consider these errors, as excessive deletion errors can make a transcript challenging to understand, whereas substitution errors can alter the meaning of the text depending on its phonetic or morphological nature (Rodman, 1999, as cited in Levis & Suvorov, 2012; Errattahi et al., 2018; Markl, 2021).

2.4.5. Previous investigation into ASR

Numerous studies have investigated ASR bias in the context of non-standard English. Empirical evidence suggests that these systems inadvertently encourage the marginalisation of certain speakers from specific groups or communities (Coniam, 1999; Derwing et al., 2000; Pasandi & Pasandi, 2022; Tatman, 2017; Koenecke et al., 2020; Chan et al., 2022).

For example, regarding accented speech, research conducted by Coniam (1999) and Derwing et al. (2000) explores the difficulties encountered by non-native speakers when utilising ASR technologies. The former investigates Cantonese-accented speech, whereas the latter examines the pronunciation difficulties faced by Spanish-speaking English learners when using ASR systems. Both findings revealed that the accuracy of the output produced by second-language speakers was notably inferior to that of native speakers.

Although ASR has improved in recent years, issues with non-native accents and users appear to persist. For example, Chan et al. (2022) highlight significant biases in ASR systems when processing different English dialects. According to the authors, there is a discernible systematic bias against speakers whose native language is tonal, i.e., Mandarin or Cantonese. For example, L1 transfer of pitch variations from their native language into their own can cause differences in prosodic patterns that automatic speech recognition (ASR) systems are trained to recognise, leading to reduced accuracy.

Moreover, there is a growing body of research highlighting ASR marginalisation due to factors such as race, gender, and dialects within native-speaker communities. Koenecke et al. (2020)

show that ASR systems display significant racial discrepancies between black and white English speakers. In 2017, Tatman's analysis of YouTube's automatic captions found significant disparities in accuracy between women and speakers from Scotland (Tatman, 2017). On the other hand, research has shown that ASR systems, which are trained to recognise dialect variations, greatly enhance accuracy. For instance, Dorn's study (2019, as cited in Wassink et al.) revealed that ASR systems trained in African American English exhibited an accuracy enhancement of over 16.6% when tested on African American samples, as opposed to those trained in Standard American English.

Moreover, Lai & Markl (2021) conducted a study on two British English automatic speech recognition (ASR) systems created by Google and Amazon. Their findings showed that these systems have difficulty accurately recognising the speech of second-language English speakers and those who speak certain (stigmatised) regional variations of British English. The research suggests that despite technological advances, ASR systems may have an inherent bias against non-standard English varieties. The study found two major challenges: phonetic and morphological/syntactic errors. Regarding pronunciation, the main issue was substitution errors. As for morphology and syntax, the Google transcripts contained tense-related errors, such as frequent replacement of "lived" with "live". Distinct phonetic substitutions are also observed. Furthermore, in several recordings, "would" is replaced with "will". The authors suggest that these mistakes could be due to training data containing present tense verb forms. and argue that the findings reveal a standard language ideology that favours an ideal standardised language, often overlooking regional accents and dialects (Lai & Markl, 2021).

This section has provided a brief overview of the fundamental aspects concerning the functionality, challenges, and metrics of ASR systems. Additionally, it has provided an overview of previous research relating to variability and accuracy concerning non-native speakers. These findings show that even with advancements in ASR, sociolinguistic factors along with variations in speaker output, can still have a negative impact on the accuracy of transcripts. Furthermore, the emergence of Automated Speech Recognition (ASR) technology is significant as its recent rise in users is aligned with a prevailing trend of global English.

2.5. Startups

The COVID-19 pandemic led to a shift in corporate communication practices, with remote collaborations and virtual meetings being conducted on online platforms like Microsoft Teams, Zoom, and Google Meet. During this period, digital tools enabled businesses to sustain operations during times of uncertainty, but it also had the unintended benefit of fostering innovation while facilitating greater access to a more diverse workforce (Forbes, 2023). This change is apparent in the startup sector, where digital advancements have enabled entrepreneurs to have an international vision from the conception of a new business idea (Cavusgil & Knight, 2015). As the preferred language for business, it is argued that English promotes effective communication with people from different linguistic and cultural backgrounds (Crystal, 2005; Seone & Suarez-Gomez, 2016). Moreover, Turunen & Nummela (2017) assert that entrepreneurial capabilities, including language proficiency, a global mindset, and cultural awareness, are pivotal to entrepreneurial success. Not to mention that studies suggest language skills play a role in recognising and taking advantage of international opportunities for small and medium-sized enterprises (SMEs) (Aston University, 2021). Therefore, it can be argued that English seems beneficial to startup founders from non-English speaking regions as it can help to facilitate the internationalisation of their product or service.

2.5.1. Startup internationalisation

The term startup is ambiguous since it includes a diverse group with varying business models, products, and resources (Carpenter, 2015). De Bernadi and Azucar (2020) define a startup as a newly emerged business with an innovative idea, developing a business model to meet the marketplace's needs. Moreover, the terms ‘Born Globals’ (BGs), ‘International New Ventures’ (INVs), and ‘Global Startups’ are used interchangeably to describe startups with a clear international focus. In other words, they are ventures that explore global opportunities from incubation (McDougall & Oviatt, 1995). According to Korhonen (1997 as cited in Englis, 2007), these startups engage in cross-border activities like international sourcing, resource building, and early international cooperation in product development.

Neubert (2018) states that startups should prioritise internationalisation at an early stage. For example, by utilising digital tools for foreign market development, they can identify new market opportunities more efficiently and save resources while focusing on the most promising markets globally. Internationalisation refers to adapting these operations, products, services, and strategies to foreign markets and cultures. Startups internationalise by considering linguistic, cultural, legal, regulatory, and market differences to tap into new growth opportunities, diversify their customer base, and achieve economies of scale (McDougal & Oviatt, 2005). In addition, Cavusgil and Knight (2015) state that processes related to globalisation, such as modern communication technology and greater internet accessibility, have reduced the cost of internationalisation, making foreign expansion more accessible to smaller, under-resourced companies. Interestingly, this is in line with the views of Gabrielsson and Pelkonen (2008), who claim that the rise of digital goods providers will lead to the earlier and faster internationalisation of startups.

Furthermore, internationalisation is viewed favourably by many institutions, including the European Union. They believe it is essential to business success, as it creates jobs, encourages innovation, and boosts profitability (Lilischkis et al., 2016). Startups contribute to technological advancement, new market creation, better service quality, and lower unemployment rates in society and the economy (Corl, 2019). For example, Europe's technology industry is expanding at twice the rate of the global economy Gauthier et al. (2017).

2.5.2 Spanish Startup Sector

A recent report shows that the value of the Spanish tech ecosystem has increased 3.8 times since 2018 and is now worth €93 billion. As a result, Spain ranks 6th in Europe and 16th globally for total investment raised in 2022 (Dealroom, 2023). According to PWC's report from 2023, Spain ranks fourth in Europe in terms of the number of startups, with a total of 11,100. These emerging companies provide employment to 140,000 people, resulting in a 20-fold increase in the value of the Spanish entrepreneurial ecosystem over the last decade (PWC, 2023). Enisa's report for 2021 states that Venture Capital (VC) investment in Spanish startups, which refers to professional management firms or investors that fund innovative ideas, has

grown faster than any other European country except for the Netherlands in the first half of the year (Enisa, 2021). The report reveals that Spanish and European investors hold the majority share of the country's venture capital, with 59% coming from domestic and cross-border investments (Enisa, 2021). The data also demonstrates that 76% of investment derives from non-Spanish-speaking countries. Moreover, according to the Startup Report (2023), 68% of Series A investment for growth comes from foreign sources, leading to over 20% of scale-ups relocating their headquarters to other markets.

The Spanish startup sector has recently undergone some recent changes which could further reinforce the importance of multilingualism. Firstly, favourable conditions regarding “digital nomads” have recently been established. The objective is to attract entrepreneurial talent and remote workers back to Spain. Secondly, the Spain Startup Law (2022), which came into effect in January 2023, is ‘one of the milestones of the Recovery, Transformation and Resilience Plan’. This new law aims to attract investments, entrepreneurship, and talent, making Spain as desirable as other European countries. The initiative also encourages a culture of innovation within the European Union, creating conditions that support the growth of new and innovative companies. (Ministry of Economic Affairs and Digital Transformation, 2022). As can be seen, conditions are being established to actively create and internationalise businesses within Spain and arguably, this could result in more multilingual interactions.

2.5.3 Global English in the Spanish Startup Sector

It is important to note that GE is relevant to both startups with an international focus and those operating domestically. Even if entrepreneurs work in their home country, their work environment is typically highly globalised. Entrepreneurs usually collaborate with foreign individuals and organisations from multiple countries, particularly at technology conferences or summits, even when their initial idea is domestically discovered (Englis et al. 2007).

Networking, both formal and informal, is crucial for Spanish entrepreneurs to enter the market, identify opportunities, and acquire resources (Coviello, 2015). As a result, international conferences, workshops, and networking events in Spain typically adopt an ‘English-first’ policy for building global partnerships and collaborations (Nevado-Peña, 2018).

For example, South Summit is a platform where leading companies, startups, investors, and institutions from around the world can showcase their innovations, build relationships, identify opportunities, and generate business (JPMorgan, 2023). This event is internationally recognised and is the biggest celebration in Spain (Hernández, 2023), attracting approximately 20,000 participants, of which 6,500 are entrepreneurs and 2,000 investors. Valencia Digital Summit, a startup gathering to promote the region's talent and prowess as an international tech hub. According to sources, it was attended by more than 12,000 attendees from over 35 different countries (Tech.EU, 2022). Food 4 Future is an innovation event held in Bilbao that showcases industrial foodtech solutions and trends that are driving transformation in the food industry.

Secondly, another important aspect is the use of English during negotiations and the procurement of funds. Major venture capital hubs like Silicon Valley, London, and New York primarily operate in English. Proficiency in English can help in securing funding through pitching. A pitch is a presentation where entrepreneurs must concisely yet charismatically convey the value of their innovation (Fairbairn et al.). It can be seen as a performance that has an impact on the success of individual startups and even entire economic sectors (Fairbairn et al.). Entrepreneurs use pitches to showcase the potential of technology, generate value, and attract investor capital. In essence, a pitch legitimises the startup idea and creates market opportunities (Lounsbury and Glynn, 2001), offering opportunities to receive feedback that can transform their idea from an envisioned project to a viable business (Benton 2020 cited in Fairbairn et al.).

In Spain, there are numerous mentorship programs available for startups to facilitate their growth. These programs vary from incubators that offer support at the initial stage to accelerator programs that focus on mentoring, training, scaling, and pitch presentations. Typically, mentorship programmes are arranged into batches or cohorts with a wide range of technological interests (Ester, 2017). Some of these programmes may take equity or commission, and they tend to have an intensive schedule lasting (approximately three to four months or less). A common feature of these programmes is that they finish with a Demo Day, during which minimum viable products (MVPs) are presented to potential investors. Generally speaking, startup founders usually pitch for 5-10 minutes to invite-only investors in various

formats. Given the emphasis on internationalisation, Spanish mentorship programmes often require Demo Day pitches to be delivered in English. As demonstrated, it can, therefore be argued that having language skills, particularly in English, may help with internationalisation.

2.5.4 Global English implications for Spanish startups

In theory, Spanish startup founders deliver their pitches in English for the purpose of internationalisation. However, in practice, members of the audience might be monolingual Spanish-speaking natives. For example, Demium Capital, which has its headquarters in Valencia, is Spain's most active investment firm, having carried out 28 operations (Bankinter, 2022). Investment pitches are delivered in English and sometimes online. Other international tech events, such as South Summit or Valencia Digital Summit adopt 'English-first' policies that encourage or require competition pitches to be delivered in English. This is also true for numerous other mentorship programmes where increasingly participation is 100% online. Interestingly, while pitching in English is conducive to internationalisation, in monolingual environments where audience English proficiency between Spanish speakers varies, the requirement to pitch in English could impede comprehension.

3. Methodology

English is becoming more important in the global startup scene and Automated Speech Recognition (ASR) plays a more prominent role in online communication. As mentioned, this TFM aims to investigate the effectiveness and potential biases of ASR tools in this context by recording Spanish startup founders presenting their pitch in English.

3.1. Research questions:

RQ1. To what extent are ASR systems tolerant of variations in the oral output of non-native English speakers, particularly Spanish founders?

RQ2. How do the errors in ASR transcription correlate with sociolinguistic factors such as age, gender, and level of English proficiency among Spanish startup founders?

RQ3. Regarding intelligibility, to what extent are variations in output intelligible for human comprehension but not ASR transcription?

RQ4. Are there any common linguistic patterns used by Spanish speakers of English that cause inaccuracies with ASR transcriptions? Should certain elements be given more focus or avoided?

3.2. Data collection

Participants: Startup accelerators, incubators, organisations, and institutions in the Valencia region were contacted to request startup participation. Moreover, Spanish CEOs, founders and entrepreneurs were contacted directly via social media. The data collection process involved a survey and an online meeting with 10 participants.

Survey: Data collection adopted a qualitative approach. The participants were required to complete a Google Survey of 15 questions focused on sociolinguistic factors such as age, gender, primary language (L1), and whether they spoke a non-Spanish mother tongue, such

as Valencian. Additional factors, such as the participants perceived level of English, were collected. Other questions relate to the experience participants have with presenting pitches and the number of times they have presented the pitch that would be recorded for this investigation. This information is pertinent as participants with more practice might present with a higher level of fluency which could impact WER scores. Moreover, recording participant data is useful as this can be cross-referenced with inaccuracies in transcriptions, which might reveal patterns between ASR and sociolinguistic information. Identifying these aspects will help to address RQ2. (See Appendix I: Sample of questions from the questionnaire)

Online Meetings:

Video recordings and transcriptions were collected via the video conferencing platform Google Meet. Participants signed up via a booking link and presented their 5–10-minute pitch online. This platform was chosen as it uses state-of-the-art ASR and is easily accessible to participants. Each candidate signed a consent form to allow the use of the recording for research. (See Appendix II: Consent form)

Format:

Various types of pitches are performed in different contexts for different audiences (Chapple et al., 2021). This investigation focuses on the Demo Day pitch. Traditionally, a Demo Day pitch lasts 5-10 minutes. Accordingly, all participants were asked to present within this time. Demo Day pitches tend to follow the same structure, e.g., the presentation of the problem, solution, product, and team. For this investigation, this is quite useful as it promotes a degree of consistency. Lastly, as the study focuses on extemporaneous speech, participants were requested not to read from notes.

Reliability:

Li et al. (2016) state that noise refers to unwanted disturbances superposed upon the intended speech signal. Background noise, voices and other sounds can interfere with ASR transcriptions. Moreover, Errattahi et al. (2018) refer to ‘mismatch factors’. They argue that

differences in hardware, transmission channels, and recording devices can introduce variability during recording and decrease system accuracy. To improve the accuracy of ASR transcriptions, participants were instructed to record in a quiet room. In addition, recordings with background noise or poor quality were discarded.

It should be noted that speakers were not required to use the same hardware. Thus, quality cannot be guaranteed 100%. To mitigate this, online meetings were recorded from the same location and under the same conditions to achieve consistency.

3.3. Data processing and analysis

A spoken corpus will be created to evaluate ASR precision, and a series of experiments will be employed.

Experiment 1: ASR Transcript Analysis – Error Prediction

Regarding RQ1 and RQ3, audio from the captured pitches will be transcribed using Google Meet's state-of-the-art ASR. Following this, the transcripts will be analysed using a rubric that aligns with the current GE paradigm. This rubric prioritises intelligibility and mutual understanding over standard native norms. During this initial stage, errors related to syntax and lexis will be identified. Firstly, this phase aims to acquire an initial overview of precision and intelligibility. Secondly, this approach can provide insights into the tolerance of the ASR system. For example, by comparing the transcript errors with the ground truth transcription to see which non-standard utterances were accurately transcribed. Overall, it will assess how much human vs. machine comprehensibility differs.

Experiment 2: Comparative Analysis – ASR tolerance

To address RQ1 and RQ4, the ASR-generated transcripts (hypothesised transcripts) will be compared against the ground truth version for each speaker (reference transcript). Comparisons between the predicted errors and reference corpus can give insights regarding the tolerance of ASR systems with non-native variations as outlined in Experiment 1).

Experiment 3: WER analysis

The hypothesised ASR transcript and HT reference transcript will be compared, and a separate corpus will be created to register the confirmed number of insertions, deletions, or substitutions. The average word error rates will be calculated globally and compared with participant details to evaluate the impact of sociolinguistic factors on ASR transcription. As discussed, it has been noted that natural language processing has faced challenges in the past due to sociolinguistic variation (Tatman, 2017). Therefore, this stage will help to address RQ2. Additionally, the same rubric used in experiment 1 will be used to evaluate the intelligibility of the speaker's utterance. This approach will help to establish the extent to which variations are understandable to a human listener but not to the ASR system (RQ3).

Experiment 4: Qualitative analysis of potential causes

This stage aims to better understand the potential causes behind ASR inaccuracies, such as variations in pronunciation, lexis, grammar, or other factors. This stage will help address RQ4. As mentioned, ASR systems use statistical probability to analyse the phoneme sequences it detects. Following this it deduces words that best match those sound strings by referencing the system's pronunciation dictionary (Pérez Castillejo, 2021). Therefore, this stage will take this into account when analysing the data by considering errors at a phoneme level.

3.4. Criteria for Errors

There is a noticeable contrast between spontaneous or natural speech and prompted speech. When people engage in conversations, they often speak at a faster pace, do not articulate as clearly, may repeat words or correct themselves, and use a more unique and varied vocabulary that is tailored to their particular social group (Google, n.d.). This investigation aims to analyse extemporaneous speech in the context of a startup pitch. Therefore, some considerations are required regarding the categorisation of errors. Startup and Spanish names will be omitted as these are unlikely to be recognised by the phonetic dictionary of the ASR system. Moreover, reformulations will not be considered as this could unfairly impact the WER score (Anastassiadis Serrat, 2021).

3.5. Rubric

To assess the output of both the ASR system and the speaker, an intelligibility rubric has been adapted from various sources, including an adequacy scale (adapted from Arnold et al. 1994 cited in Calefato 2015) and the updated version of the Common European Framework of Reference Companion Volume (2018). As mentioned, the CEFRCV was revised to prioritise intelligibility over native-like speech, making it more suitable for a Global English context.

Value	Description
1	**Adequate** It is clear, intelligible, mostly grammatically correct, causing almost no problems for the reader.
2	**Fairly adequate** It is generally clear and intelligible, and one can (almost) immediately understand what it means. Despite errors in grammar or lexis the reader is likely to interpret the message correctly based on the context.
3	**Poorly adequate** It contains grammatical errors and/or poor word choices. The general idea is partly intelligible but may be difficult to discern. It requires careful attention from the reader and is likely to cause misunderstanding.
4	**Completely inadequate** It is unintelligible or not possible to deduce meaning. It contains grammatical errors and/or poor word choices which are highly likely to cause a loss of meaning and impede reader comprehension.

Figure 1. Adequacy scale (adapted from Arnold et al. 1994 cited in Calefato 2015 & CEFRVC updated scales 2018)

Regarding errors related to the transcriptions themselves, it is not clear which errors will be present therefore all occurrences such as syntax, morphology.

4. Analysis and results

Firstly, results from the questionnaire will be discussed. This information is useful in helping to understand sociolinguistic information related to the participants and the data sample. The questionnaire collected quantitative and qualitative data related to English level and experience within the Spanish startup sector. This information can provide insights regarding English usage in the Spanish startup community and to a greater extent a wider global English context.

4.1. Questionnaire

4.1.1 Demographics

The table below presents the demographics of participants involved in the experiment.

Demographics	**Description**
Age	Three distinct age groups are present in this experiment: 18-24 (n=1) 25-34 (n=4) 45-54 (n=5)
Gender	Participants. n=8 males and n=2 females.
Mother Tongue	The majority are native monolingual Spanish speakers (n=9). One participant is multilingual and speaks Valencian as their mother tongue (n=1).
English Level	Levels are almost evenly distributed between B2 (n=4) and C1 levels (n=6).
Company position	The majority of the participants are CEOs (n=7). CFO (n=1), CMO (n=1) and Other (n=1)

Table 1: Summary of participant data

The data provides valuable insights into the demographics of the participants. Nevertheless, it is important to note some limitations. Firstly, the sample size is relatively small (n=10), which means that the findings may not be representative of the broader population. Secondly, there is a significant gender imbalance in the sample, with a higher number of male participants. Therefore, this should be taken into account when drawing conclusions between the efficacy of ASR systems and sociolinguistic factors.

4.2. Use of English in the Spanish Startup Ecosystem

The following section provides information related to participant use and experiences of English use in the startup sector.

4.2.1. English Context

Situation	**Nº Frequency**	**% of participants**
Networking	N = 7	70%
Presentations	N = 7	70%
Seeking foreign investment	N = 6	60%
Online client meetings	N = 5	50%
General meetings	N = 5	50%
Recruitment	N = 3	30%

Table 2: Results based on situational context in which participants use English.

As can be seen from the table above, the majority of the participants (n=7) use English for networking opportunities and for giving presentations. Similarly, 60% of the participants utilise English when seeking foreign investment, underscoring the language's importance in international financial interactions. In a similar vein, 50% of the participants use English for online client meetings and general meetings, highlighting its significance in everyday professional communication. Lastly, 30% of the participants use English in recruitment processes, which indicates that some teams might be formed of multilingual members.

Overall, the data appears to support the viewpoint English is used as a business lingua franca in startup sector.

4.2.2 Native vs. Non-native Interactions

The following information relates to the frequency of English use per week in hours contrasted with the number of hours exposure speaking English to native speakers.

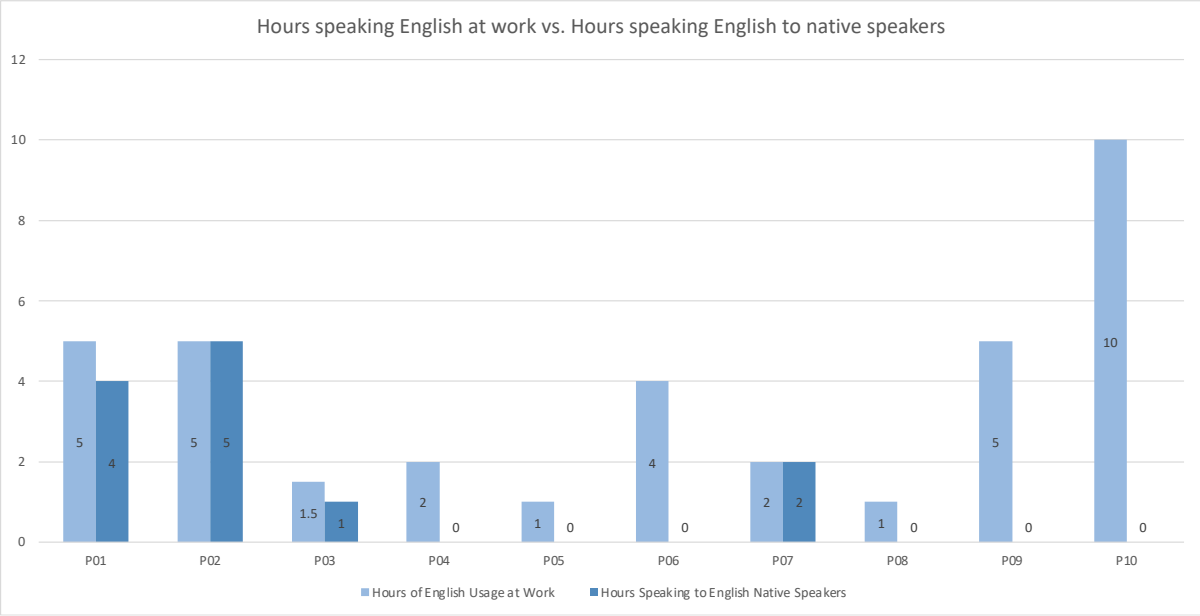

Figure 2: Hours per week speaking English vs. Hours per week interacting with native-English speakers.

Notably, one participant (P10) uses English at work for ten hours per week, which is the highest among the participants, whereas other participants (P05 & P08) speak English for only 1 hour a week. Interestingly, in terms of interactions with native English speakers, the majority of participants (6 out of 10) do not spend any hours per week doing so. Those who do, spend between 1 to 5 hours per week. It is worth noting that despite using English at work, the majority of participants do not interact with native English speakers (60%). However, the data specifically refers to oral interaction and does not consider reading or writing communication such as email. Lastly, the two participants (P01 and P02) who spend the most time speaking to native English speakers also have a relatively high weekly usage of English at work (5 hours), suggesting that their type of work or role possibly involves more interaction with native English speakers compared to the others.

However, comparisons of native vs. non-native interactions and the situational contexts in which English is used reveals some interesting trends. For example, P01 and P02 use English for 5 hours per week and interact with native English speakers for 4 and 5 hours, respectively. Their usage of English spans across multiple contexts, including online client meetings, networking, presentations, seeking foreign investment, and recruitment. This high interaction with native speakers could be attributed to the nature of their work, which involves diverse situations requiring English proficiency, and possibly, interactions with international clients or stakeholders who are native English speakers.

On the other hand, P04, P05, P06, P08, P09 & P10 do not interact with native English speakers despite using English at work. These participants use English in various contexts such as seeking foreign investment, online client meetings, networking, presentations, general meetings, and recruitment. As a result, this data appears to align with the paradigm of an increasing number of non-native-to-non-native interactions from expanding circles (Crystal 2005; Rose and Galloway, 2015; David Graddol, 1999). For example, P10 uses English for 10 hours per week but does not interact with native English speakers, indicating that a significant amount of time is spent using English in professional contexts that are absent of native speakers.

4.2.3. Audience Demographic

There is an assumption that English serves as a lingua franca in a Spanish business context, especially during a pitch and networking gatherings. The data below represents whether participants have experience pitching in English to native Spanish speakers.

Response	**Number**	**%**
Yes, members of the audience. Yes, members of the panel.	N=9	90%
Yes, members of the audience.	N=1	10%

Table 3. Audience demographics – Pitching in English to native Spanish speakers

As can be seen, every participant in this sample has pitched in English at least once to an audience that contains native Spanish speakers. Moreover, nine of the participants, responded stating that they have presented their pitch to panel members. For clarification, panel members refer to judges or competition pitches or investors during demo day pitches. The trend from this sample suggests a high frequency of situations where professionals find themselves needing to pitch in English to native Spanish speakers.

4.2.4. Experience Pitching Online

Response	Number	%
Yes	N=9	90%
No	N=1	10%

Table 4. Experience pitching in English in an online context.

The data indicates that almost all participants (90%) have experience pitching online. This could be viewed as supporting the relatively recent trend of digital transformation that has led to a rise in online meetings and pitches.

Comparisons between audiences and experiences pitching online serve to support some of the hypotheses of this investigation. Firstly, it demonstrates that Spanish startups sometimes present in English to audience or panel members, potentially L1 monolingual Spanish speakers. Secondly, it underscores the use of videoconferencing platforms given that 90% of participants having experience pitching online.

In this context there could arguably be a greater use of ASR tools and this, in turn, further underscores the importance of ensuring that these tools are effective and accessible for non-native English speakers. The focus on this investigation is on the efficacy of ASR tools in multilingual contexts. However, this does raise the question as to whether inaccuracies in transcription could impede comprehension between speakers with the same mother tongue when participating in pitches in English.

4.3. ASR Corpus Overview

The automated speech recognition transcripts from each meeting were saved for analysis. The specification for the corpus is as follows:

Corpus statistics	
Speakers	10
Total time recorded.	70 minutes, 9 seconds
ASR words transcribed.	7,932
ASR tokens recorded.	9,006
Domain	Startups, investment

Table 5. ASR corpus summary

5. Corpus Analysis and Findings from Experiments

Using the data from the corpus a series of experiments were conducted during four different stages as outlined in the methodology. These will be discussed and analysed in the following sections.

5.1. Experiment 1: ASR Accuracy and Intelligibility Scores

The hypothesised transcripts were reviewed, and errors were recorded without reference to the ground truth versions. The objective is not to draw conclusions from the data, instead this approach aims to facilitate a preliminary understanding of the transcript's overall comprehensibility from the reader's standpoint. Furthermore, potential errors were identified. These can be compared at a later stage to provide insights regarding ASR tolerance. Errors were categorised according to three main categories: Syntax (S), Lexis (L), and Morphology (M). The intelligibility of the written output was assessed using an adapted rubric. This section addresses RQ1 and RQ3.

Results:

In total, there are 265 suspected instances of errors present in the corpus. These are divided into (S) n=106, (L) n=145 and (M) n=18. A breakdown of the error type can be found in the following tables:

Syntax	**Number**	**%**
Sentence fragment (Sf)	N=41	40.59
Word order (Wo)	N=19	18.81
Article (A)	N=13	12.87
Preposition (P)	N=12	11.88
Agreement (Ag)	N=9	8.91
Missing word (Mw)	N=7	6.93

Table 6. Summary of potential errors related to syntax.

Lexis	**Number**	**%**
Wrong word (Ww)	N=102	70.34
Word choice (Wc)	N=43	29.65

Table 7. Summary of potential errors related to lexis.

Morphology	**Number**	**%**
Verb form (Vf)	N=8	44.44
Plural / singular (P/S)	N=5	27.77
Word form (Wf)	N=3	16.66
Tense (T)	N=2	11.11

Table 8. Summary of potential errors related to lexis.

Overall, the data on predicted errors suggest that wrong word (L) followed by sentence fragments (S) and word choice errors (L) are the most common inaccuracies. In contrast, tense-related errors in the morphology category appear to be the least common.

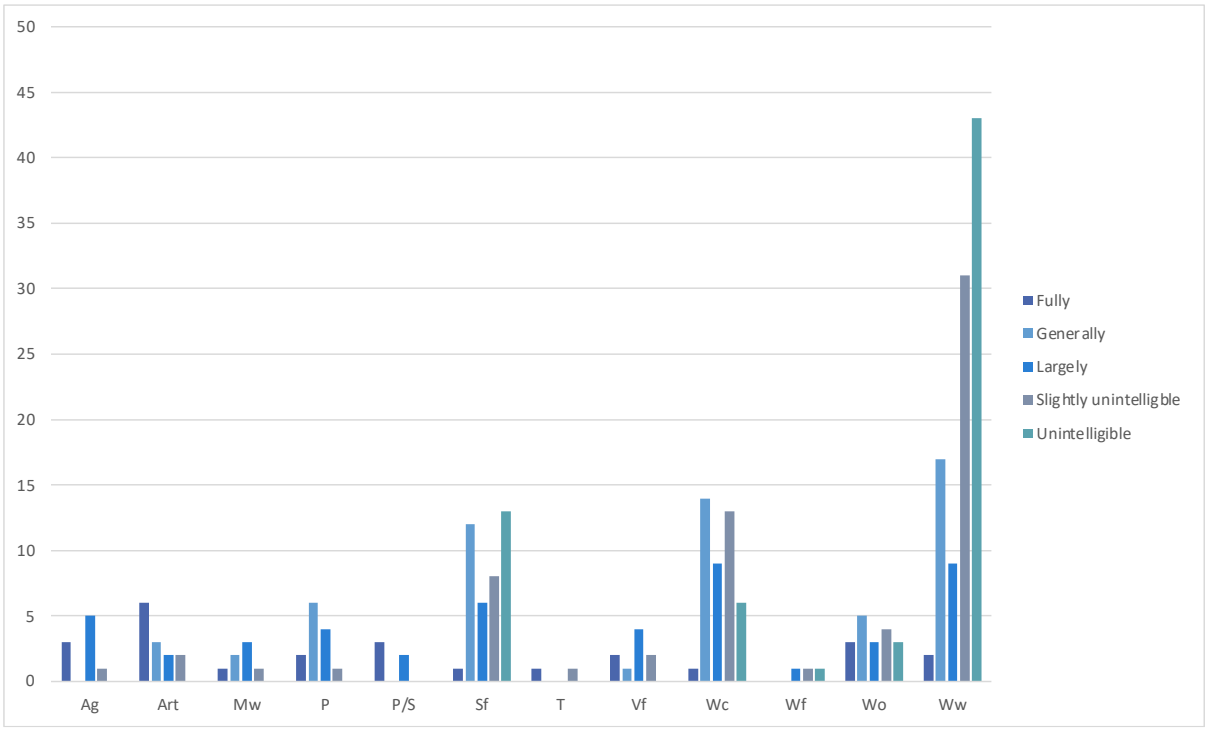

Figure 3. Potential errors related to intelligibility.

The graph above categorises potential errors by their type and the degree of intelligibility they are associated with (e.g., 'Fully', 'Generally', 'Largely', 'Reasonably', 'Slightly unintelligible', 'Unintelligible'). The data shows that 'Wrong Word' (Ww) errors are the most frequent, with a total of 102 occurrences. Furthermore, most of these errors are either ‘Unintelligible’ n=43, or 'Slightly Unintelligible' n=39. 'Sentence Fragment' (Sf) errors are the next most common, with a total of 41 occurrences. Intelligibility varies, although it is notable that n=21 are either ‘Unintelligible’ n=43, or 'Slightly Unintelligible'. With the exception of ‘Sentence fragments’, a general trend suggests that errors related to syntax and morphology are mostly intelligible and do not cause issues for the reader. For example, error types, such as 'Preposition' and 'Plural/Singular' errors, occur less frequently and are associated with higher levels of intelligibility. On the other hand, 'Wrong Word' is significantly more common and tend to be associated with lower levels of intelligibility. Perhaps unsurprisingly, it can be suggested that semantic errors related to meaning appear to be more disruptive to reader comprehension than those related to function.

5.2. Experiment 2: Comparative Analysis- ASR tolerance

During the second experiment audio recordings were reviewed, and human transcriptions (HT) were created. Firstly, this enables the calculation of Word Error Rates (WER). Secondly, comparisons between the potential errors in the previous experiment and the reference corpus can give insights regarding the tolerance of ASR systems in respect to non-native variation. While not conclusive, the extent to which speaker utterances are intelligible to humans but not ASR systems can be analysed. This section refers to RQ1 and RQ4.

Results:

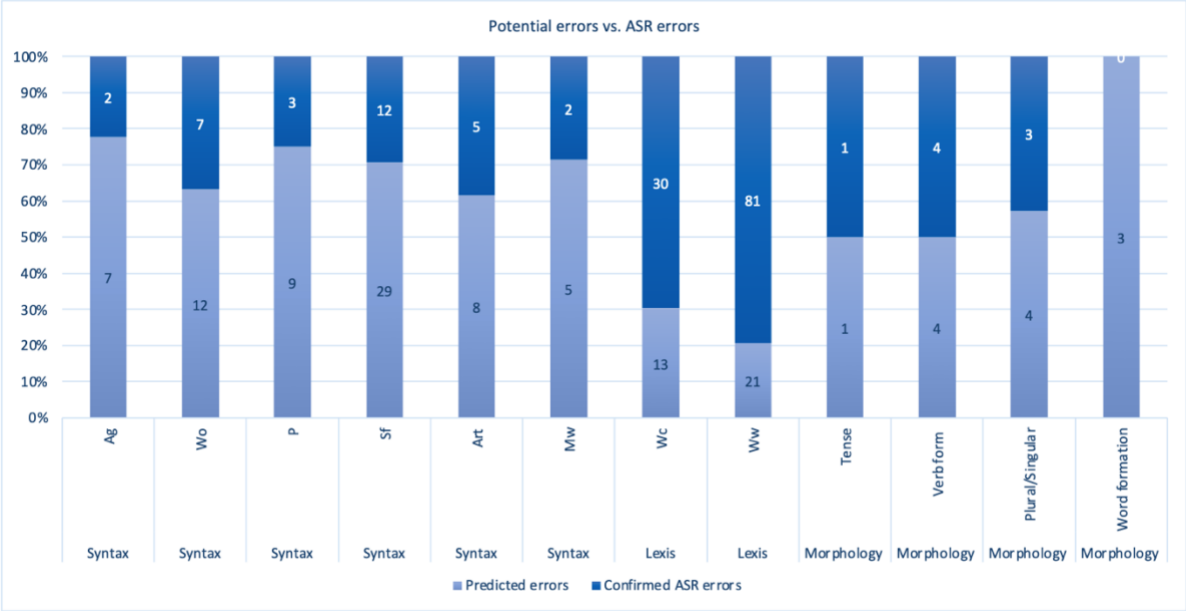

Figure 4. Potential errors vs. confirmed ASR errors

Error	Total error	Speaker	ASR
Agreement	9	7	2
Word order	19	12	7
Preposition	12	9	3
Sentence fragment	41	29	12
Article	13	8	5
Missing word	7	5	2

Table 9. Syntax tolerance

Based on the data above (Figure 4), the ASR system seems to exhibit greater tolerance to 'Syntax' and 'Morphology' errors than 'Lexis' errors. For example, in the 'Lexis' category, 30/43

instances of ‘Word choice’ and 81/102 instances of ‘Wrong word’ were due to inaccuracies in the ASR system. Conversely, in the 'Syntax' category, 29/41 instances were the result of speaker utterances. These were accurately transcribed by the ASR system despite the non-standard syntax (as observed in the audio recordings).

Error	Total error	Speaker	ASR
Word choice	43	13	30
Wrong Word	102	21	81

Table 10. Lexical tolerance

Error	Total error	Speaker	ASR
Tense	3	2	1
Verb form	8	4	4
Plural/Singular	5	2	3
Word formation	3	3	0

Table 11. Morphological tolerance

Overall, in response to RQ1, the high tolerance for 'Syntax' and 'Morphology' variations suggests that the ASR system is particularly adept at handling these types of language variations. The lower tolerance for 'Lexis' variations, particularly 'Wrong Word' errors, suggests that the ASR system may need refinement in this area to better handle lexical variations. However, this may also be due to other factors and causes (See experiment 4).

5.3. Experiment 3: WER Analysis

The next section will outline the findings based on the entire number of Word Error Rate (WER). During this process, the hypothesised ASR transcript and HT reference transcript were compared, and the number of insertions, deletions, or substitutions was recorded. Accordingly, a separate corpus was created so as to better represent the total amount of confirmed ASR inaccuracies. In this corpus, n=383 individual errors are recorded from a total of n=246 utterances. Utterances are considered as uninterrupted stretches of spoken language preceded by silence.

WER	Number	%
Deletions	N=93	24.28
Substitutions	N=247	64.49
Insertions	N=43	11.22
Total	N=383	

Table 12. Number of total WER instances

The majority of the errors were substitutions (64.49%), followed by deletions (24.28%), and then insertions (11.22%). This distribution of errors seems to imply that the system tends to substitute words or phrases more frequently than inserting or deleting words.

5.3.1. WER score per speaker

The WER score is calculated by adding the total number of insertions, substitutions, and deletions from the reference transcript and dividing this by total number of words per person.

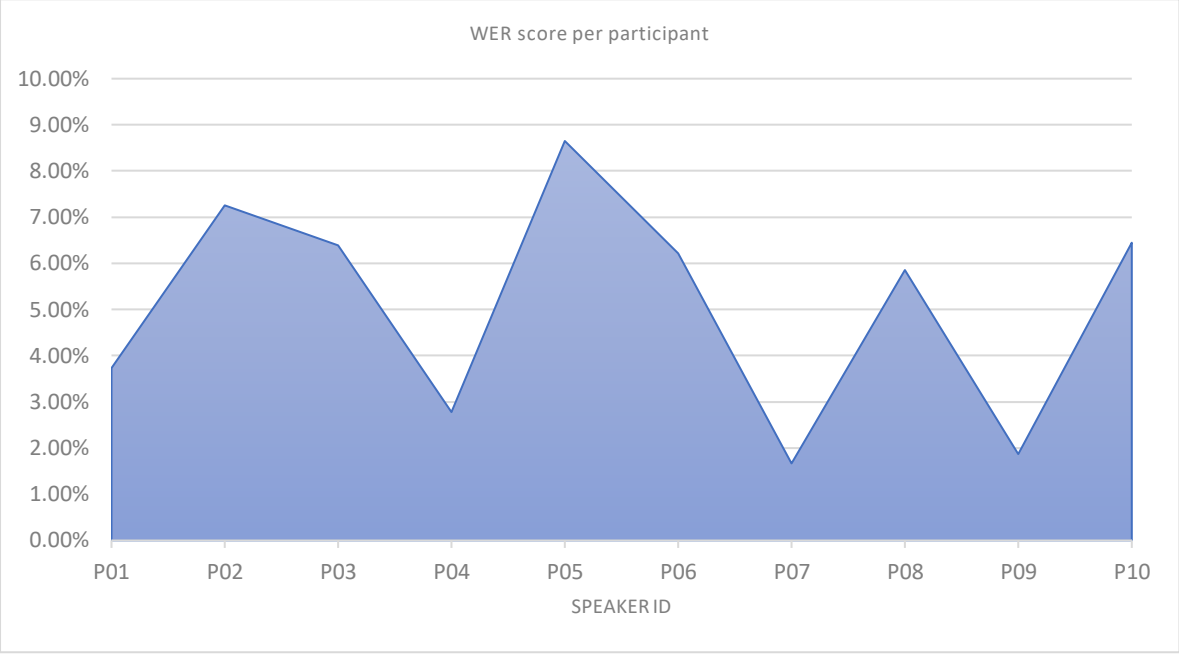

Figure 5. WER score per speaker

The provided data shows the Word Error Rate (WER) scores for 10 different speakers, identified as P01 to P10. WER scores range from 1.66% to 8.66%. Speaker P07 has the lowest WER score of 1.66%, indicating the highest level of accuracy in ASR, while speaker P05 has the highest WER score of 8.66%, indicating the lowest level of accuracy.

Overall, the WER scores appear to be relatively low. In the context of ASR systems, a score between 5-10% is of satisfactory quality and industry approved (Microsoft, 2023). Since the highest WER score among the participants observed is 8.66%, and most are below 7%, these scores can be considered quite good, especially for non-native speakers that display English variation.

5.3.2. WER in relation to sociolinguistic factors

Part of this investigation aims to better understand the relationship between ASR accuracy and sociolinguistic factors. This directly responds to research RQ2 of this study.

Age and ASR Performance:

Age	Number	Speaker ID	WER score	Average
18-24	1	P04	2.78%	2.78%
25-34	4	P02 P05 P08 P09	7.26% 8.65% 5.85% 1.87%	5.90%
45-54	5	P01 P03 P06 P07	3.73% 6.39% 6.21% 1.66%	4.88%

Table 13. Age and WER score

The data provided displays the WER scores of participants from three different age groups:

- **Age Group 18-24:** There is only one participant in this age group, P04, with a WER score of 2.78%.

- **Age Group 25-34:** There are four participants in this age group: P02, P05, P08, and P09. Their WER scores are 7.26%, 8.65%, 5.85%, and 1.87%, respectively. The average WER score for this age group is 5.90%.

- **Age Group 45-54:** There are five participants in this age group: P01, P03, P06, P07, and P10. Their WER scores are 3.73%, 6.39%, 6.21%, 1.66%, and 6.45%, respectively. The average WER score for this age group is 4.88%.

The 18-24 age group has the lowest average WER score (2.781%), but it only includes one participant, thus it is not representative. The 25-34 age group has the highest average WER score (5.90%) and the most variation in scores, ranging from 1.87% to 8.65%. The 45-54 age group has an average WER score of 4.88%, with scores ranging from 1.66% to 6.45%.

It is notable that the 25-34 age group, despite having a participant with one of the lowest WER scores (1.877%), has the highest average WER score. This disparity within the same age group suggests that individual linguistic nuances or other factors may play a more significant role than age. In a similar vein, the 45-54 age group had varying WER scores, from as low as 1.66% to as high as 6.45%. As with the previous group, perhaps this indicates that age is not a predominant factor affecting ASR accuracy.

Gender and ASR Performance:

The provided data represents the average Word Error Rate (WER) for male (n=8) and female (n=2) participants.

Gender	Number	Average WER
Male	8	5.04%
Female	2	5.26%

Table 14. Gender and WER score.

As can be seen, the average WER for both genders is quite similar, with a slightly higher average error rate for the female participants (5.26%) compared to the male participants (5.04%). However, this difference is minimal and given the small sample size it is difficult to

draw any definitive conclusions from this difference. In both groups WER scores varied widely. For example, among the males scores ranged from 1.66% - 7.26%, while among the females, scores ranged from 1.87% - 8.65%. While there is a slight difference in the average WER scores between males and females, the variation within each gender and age group is more significant than the variation between the groups. This suggests that age and gender are not strong predictors of ASR performance, and individual variations, such as linguistic background and English proficiency, may have a more significant impact.

Additional factors:

Factor	Result
Role in company	CEOs n=7, CFO n=1, CMO n=1 and 'Other' n=1.
English proficiency	B2 n=4, C1 n=6
Use of English at work	All participants use English at work.
Interactions with natives	Yes, n=2. No, n=8.
Previous pitch practice	Yes, n=9, No, n=1

Table 15. Other factors

English proficiency:

The speakers have either a B2 or C1 level of English proficiency. There are 4 speakers with a B2 level and 6 speakers with a C1 level. Interestingly, the two lowest WER scores are from speakers who have a C1 level of English proficiency. The highest WER score is from a speaker who has a B2 level of English proficiency. The average WER score for the B2 group (6.17%), is higher than the average WER for the C1 group (4.36%). This suggests that, on average, the C1 group has a lower WER score than the B2 group, which could indicate that a higher English proficiency level is associated with a lower WER score.

Moreover, cross-references between age, proficiency and WER score also reveal another trend. For example, there are differences between B2 and C1 speakers in age group 25-34 and

45-54. The difference in average WER score between B2 and C1 speakers in the 25-34 age group is 2.68%. Whereas the difference in average WER score between B2 and C1 speakers in the 45-54 age group is 0.59%. As a result, there appears to be a larger difference between the average WER of B2 and C1 speakers in the 25-34 age group. This could be interpreted as suggesting that in the younger age group, English proficiency level has a more pronounced impact on WER score compared to the older age group.

Hours of English Usage at Work:

The number of hours that speakers use English at work varies from 1 to 10 hours. One speaker who has a B2 level of English proficiency, uses English the most at work (10 hours) but has a relatively high WER score of 6.45%. On the other hand, another speaker who uses English for 5 hours at work and has a C1 level of English proficiency, has one of the lowest WER scores (1.87%). Again, this appears to support proficiency rather than the frequency in use of English plays a more pivotal role.

Hours Speaking to English Native Speakers:

The following data refers to the number of hours per week that are spent speaking to native speakers as opposed to non-natives. P02, who speaks to native English speakers the most 5/5 hours per week, has a comparatively higher WER score of 7.26%. Conversely, P07 who speaks to native English speakers for 2 hours, has the lowest WER score of 1.66%. Initially, there is no clear link between the number of hours spent speaking to native English speakers and WER score. However, a higher average ratio of time spent speaking with natives appears to be associated with a slightly lower average WER score. Although, the difference in the WER scores (4.76% vs 5.28%) is not very large, suggesting that while speaking with natives may have some impact on WER score, it is not the only significant factor.

Summary of WER findings:

In response to RQ2, the following findings are summarised. The data suggests that age is not a predominant factor affecting ASR accuracy since there is a significant variation in WER scores within each age group.

The average WER for male participants is 5.04%, and for female participants, it is 5.26%. Gender, based on this dataset, does not show a strong influence on WER scores. The slight difference in the average WER scores between males and females and the significant variation within each gender group suggests that gender is not a strong predictor of ASR performance.

The average WER score for the B2 group is 6.17%, while for the C1 group, it is 4.36%. This suggests a relationship between English proficiency and WER score, with higher proficiency associated with a lower (better) WER score. While proficiency (B2 vs. C1) is a clear differentiator in WER scores across age groups, age itself seems to also play a role, though in a subtler manner. For instance, the younger age group (25-34) shows a more noticeable gap between B2 and C1 WER averages than the older age group (45-54). This seems to have a more pronounced impact of English proficiency on WER score in the younger age group. However, it could be speculated that the older group (45-54), regardless of level, might have more experience in speaking English in specific contexts, leading to slightly better ASR accuracy than their younger B2 counterparts. One caveat is that the data is open to subjectivity as the question requires participants to provide their level and this could be based on their own perception not qualification and lacks reliability.

Initially, there is no clear link between the number of hours spent speaking to native English speakers and WER score. However, a slight correlation is revealed when considering the average ratio of interaction with natives vs. non-natives. A higher average ratio of time speaking with natives is associated with a slightly lower average WER score, but the difference in WER scores (4.76% vs 5.28%) is not very large, suggesting that speaking with natives may have some impact, but it is not the only significant factor.

5.4. Experiment 4: Qualitative Analysis of Causes

Error Category

To achieve a more nuanced understanding of the elements that impact WER scores in the context of Spanish startup founders, a qualitative analysis of speaker utterances and ASR inaccuracies was conducted. This section relates to RQ4.

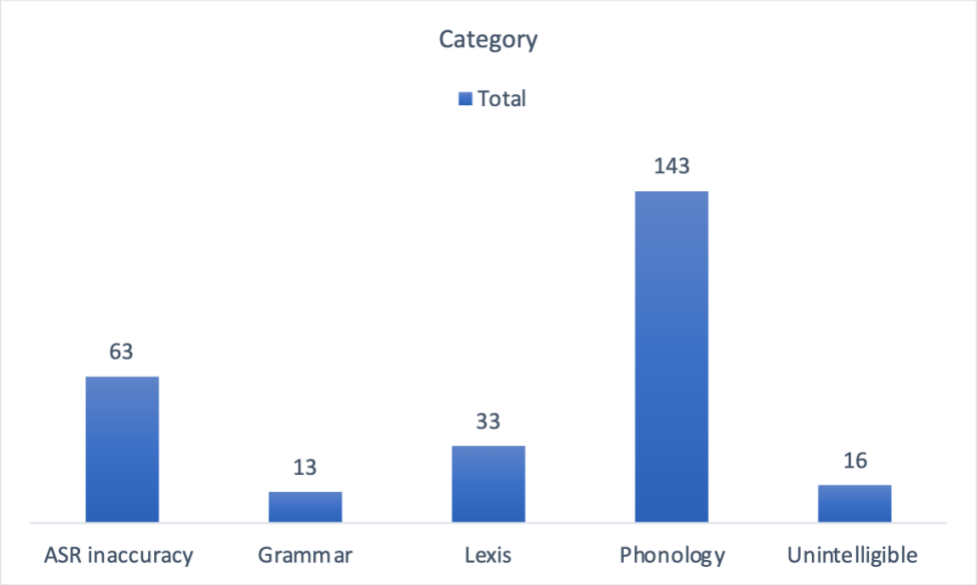

Figure 6. Breakdown of causes

From the initial 383 WER errors, a total of 268 utterances were tagged as containing errors. The main causes were labelled according to five different sub-categories: ASR inaccuracy, Grammar, Lexis, Phonology, or Unintelligible. The latter referring to instances where a speaker utterance was not intelligible enough to manually transcribe, thus these errors are classified as pertaining to the speaker. Moreover, utterances were scored for intelligibility.

Results:

Phonology is the largest category with 143 errors, followed by ASR Inaccuracy with 63 errors. Regarding Grammar, there are 13 errors in this category while Lexis accounts for 33 errors. Lastly, Unintelligibility is the smallest category, with 16 errors.

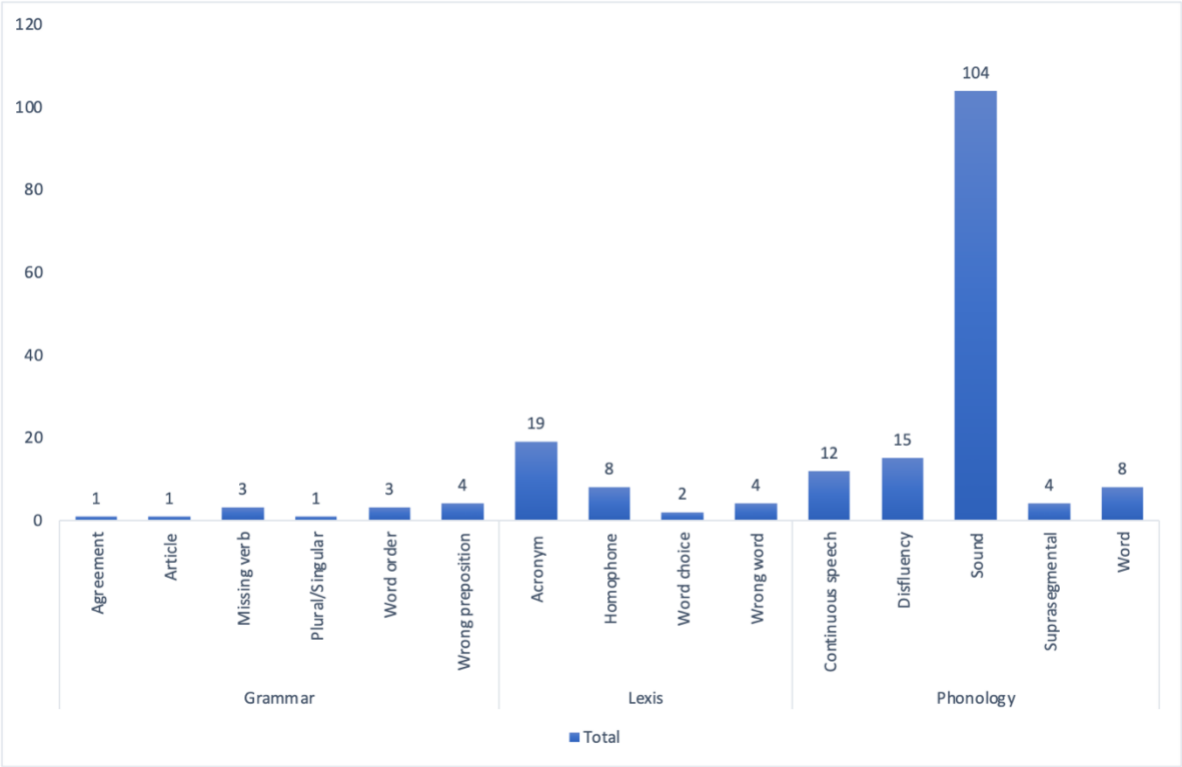

Figure 7. Sub-categories of causes

The data above represents the count of different types of language errors categorises into three main categories: Grammar, Lexis, and Phonology. Grammar errors relate to wrong preposition, agreement, missing preposition, word order, missing verb or plural/singular etc. Lexical errors include word choice, wrong word, acronym, or homophone. Phonological errors are categorised at the level of occurrence e.g., phoneme (sound) word or sentence (continuous speech). In addition, prosodic features such as intonation patterns and disfluencies such as hesitation or circumlocution were also observed.

Grammar: This category has a total of n=13 errors. As a result, it is the category with the least number of errors of the three categories. The most common error within this category is Wrong preposition with n=4 occurrences, followed by Missing verb and Word order with n=3 occurrences each. Agreement, Article, and Plural/Singular each have only n=1 occurrence, making them the least common types of grammar errors in the data.

Lexis: This category has a total of n=33 errors. Acronym is the most common error type in this category, with n=19 occurrences, and this equates to more than half of the total errors in this

category. Homophone errors are the second most common error in this category with n=8 occurrences, followed by Wrong word n=4 and Word choice n=2.

Phonology: This category includes n=143 errors which is significantly higher than the other categories. As can be seen from the data above, 'sound' accounts for the most common sub-category of errors with a total 104 instances. This represents nearly 72.72% of the phonology errors and 55.02% of the total errors across all categories. This indicates that errors related to phonemes are by far the most common type of error in the data. Disfluency is the second most common error type in this category with n=15 occurrences, followed by Continuous speech (n=12 occurrences), Word (n=8 occurrences), and Suprasegmental (n=4 occurrences).

Overall, with a grand total of 189 errors, the most common category of errors is Phonology (n=143 occurrences), followed by Lexis (n=33 occurrences), and then Grammar (n=13 occurrences). This data suggests that ASR systems struggle most with phonological variations, specifically related to phonemes. Moreover, lexical errors, specifically with acronyms and homophones, also seem to stretch the capabilities of the system. Grammar errors were the least common among the three categories, with wrong prepositions, missing verbs, and word order being the most frequent but they do not appear to be the main cause of ASR inaccuracies. The following section will explore the findings in each area below:

5.4.1. Grammar

The following section will discuss aspects related to the type of grammatical errors and occurrences of WER observed in the investigation.

Word order is a relatively infrequent occurrence given the suggested ASR tolerance for non-standard syntax.

WER type	Cause	Hypothesised transcript	Reference transcript
Deletion	Word order	[ ] our concrete.	How is evolving our concrete.

Table 16. Speaker ID P08 – Error 308 - ASR199

Agreement did not normally impede comprehension, however in the instance below this was combined with a speaker utterance that was scored as slightly unintelligible.

WER type	Cause	Hypothesised transcript	Reference transcript
Insertion	Agreement	So, what would make acid?	So, what make us different?
Deletion			
Substitution			

Table 17. Speaker ID P03 – Error C11 – Utterance ASR073

There are two instances of the missing verb to be. Omissions such as these resulted in changes to syntax or a loss of meaning.

WER type	Cause	Hypothesised transcript	Reference transcript
Substitution	Missing verb to be	they receive in around 10,000 emails per month.	they receiving around 10,000 emails per month.
Insertion			
Substitution	Missing verb to be	burns in the eight piece due to…	borns in the 80s due to...
Substitution			
Insertion			

Table 18. Speaker ID P08 – Error C275 – Utterance ASR187 / Speaker ID P01 – Error C0002 – Utterance ASR002

Surprisingly, the most common error observed was 'wrong preposition'. Normally, as function words, prepositions do not carry meaning and rarely impede human comprehension when used inaccurately. However, the inclusion of the wrong preposition often resulted in semantic changes (due to substitutions) which impede understanding of the overall, ASR transcription.

WER type	Cause	Hypothesised transcript	Reference transcript
Substitution	Wrong preposition	the Mango Sisters	the Man goes to the stairs.
Insertion			

Table 19. Speaker ID P02 – Error C058 – Utterance ASR044

As with the previous example, non-standard grammatical output was accompanied by slightly unintelligible pronunciation, suggesting that a combination of these factors almost always leads to inaccuracies in ASR transcription.

However, some instances of incorrect preposition use or missing articles led to ASR inaccuracies despite the output being fully intelligible.

WER type	Cause	Hypothesised transcript	Reference transcript
Substitution Deletion	Wrong preposition	We are now more focusing Artificial intelligence.	We are now more focused in Artificial Intelligence.

Table 20. Speaker ID P01 - Error C002 – Utterance ASR002

WER type	Cause	Hypothesised transcript	Reference transcript
Substitution	Missing article	Basically, we are cast generator.	Basically, we are cash generator.

Table 21. Speaker ID P06 – Error C236 – Utterance ASR161

In the first example, the non-standard verb + preposition collocation of ‘in’ rather than ‘on’ appears to lead to confusion and a consequent morphological change to the base form of the verb. It could also be argued that features of connected speech, namely catenation between the consonant sound /d/ and the vowel sound /I/ may have caused the ASR system to perceive this as one word. In the second example, the lack of an article causes a substitution. The word ‘cast’ acts as both a noun and a verb. If it is considered as being a verb, then perhaps this is an indication of that the ASR system’s predictive model. Therefore, this might suggest that the word preceding a potential error might be influential in determining the outcome given that ASR systems uses predictive modelling.

Overall, the trends suggest that 'Wrong Preposition' errors are the most varied and complex, being associated with all WER strategies and a range of intelligibility levels. In contrast, 'Missing Verb to Be' and 'Word Order' errors are more consistent, typically resulting in deletions or substitutions and 'Generally' intelligible utterances. However, insertion errors seem to have a more significant impact on intelligibility compared to substitutions or deletions, as they are associated with lower levels of intelligibility ('Slightly Unintelligible' and 'Largely' intelligible). Based on the information above, Google Meet generally exhibits a high tolerance for non-standard grammatical forms. However, the analysis reveals that slight variations, such as wrong prepositions or missing articles, can lead to inaccuracies in ASR

transcription, especially when combined with utterances that display lower levels of intelligibility.

5.4.2. Lexis

Regarding lexical errors, this study observed issues related to word choice, wrong word, acronyms, and homophones.

Lexis	Number
Word choice	2
Wrong word	4
Acronym	19
Homophone	8
Total	**33**

Table 22. Breakdown of lexical errors

Word choice and wrong word

Word choice refers to instances where the speaker chooses a word or item which while not incorrect is not fully appropriate or frequent. Wrong word however refers to situations where the choice does not appear to be accurate, displays L1 transfer or does not exist.

WER type	Cause	Hypothesised transcript	Reference transcript
Substitution	Word choice	reassign the returns	reassign the turns
Substitution	Wrong word	And AGGREGORY in sports science.	And a degree in sports science.
Deletion			
Deletion	Wrong word	We cannot plannificate.	We cannot plannificate well.
Substitution	Wrong word	The Erica, I'm putting the parties with the technology.	Theoretical, I'm putting the parameters with the technology.
Substitution			

Table 23. Word choice and wrong word

Substitution was commonly employed by the ASR system as a strategy when dealing with unrecognisable or less frequent words. In the first example, the speaker uses the word ‘turns’.

While not inaccurate perhaps an alternative item related to the domain of work would have been more appropriate such as ‘shift’. In the second example, the speaker produces an utterance which is understandable, but placement is not accurate. Generally speaking, substitutions are expected, especially in the case of ‘wrong words’ given that they are unlikely to be included in the training data and therefore outside of the pronunciation dictionary of the system. Nevertheless, the system did manage to accurately transcribe one example of L1 transfer ‘*plannificate’* as seen in the example. However, this was followed by a deletion of the word ‘well’. This raises the question as to whether examples such as this impact the capability of the system to accurately predict the following words or n-gram. Finally, all of these examples were assigned an intelligibility score of either slightly unintelligible or unintelligible. As with grammatical errors, perhaps a combination of both inappropriate vocabulary choices and poor intelligibility can lead to errors.

Acronyms

Like all sectors and industries, the startup ecosystem uses domain specific terminology and acronyms to refer to key concepts and ideas. Given the innovative nature of this sector, there are lots of emergent neologisms and in turn anglicisms which are adopted by the Spanish startup community. The use of specialist terminology helps to demonstrate credibility and authority. Arguably, this is true during a pitch, where investors, who are well-versed in this terminology, assess not only the viability of the value proposition but also an entrepreneur’s knowledge and competence in the sector. Using domain-specific terminology accurately helps founders to show that they belong to this sector and understand its nuances.

However, the data highlights that the ASR system struggled with the transcription of unfamiliar acronyms commonly used in the startup sector. For instance, the acronym "SAAS" (Software as a Service) was inaccurately translated by the ASR system 6 times, making it the most common error in the dataset. Comparatively speaking, other specialist investment terms such as "SOM" (Share of Market), "TAM" (Total Addressable Market), "SAM" (Serviceable Available Market), were also frequently misinterpreted with n=3 occurrences each. "SME" (Small and Medium-sized Enterprises), "LATAM" (Latin America), "MRR" (Monthly Recurring

Revenue), “B2B” (Business to Business) were the least common errors, with 1 occurrence each. In total, there were n=19 inaccuracies related to startup acronyms.

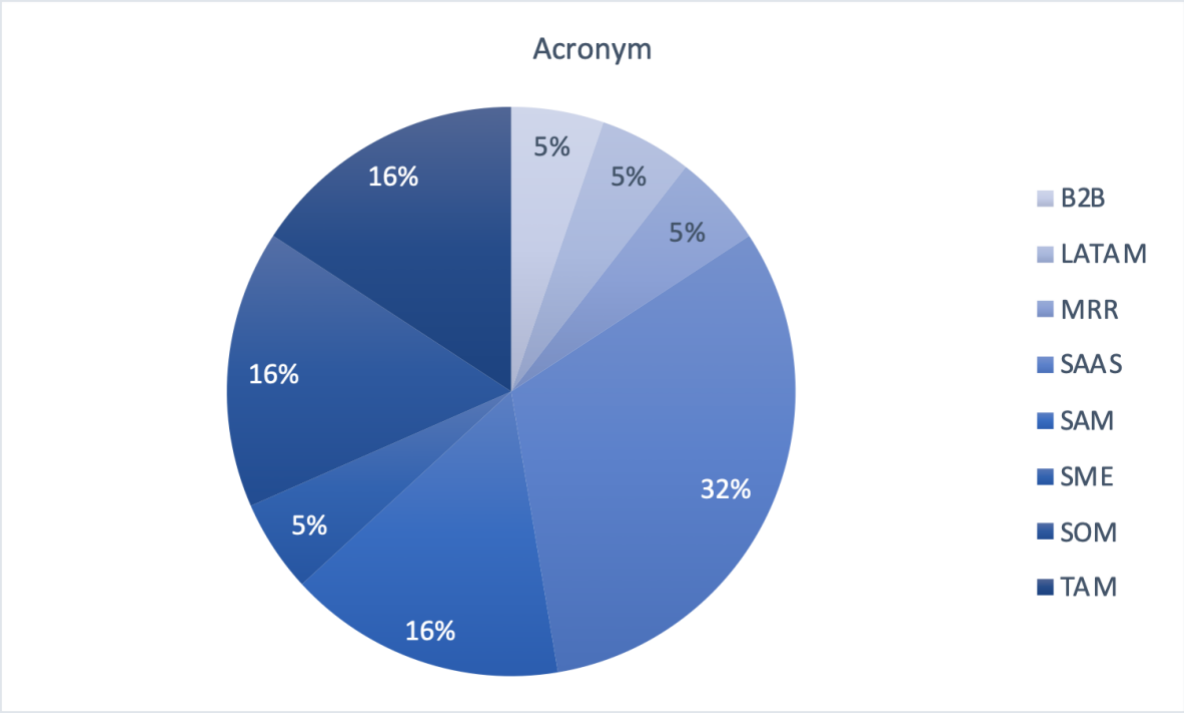

Figure 8. Errors relating to acronyms

Hypothesised transcript	Reference transcript
sell this technology through assess…	sell this technology through a SAAS…
our real-time sales, analyze the movement…	our real-time SAAS, analyse the movement…
have access for the review of videos…	have a SAAS for the review of videos…
In relation to the Sun…	In relation to the SAM…
our time refers to the three markets.	our TAM refers to the three markets.
as USB to be our customer represent…	as a SAAS B2B our customer represents…

Table 24. Example acronym errors

These results seem to support the literature that industry-specific terminology can pose a challenge to ASR systems. Owing to this, the use of acronyms should be considered especially in multilingual online contexts where users might use ASR tools for to enhance accessibility.

Homophones

Intelligibility was not the issue in the sample collected as all utterances were either largely or fully intelligible. In fact, the issue appears to be the opposite. Accurate pronunciation of words which had a homophone equivalent appear to confuse the ASR system. The extent to which these systems account for context is unclear and a lack of contextual cues may lead to erroneous choices when encountering different words with the same sound.

WER type	Cause	Hypothesised transcript	Reference transcript
Substitution	Homophone	the inland transport of your roots	the inland transport of your routes

Table 25. Speaker ID P03 – Error C126 – Utterance ASR081

WER type	Cause	Hypothesised transcript	Reference transcript
Substitution	Homophone	We hear we're talking about	We here we're talking about

Table 26. Speaker ID P09 – Error C339 – Utterance ASR218

Both the examples above refer to substitutions. The meaning does not appear to be overly distorted given that only one word has been substituted per utterance. Moreover, given the similarity in sound, it is unlikely to cause issues with comprehension. In the second example, the speaker's omission of the verb to be may have caused the system to predict a verb after the personal pronoun.

WER type	Cause	Hypothesised transcript	Reference transcript
Deletion	Homophone	along term are and record revenue.	a long term and recurrent revenue

Table 27. Speaker ID P06 – Error C231 – Utterance ASR158

WER type	Cause	Hypothesised transcript	Reference transcript
Deletion	Homophone		which are our customers?

Table 28. Speaker ID P03 – Error C087 – Utterance ASR160

Both examples above relate to deletions. The reason why the ASR system chose to delete part of these utterances is unclear. However, in the second example, perhaps the choice of relative clause 'which' rather than 'who' may have triggered the system to predict another structure. Moreover, the adjacency of the unigrams "are" and "our" may have caused the system to register this as hesitation or reformulation, thus leading to deletion. Overall, while leading to errors in transcripts the homophones observed in the data do not seem to have a significant impact on meaning.

5.4.3. Phonological

The following section will describe the possible causes of ASR transcription from the perspective of phonological features. There are 143 errors in total and the findings will be outlined below:

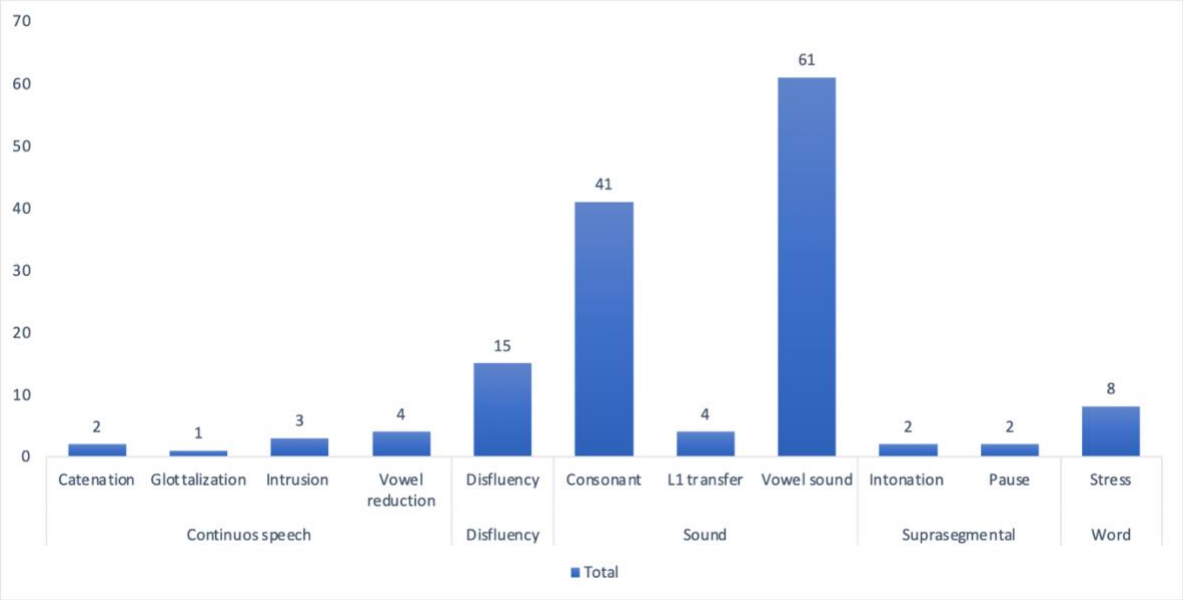

Figure 9. Breakdown of phonology errors

Continuous Speech

There are 10 instances where continuous speech contributed to ASR transcription inaccuracies. As mentioned in the literature, ASR systems can struggle to accurately transcribe words in a continuous flow of speech where word boundaries are not always clear.

Catenation: There are 2 instances where catenation contributed to ASR inaccuracies.

Cause	Hypothesised transcript	Reference transcript
Catenation	So, we're not a problem played.	So, we're not a plug and play.
Catenation	You will have to zoom in a name well.	You will have to zoom and aim well.

Table 29. Examples of catenation

Glottalization: There is 1 instance where glottalization contributed to ASR inaccuracies.

Cause	Hypothesised transcript	Reference transcript
Glottalization	Was is [company name]?	What is [company name]?

Table 30. Examples of glottalization

Intrusion: There are 3 instances where intrusion contributed to ASR inaccuracies. In the example, below the speaker joins two vowel sounds with the inclusion of /w/.

Cause	Hypothesised transcript	Reference transcript
Intrusion	tour customers on how to act	to our customers on how to act
Intrusion	they wanted to addressability	they wanted to add traceability

Table 31. Examples of intrusion

Vowel Reduction: There are 4 instances where vowel reduction contributed to ASR inaccuracies.

Cause	Hypothesised transcript	Reference transcript
Vowel reduction	Okay, so just comment last year…	Okay, so just as a comment last year.

Table 32. Examples of vowel reduction

Disfluency

There are 15 instances where disfluencies contributed to ASR inaccuracies. Disfluencies are interruptions in the flow of speech, such as hesitations or repetitions. 50% of the speakers (P05, P07, P09, P10) did not have any recorded disfluencies and all disfluencies recorded were from male participants.

WER type	Cause	Hypothesised transcript	Reference transcript
Deletion	Disfluency	[ ]	Involves like five companies already

Table 33. Examples of disfluencies

As for the relationship between Words Per Minute (WPM) and disfluencies, there is no clear trend in this dataset. For example, the speaker with the highest WPM (P07) had no disfluencies, while the speaker with the lowest WPM (P03) had 3 disfluencies. However, it is interesting to note that all speakers with zero disfluencies had a WPM greater than the average (116.15), which might suggest that a higher speaking rate is associated with fewer disfluencies.

Sound

Short vs. long vowel sounds

The data indicates that the most common phonological errors are related to vowel sounds.

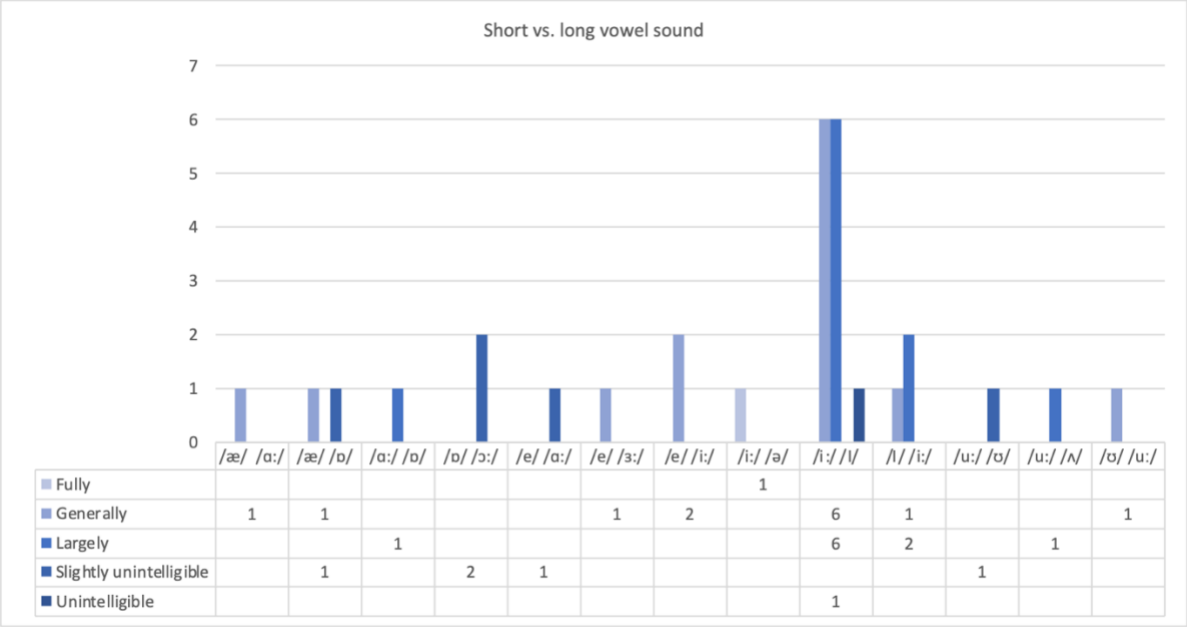

Figure 10. Breakdown of short vs. long vowel sound substitution

The data presents the frequency of specific phoneme variations (speaker output versus standard form) and their corresponding intelligibility scores. The most frequent phonemic

substitution occurred between "/iː/" and "/ɪ/", with a total of 16 instances. These were mostly 'Generally' or 'Largely' intelligible, with only one instance being 'Slightly unintelligible'. The next most common substitution was between "/e/" and "/iː/", with a total of 2 instances, all of which were 'Generally' intelligible. Most of the inaccuracies (13 out of 30) were classified as 'Generally' intelligible, followed by 'Largely' intelligible with 10 instances. One attempt was classified as 'Unintelligible'.

Overall, the data suggests that the ASR system experiences difficulties with differentiating between the "/iː/" and "/ɪ/" sounds in non-native speech. Interestingly, the majority of these confusions are still generally or largely intelligible. This indicates that while the ASR system does make errors in transcribing non-native speech, these errors are usually not severe enough to render the output unintelligible for human comprehension. Interestingly, the /ɒ/ /ɔː/ variation, with 2 occurrences, is rated as 'Slightly unintelligible', indicating a clear pattern of reduced intelligibility for this specific variation.

Moreover, of the five ‘slightly unintelligible’ utterances four are categorised as being short vs. long vowel sounds. For example, 9 of these where due to the production of ‘this’ vs. ‘these’. This suggests that while some variations in pronunciation by non-native speakers do not severely impact intelligibility, others consistently do. Therefore, although intelligible to human transcription, all the above led to errors in ASR outputs suggesting a lack of tolerance regarding an ability to discern variations in short and long vowel sounds.

Variation	Standard	Reference item	Hypothesised transcript
/iː/	/ɪ/	inbox	they have any box where they receive all the invoices.
/iː/	/ɪ/	Shipping lines	And our solution is for sleeping lines
/ɪ/	/iː/	these	you get this fruits and veggies
/ɒ/	/ɔː/	Haulage	What is container college?
/e/	/ɜː/	third	the fear is getaway
/æ/	/ɑː/	tasks	continue with the next task

Table 34. Breakdown of short vs. long vowel sound substitution

Vowel vs. diphthong

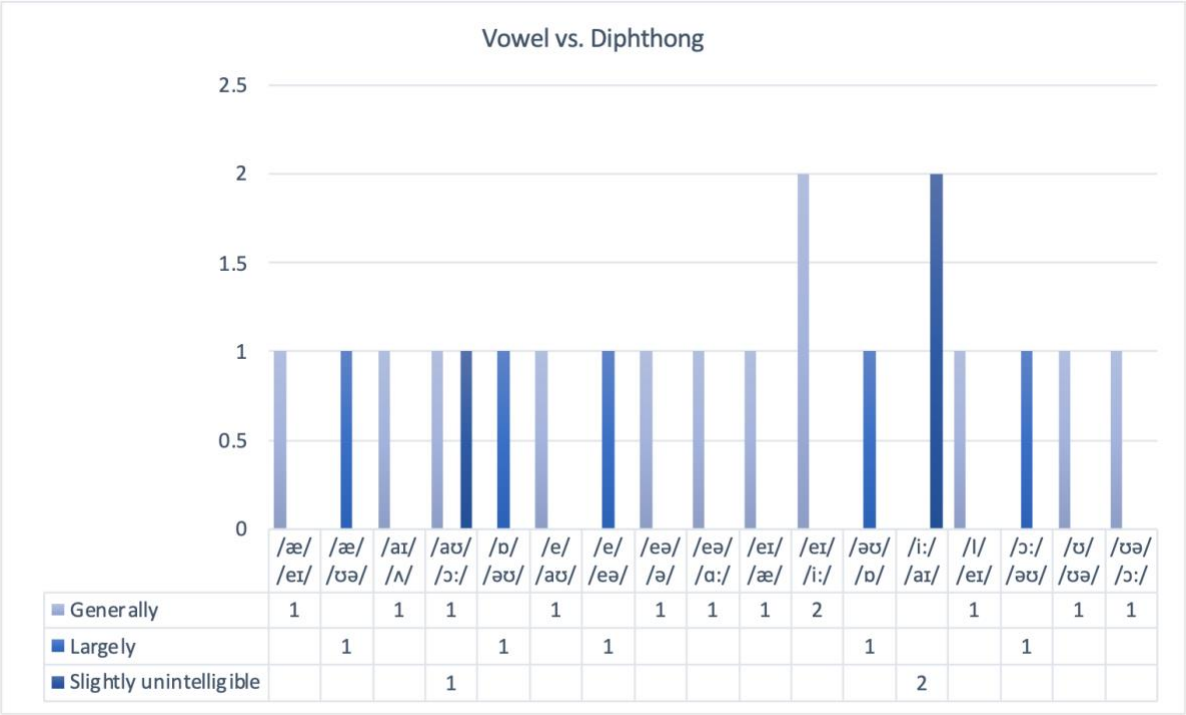

Figure 11. Breakdown of vowels vs. diphthong substitution

There are a total of 17 different combinations of vowel sounds in the data. Overall, the data demonstrates that there are a total of 20 instances of vowel sound substitutions.

Generally speaking, substitutions between these combinations did not appear to impede human comprehension as 17 out of the 20 examples were labelled 'Generally' or 'Largely' intelligible. However, some vowel sounds appear more frequently in the data than others, for example, /eɪ/ and /ɔ:/ appears in three different pairs. This provides further support that the long vowel sound /ɔ:/ is particularly difficult to articulate and thus misunderstood.

Wrong vowel sound or diphthong

There were relatively few examples in the data. Overall, the data indicates that there were six different pairs of vowel sounds. Vowels involving the schwa sound (/ə/) are the most common.

L1 transfer

Variation	Standard	Reference item	Hypothesised transcript
<r>	/r/	enter	The user **entered**, the fitness app
<es>	/s/	Scalability	focus on **A software**
<es>	/s/	stay	get better results and **a state** motivated.

Table 35. Examples of alveolar trill and epenthesis

There were four examples of L1 transfer, namely the alveolar trill n=1 and epenthesis n=3. The latter seems to have a greater impact on the output of the transcription. Epenthesis is the addition of one or two sounds before a word. As can be seen in both examples above, the system appears to register the speakers' release of air as a phoneme and transcribes this as the article 'a'. Consequently, the predictive model seems to substitute the remaining sound as a noun as this is syntactically correct e.g., article + noun. It should be noted that the epenthesis of /e/ before consonant clusters is a phonological feature of Spanish.

Consonant

There are 41 instances of consonant-related errors. This indicates that consonant pronunciation by non-native speakers often causes inaccuracies in ASR transcription.

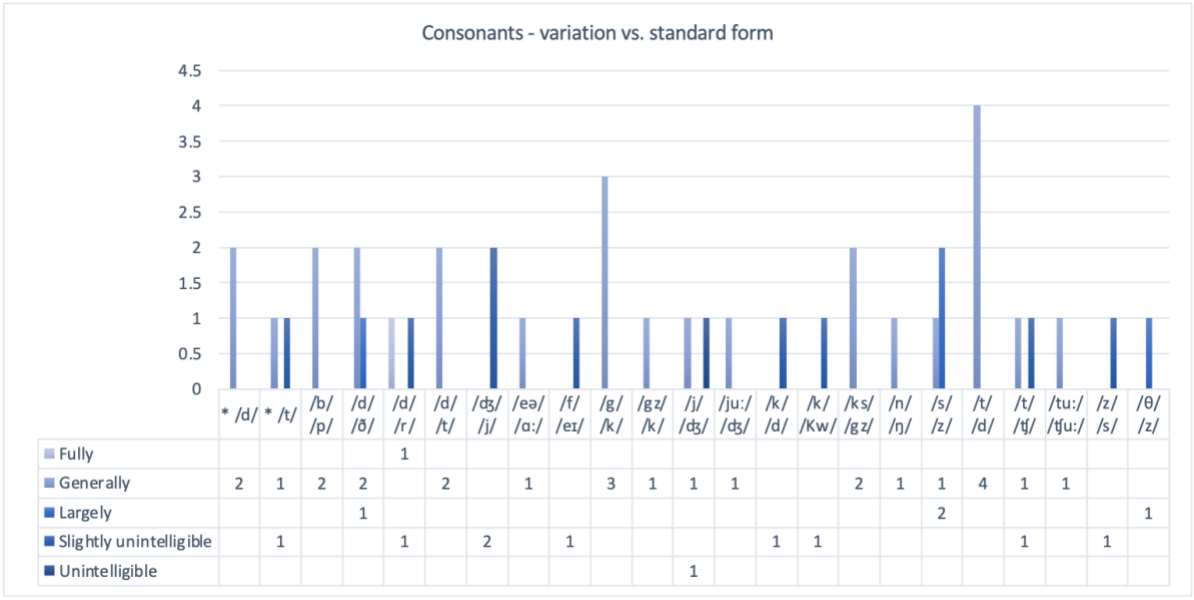

Figure 12. Breakdown of consonant substitution

The most common phonetic substitution is between /t/ for /d/, with a total count of n=6, all of which are classified as generally intelligible. Substitutions between /g/ for /k/ also account for a total count of n=6. Phonetic substitutions between /s/ for /z/ and /d/ and /ð/ have n=4 and n=3 errors respectively.

Most of the substitutions are classified as 'Generally' intelligible, with a total count of 26 out of 41. There is only 1 instance each of 'Fully' and 'Unintelligible' substitutions, which suggests that most of the substitutions, while not accurate, do not completely hinder the understanding of the speech. However, 10 errors are recorded as Slightly unintelligible, indicating that these substitutions can cause difficulty to both the listener and ASR systems.

Many of the substitutions involve phonemes that are articulatory or acoustically similar. For example, /s/ for /z/, /t/ for /d/, and /g/ for /k/. This suggests that speakers have difficulty in distinguishing between voiced and voiceless sounds and that ASR systems are less tolerant with these types of variations.

/t/ and /d/ phonemes accounts for a total of n=6 occurrences or 14.63% of total.

Variation	**Standard**	**Hypothesised transcript**	**Reference transcript**
/t/	/d/	Put ways of course in warehouses	food waste occurs in warehouses
/t/	/d/	And this is all from my site.	And this is all from my side.
/d/	/t/	I am more focus on	I am more focused on

Table 36. Example of /t/ vs. /d/ substitution

/k/ and /g/ phonemes account for a total of n=6 occurrences or 14.63% of total.

Variation	Standard	Hypothesised transcript	Reference transcript
/g/	/k/	not even the drug driver	not even the truck driver
/gz/	/ks/	GPS of the drugs	GPS of the trucks
/ks/	/gz/	Then there's the cost of the tax themselves	Then there's the cost of the tags themselves
/ks/	/gz/	RFID tax are another alternative	RFID tags are another alternative

Table 37. Examples of /g/ vs. /k/ substitution

/s/ and /z/ phonemes account for a total of n=4 occurrences or 9.75% of total.

Variation	Standard	Hypothesised transcript	Reference transcript
/z/	/s/	in this process of this document presentation.	in this process of document processing
/s/	/z/	with different feast	with different fees.
/z/	/s/	that I us competition, there's global.	that as competition turns global.

Table 38. Examples of /s/ vs. /z/ substitution

Substitution between /ʤ/ and /j/ also causes inaccuracies in the transcription and accounted for 7.31% of the errors observed.

Variation	Standard	Hypothesised transcript	Reference transcript
/j/	/ʤ/	private for customers	projects for customers
/ʤ/	/j/	We are. Maybe freak junk,	We are so, maybe so freak, young
/ʤ/	/j/	It's a law due to share your result	It also allow you to

Table 39. Examples of /ʤ/ vs. /j/

Variations between /d/ and /ð/ were observed as causing inaccuracies in the transcription and accounted for 7.31% of the errors observed.

Variation	**Standard**	**Hypothesised transcript**	**Reference transcript**
/d/	/ð/	Destruction brings us to a break-even point.	This traction brings us to a break-even point.
/d/	/ð/	Destroy is led by four senior professional	This project is led by four senior professional

Table 40. Examples of /d/ vs/ð/

Variation	**Standard**	**Hypothesised transcript**	**Reference transcript**
/b/	/p/	just to go through some symbols	just to go through some samples
/b/	/p/	we are not an obligation.	we are not an application

Table 41. Examples of /b/ vs/p/

While not as common some errors appear to be caused by similarities with the place of articulation. For example, both /b/ and /p/ are bilabials. Moreover, the manner of articulation

is the same as both are plosives which means the airflow is completely blocked prior to release.

Overall, the data suggests that while there are several phoneme substitutions or mispronunciations, these are generally intelligible and do not severely impact the understandability of the speaker from a listener perspective. However, it should be noted that this is due largely to these utterances being contextualised. It was observed that these variations or misplacements did cause numerous substitutions and seem to be exacerbated by phonemes that are similar in place and manner of articulation and have a voiced/unvoiced equivalent. Furthermore, WER substitutions accounted for 90.24% of the total errors. This highlights limited tolerance and a strong likelihood that the ASR output will demonstrate significant semantic changes and a loss of meaning as can be seen in the examples above.

Word

There are 8 instances of word-related errors, all related to stress. Compared with the overall error count, they provide a small representation of potential inaccuracy causes. Nevertheless, when observed they impacted the accuracy of the transcription.

Hypothesised transcript	Reference transcript
is our Serco	is our CEO
it is enormous only for fruits and pitch tables markets.	it is enormous only for fruits and vegetables markets.
a week to deliver the residence.	a week to deliver the results

Table 42. Inaccuracies caused by word stress.

ASR transcription errors as a result of potential L1 transfer and inaccurate word stress placement.

Suprasegmental:

There are 4 instances of suprasegmental errors, this category includes n=2 instances of intonation and n=2 instances of pause. Based on the number of occurrences they appear to have a limited impact on ASR inaccuracies on the whole.

Hypothesised transcript	Reference transcript
And why.	And why us?
in the justic sector	in the logistics sector

Table 43. Suprasegmental influence – pause

Hypothesised transcript	Reference transcript
were houses, they ask per year per month	warehouses pay us per year per month
were houses, they ask per year per month	warehouses pay us per year per month

Table 44. Suprasegmental influence - intonation pattern

Summary

In summary, the data shows that the ASR system has difficulties with various phoneme variations, leading to errors in transcription. However, most errors are classified as 'Generally' or 'Largely' intelligible, indicating that despite the variations in pronunciation by non-native speakers, the utterances are mostly understandable to human transcribers. However, ASR systems show a lack of tolerance in discerning variations in vowel sounds, particularly short and long vowels. Moreover, the ASR system struggles to differentiate between phonemes that are articulatory or acoustically similar, such as /iː/ and /ɪ/, /t/ and /d/, /g/ and /k/, etc. Furthermore, certain specific variations, like the /ɒ/ /ɔː/ pair, consistently result in reduced intelligibility.

Epenthesis and L1 transfer also cause transcription inaccuracies, although they are less common in the dataset. As a phonological feature in Spanish, epenthesis potentially poses challenges to ASR systems which can incorrectly recognise this sound as a phoneme resulting in incorrect transcription of utterances, usually an insertion error.

Consonant-related errors are frequent, with the most common substitutions involving phonemes that are similar in place and manner of articulation and have a voiced/unvoiced equivalent. Although most of these errors are generally intelligible, they lead to significant semantic changes and a loss of meaning in the ASR output. These include /t/ and /d/, /g/ and /k/, and /s/ and /z/, supporting previous research such as the Lingua Franca Core (Walker, 2010).

Lastly, word-related errors such as incorrect stress placement and suprasegmental such as non-standard intonation patterns although less frequent, have a notable impact on transcription accuracy.

Overall, the data suggests that while non-native speaker variations in pronunciation do not severely impact intelligibility for human transcribers, they lead to significant errors in ASR transcriptions. This highlights the limited tolerance of ASR systems to phoneme variations and the resultant semantic changes and loss of meaning in the transcriptions. Despite the WER scores being relatively low, it could be argued that there is a greater need for ASR systems to be trained with a more diverse range of phoneme variations and accents to improve their performance and accuracy in transcribing non-native speech in Global English context.

5.4.4. Final WER Considerations

Analysis of WER by Cause

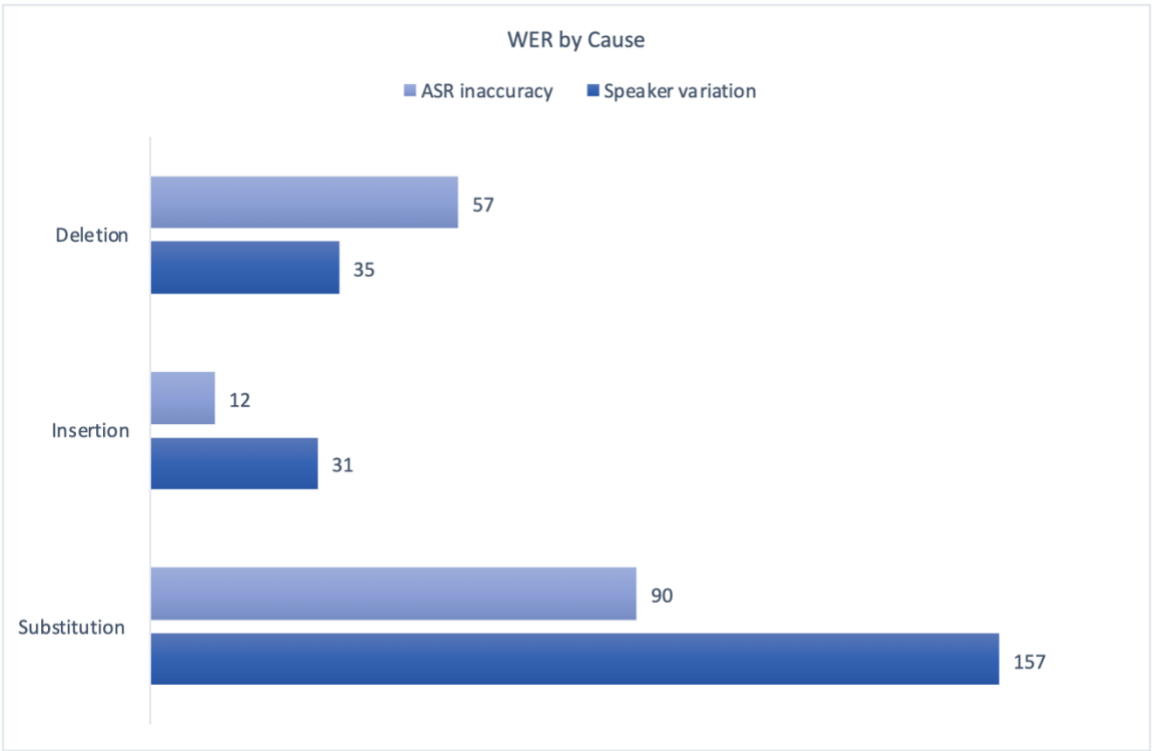

Figure 13. WER by Cause

The data above highlights the number of Word Error Rate (WER) scores based on the cause. Errors identified from the corpus have been divided into two categories: ASR inaccuracy and speaker variation.

- **Deletion:** There are a total of 92 deletion errors, out of which n=57 are attributed to ASR inaccuracy and n=35 to speaker variation.
- **Insertion:** There are a total of 43 insertion errors, out of which n=12 are attributed to ASR inaccuracy and n=31 to speaker variation.
- **Substitution:** There are a total of 247 substitution errors, out of which n=90 are attributed to ASR inaccuracy and n=157 to speaker variation.

The findings indicate that speaker variation is the most common cause of insertion and substitution errors. This is in line with the higher of number of errors related to sounds and phonological variation. As observed, the output of speaker variations related to phonology and lexis are typically registered by the ASR system but inaccurately transcribed. On the other hand, ASR inaccuracy is the most common cause of deletion. Interestingly, these utterances were observed as being either generally, largely or fully intelligible. As observed factors such as disfluencies, hesitations and pauses typically lead to deletions in the utterances that were otherwise intelligible.

WER by Intelligibility

Below is the different types of errors (deletion, insertion, and substitution) categorised into five levels of intelligibility: Fully, Generally, Largely, Slightly Unintelligible, and Unintelligible.

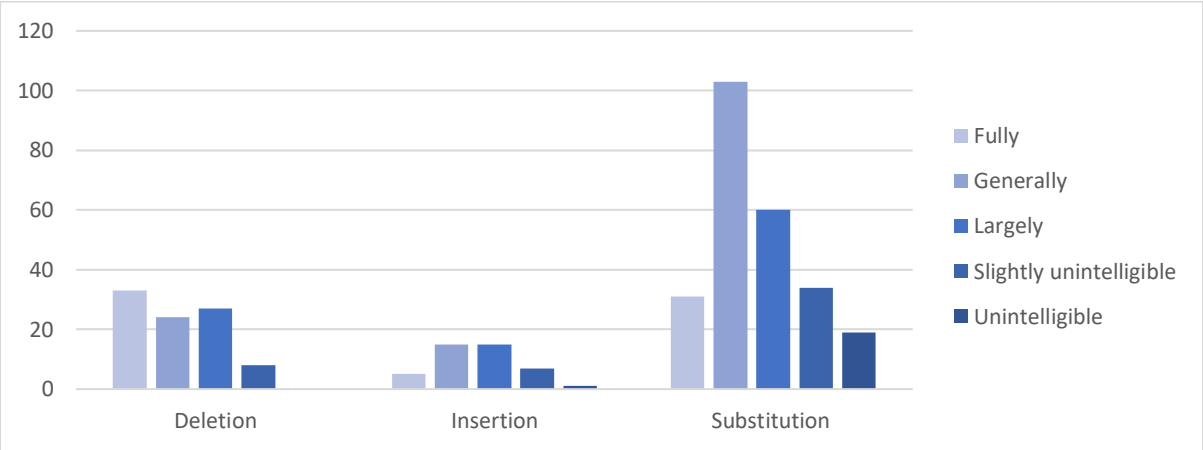

Figure 14. WER by Cause

Substitution errors are the most common with a total of n=247 substitution errors observed. Despite their frequency, most of these substitutions still result in generally intelligible content, indicating that while the ASR system misinterprets the speaker their utterances can still be discerned. There was a total of n=92 deletion errors. N=33 of these were categorised as 'Fully' intelligible, n=24 as 'Generally' intelligible, and n=27 as 'Largely' intelligible. There were n=8 instances categorised as 'Slightly Unintelligible' and no instances categorised as 'Unintelligible'. The variety suggests that even as content is omitted it remains understandable given the context. Although insertion errors are not as common, when they do occur, they do not appear to have a significant impact on meaning. The 'Generally' intelligible category contained the most substitution errors (103 out of 247), which is also the highest count among all of the categories. However, interestingly they also have the highest potential to distort meaning given that they substitute one word for another.

Intelligibility vs. Cause

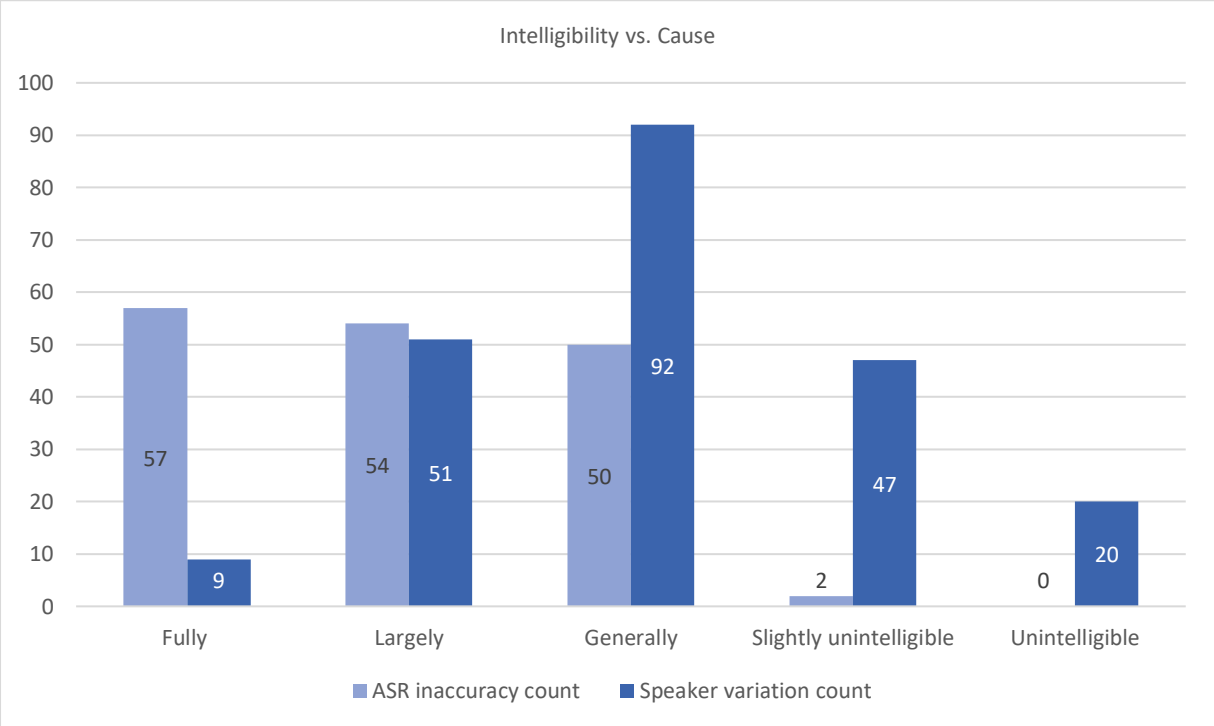

Figure 15. WER by Cause

In summary, the data suggests that ASR inaccuracies are a major cause of transcription errors when the speech is fully or largely intelligible. However, as the speech becomes less

intelligible, speaker variations become a more significant cause of transcription errors. This indicates that the ASR system struggles with both fully intelligible speech and non-native variations in the spoken output, but the impact of speaker variation becomes more significant as the speech becomes less intelligible.

6. Conclusions

Technological advancements and the recent accelerated adoption of video communication platforms have made Automated Speech Recognition software (ASR) omnipresent in digital working environments. English is becoming more important in the global startup scene and Automated Speech Recognition (ASR) plays a more prominent role in online communication. This investigation hypothesised that ASR systems are trained using monolingual models, thus not accurately representing the 'Global English' paradigm. This study aimed to investigate the effectiveness and potential biases of ASR tools in the specific context of Spanish startup founders delivering an online investment pitch in English. Accordingly, a summary of the findings will be outlined with reference to the research questions below:

RQ1. To what extent are ASR systems tolerant of variations in the oral output of non-native English speakers, particularly Spanish startup founders?

The first experiment evaluated the tolerance of ASR systems (Google Meet) when capturing non-native speaker variation. Based on the data presented, it is clear that these systems experience a range of errors when attempting to process utterances displaying divergence. The total errors amounted to 265 instances across three main error categories: Syntax, Lexis, and Morphology. Notably, lexical issues such as 'Wrong word' (n=102) and syntax errors such as 'Sentence fragments' (n=41) appeared to be the most challenging aspects for ASR transcription. When reviewing the data, it appears that sentence fragments were caused by the speaker's hesitations and disfluencies, leading to deletions by the system which could suggest that using simpler syntax can prevent these errors.

Omissions also have a significant impact on ASR transcriptions. For example, the verb "to be". This causes changes to syntax and is likely the result of the system's predictive modelling.

Similarly, this could be seen with omissions or the wrong choice of articles and prepositions. Interestingly, as function words, the incorrect use of prepositions and articles does not influence human comprehension in most contexts. In the case of ASR comprehension, though, this often resulted in inaccuracies, which may serve as further evidence regarding this impact on the predictive model.

However, some limitations were observed concerning morphological variations. The ASR system sometimes substitutes the wrong form, e.g., gerund vs. infinitive. It should also be highlighted that various factors should be considered when interpreting ASR efficacy, and these grammatical errors should not be analysed in isolation. For example, errors in ASR output tend to be compounded by the occurrence of variation in pronunciation. Overall, the system appears more tolerant of grammatical variations than lexical ones.

RQ2. How do the errors in ASR transcription correlate with sociolinguistic factors such as age, gender, and level of English proficiency among Spanish startup founders?

The Global English framework emphasises the diversity of language, sociolinguistics, and sociocultural aspects associated with the global use of English (Rose and Galloway 2019: 4). This investigation, therefore, aimed to discern how ASR transcription errors might be influenced by sociolinguistic factors, particularly age, gender, and English proficiency level, among Spanish startup founders.

Regarding age, it was not possible to draw concrete conclusions from the data, given the limited sample size and significant variations in scores within individual groups. An intriguing observation was that while the 25-34 age group included a participant with a low WER score (1.87%), this group had the highest average WER score overall. This suggests that factors beyond age, potentially individual linguistic idiosyncrasies, might play a more significant role in ASR performance.

In terms of gender, the average WER scores for both males and females were ‘satisfactory’ according to ASR industry standards. The data did not reveal a considerable gender-based performance variation in the ASR system. Therefore, to make definitive statements about the

role of gender on ASR accuracy, a larger, more varied sample would be required. However, the observed variations in WER scores within each gender and age group were more pronounced than the differences between these groups, indicating that individual differences might have a more impactful role than age or gender.

However, a discernible pattern emerged when comparing English proficiency levels. The B2 proficiency group had an average WER score of 6.17%, whereas the C1 group averaged 4.36%. This implies that as proficiency improves, the WER score decreases. While proficiency did not affect WER scores across ages, a combination of age and proficiency revealed a pattern. For instance, the gap between B2 and C1 WER averages was more prominent in the younger 25-34 age group compared to the older 45-54 age bracket. This could indicate a stronger influence of English proficiency on WER scores among younger individuals. Furthermore, older participants, even those at a B2 level, might have more experience using English in specific situations, such as delivering their pitch or networking events. However, this interpretation should be cautiously approached, especially given that participants self-evaluated their level of English which might be based on personal perceptions rather than formal qualifications.

RQ3. Regarding intelligibility, to what extent are variations in output intelligible for human comprehension but not ASR transcription?

The study assessed how variations in spoken output, when produced by non-native speakers, were transcribed by ASR systems compared to how they were perceived in terms of intelligibility by human listeners. GE underscores intelligibility and mutual understanding, and in this context, ASR systems performed remarkably well in transcribing variations in non-native speech. A majority of the participants met industry-standard WER scores. For example, four participants recorded WER scores below 5%, while the remaining six speakers had WER scores ranging from 5% to 10%. This robust performance suggests that ASR systems adapt well to certain non-native variations. On reflection, the conditions of the investigation may explain these scores. This investigation focused on extemporaneous speech, which is prepared but not memorised. Participants were required to pitch without notes, and while there was a degree of improvisation, on reflection it was unclear how much rehearsal participants had undergone. All participants, apart from one speaker, who coincidentally had the highest WER

score, had prior practice with the pitch delivered for this investigation. This could suggest that experience or practice does impact the WER scores.

Substitution errors emerged as the most common type, with 247 instances. By nature, these lexical substitutions result from the system registering output but failing to interpret variation. Interestingly, the high frequency of these errors did not severely impede human comprehension. Most of these substitution errors still resulted in generally understandable content for human listeners, even if the ASR system did not transcribe them correctly. This raises an essential limitation of the WER metric as it does not appear to consider context. Contextualised utterances enable human listener to follow a conversation, and this should be considered when assessing ASR. Nevertheless, the data indicates that the ASR system can produce inaccuracies even when speech is predominantly intelligible. However, as speech intelligibility declines, errors stemming from speaker variations become more pronounced.

RQ4. Are there any common linguistic patterns used by Spanish speakers of English that cause inaccuracies with ASR transcriptions? Should certain elements be given more focus or avoided?

As mentioned, GE advocates for linguistic diversity and fluidity in various sociolinguistic environments over fixed or standardised language (Rose & Galloway, 2019). Moreover, the GE paradigm acknowledges the plurality and diverse variations of the English language, such as WE and ELF. This investigation aimed to understand better the relationship between these non-standard variations and the recorded transcription errors. The following section will refer to previous studies and findings with reference to the Lingua Franca Core.

One of the primary areas where the ASR system showed limitations was its capacity to transcribe startup-specific acronyms correctly. For instance, the acronym "SAAS" (Software as a Service) was constantly misunderstood. This was similarly observed for other specialist investment terms such as "SOM", "TAM", and "SAM". Cumulatively, inaccuracies related to these specific acronyms amounted to n=19. These words might not be part of the training data and, therefore, not included in the phonetic dictionary of the system. Alternatively, the system might not recognise the word as an acronym due to WER not considering the context.

The findings align with the literature and highlights ASR systems' potential challenges when encountering industry-specific terminology (Anastassiadis Serrat, 2021; Kohl, 2008).

Homophones, words that sound alike but carry different meanings, also presented challenges. These words rely on context rather than clear phonetic differentiation, so they became a clear source of transcription errors. This observation is consistent with the academic findings of Levis and Suvorov (2012), who identify the inherent ambiguity in homophones as a complicating factor. Overall, they did not impede listener comprehension.

Overall, individual sounds, i.e., phonemes, significantly impacted transcription accuracy. This is particularly true when distinguishing between short and long vowel sounds such as /i:/ and /I/. These proved challenging to both the speaker and the ASR system. Nevertheless, despite consistent substitutions, utterances were intelligible. Given this and the high frequency (n=30) in the data sample, it could be argued that in terms of diversity and fluidity, ASR systems could be improved to demonstrate more inclusivity of non-standard forms. However, utterances scored as 'slightly unintelligible' were recorded n=7 times. This suggests that while some variations in pronunciation by non-native speakers do not severely impact intelligibility, others consistently do.

Interestingly, substitutions between 'this' and 'these' were frequently observed. Nevertheless, most of these instances are intelligible to human transcription, suggesting a lack of ASR tolerance regarding an ability to discern variations in short and long vowel sounds. Furthermore, variations such as /ɒ/ and /ɔ:/ presented intelligibility issues and could be potential areas of improvement from the speaker's perspective. These findings are in a similar vein to the guidelines in the LFC, which comment on the need to differentiate between two characteristics of English vowels and emphasises vowel length, where learners need to focus their attention on the long-short differences between vowels instead of sounding like a native.

Moreover, consonant sounds presented challenges. The data suggests that sounds with similar articulation, especially the pairings between voiced and voiceless phonemes, often resulted in transcription errors. The interchangeable transcription of sounds such as /s/ and /z/, or /t/ and /d/ highlight this. The/ʤ/ and /j/ phonemes also proved problematic for the

ASR system. In addition, some errors appear to be caused by similarities with the place of articulation. For example, both /b/ and /p/ are bilabials and plosives, meaning airflow is blocked before release. This supports the findings of the LFC.

Of the L1 examples observed, epenthesis significantly impacts the transcription output. The system appears to register the speakers' release of air as a phoneme and transcribes this as the article 'a'. Consequently, the predictive model substituted the remaining sound as a noun as this is syntactically appropriate, e.g., article + noun, but semantically wrong. Epenthesis of /e/ before consonant clusters is a phonological feature of Spanish. While there were only three examples of this in the entire corpus, these lead to insertions and could present issues to ASR systems.

ASR's transcription capabilities are significantly impacted by disfluencies, particularly inadequate pausing and hesitations, resulting in unexpected breaks in speech flow. Moreover, inadequate pausing appears to negatively affect the ASR transcription's intelligibility. In the data, errors were commonly viewed as deletions where the system did not transcribe otherwise intelligible utterances. There also appears to be a correlation between disfluencies and the following unigrams. For example, in the data, it was observed that speech disfluencies frequently precede errors, suggesting a causal relationship between the two. Notably, while male participants occasionally exhibited these disfluencies, their female counterparts did not. This observation, however, must be approached with caution given the smaller sample size and the gender imbalance within it.

While their occurrence was limited, word stress and differences in prosodic patterns appeared to have a significant impact on the ASR transcription. The data suggests that correct nuclear stress placement is essential for maintaining intelligibility and accurate ASR transcriptions. Interestingly, this is further supported by the substitutions that were observed. For example, substituted words often had the same sound, stress pattern and/or number of n-grams. This suggests that a combination of accurate phoneme and stress placement is essential in mitigating ASR transcription errors.

Based on the above findings the following guidelines have been suggested to promote greater ASR accuracies in the context of Spanish startups pitching in English:

Area/focus	Guideline / Recommendation
General speech: Intelligibility	Ensure consistent speech flow and include adequate pauses for better transcriptions. Disfluencies and hesitation can lead to ASR deletions.
Grammar: Sentence structure	Use simpler sentence structure to avoid deletions caused by hesitation or false starts.
Grammar: Omissions	Avoid omitting the verb 'to be'. This can lead to substitutions which in turn alter meaning.
Grammar: Wrong preposition or article	Use articles and prepositions accurately. Variations result in the ASR system incorrectly predicting the following word.
Lexis: Acronyms	Specific terms e.g., SAAS. TAM, SOM, SAM can pose challenges. Consider pronouncing these slowly to mitigate errors due to connected speech.
Lexis: Buzz words	Avoid unnecessarily long buzz words e.g., multidisciplinary. ASR systems show limited tolerance with inaccurate word stress patterns. Be concise.
Pronunciation: Vowel Sounds	Focus on the differences in length between short and long vowel sounds. ASR systems are sensitive to these phonemes e.g., /iː/ and /I/.
Pronunciation: Voiced vs. unvoiced.	Articulate differences between voiced and unvoiced consonants e.g., /s/ and /z/, /t/ and /d/, /k/ and /g/. These are often misinterpreted.
Pronunciation: Similar articulation	Be careful with similar sounding consonants e.g., /b/ and /p/ (both bilabials) or /ʤ/ and /j/ (similarities in manner and place).
Pronunciation: Epenthesis	Avoid the inclusion of /e/ before consonant clusters e.g., /e/scalability. ASR recognises the /e/ sound, and this can lead to insertion errors.
Practice	Improvisation is likely to encourage inadequate pausing, disfluencies, and variation in rate of speech which can cause deletion and insertion errors.

Table 45. Guide to clearer ASR transcription - Spanish startups pitching in English.

Limitations

The data above has offered some insights into non-native speaker variation and ASR inaccuracies. Nevertheless, it is important to note some limitations. Firstly, the sample size is relatively small (n=10), which means that the findings may not be representative of the broader population. Secondly, there is a significant gender imbalance in the sample, with a higher number of male participants. Therefore, while the data can be useful for the purpose

of analysis, a lack or representation has prevented the investigation from drawing conclusive findings between the efficacy of ASR systems and sociolinguistic factors.

Further areas of study could include the collection of data from more participants. Moreover, requesting additional factors could help to improve the overall quality of the data. For example, establishing how often startups have practised their pitch might help to reveal correlations between practice and the WER score. Other factors, such as the onset of English learning could be useful given that it is often associated with a clearer level of pronunciation. Finally, analysing the entire oral corpus to establish the extent to which the ASR system correctly transcribes all utterances could prove insightful. However, this was outside of the scope of this TFM.

References

Alderson, J. C. (2007). The CEFR and the need for more research. *The Modern Language Journal, 91*(4), 659-663. https://doi.org/10.1111/j.1540-4781.2007.00627_4.x

Anastassiadis Serrat, V. (2021). *The impact of disfluencies on ASR transcriptions. An evaluation of WIPO S2T.* [MA thesis, Université de Genève]. https://archive-ouverte.unige.ch/unige:164334

Archer, G. (2023) Integrating student-centred assessment practices. *Testing, Evaluation and Assessment Today, 8*. https://www.researchgate.net/publication/370652600_Integrating_student-centred_assessment_practices

Aston University (2021) *LO-C 30 Report.* https://www.aston.ac.uk/research/bss/abs/loc30-report

Baker, W., Dewy, M., Jenkins, J. (Eds). (2018). *The Routledge Handbook of English as a Lingua Franca*. Routledge. https://doi.org/10.4324/9781315717173

Bankinter. (2022). *Spanish Investment Tendencies.* https://www.fundacionbankinter.org/wp-content/uploads/2023/02/Tendencias-de-Inversion-Espana-2022.pdf

British Council. (2023). *The Future of English: Global Perspectives.* https://tinyurl.com/2xxnw9tj

Calefato, F., Lanubile, F., Conte, T., & Prikladnicki, R. (2016). Assessing the impact of real-time machine translation on multilingual meetings in global software projects. *Empirical Software Engineering*, *21*(3). https://doi.org/10.1007/s10664-015-9372-x

Carpenter, S. (2015). *A Startup's Guide to International Expansion.* https://techcrunch.com/2015/12/23/a-startups-guide-to- international-expansion/

Cavusgil, S. T., & Knight, G. A. (2015). The born global firm: An entrepreneurial and capabilities perspective on early and rapid internationalisation. *Journal of International Business Studies. 46*(1). https://doi.org/10.1057/jibs.2014.62

Canagarajah, A. Suresh. (2013). *Translingual Practice: Global Englishes and Cosmopolitan Relations*. Routledge. https://ebookcentral.proquest.com/lib/bibliotecaupves-ebooks/detail.action?docID=1101434&pq-origsite=primo

Chan, M. P. Y., Choe, J., Li, A., Chen, Y., Gao, X., & Holliday, N. (2022). Training and typological bias in ASR performance for world Englishes. *Interspeech* 2022, 1273–1277. https://doi.org/10.21437/Interspeech.2022-10869

Clavel, C., Adda, G., Cailliau, F., Garnier-Rizet, M., Cavet, A., Chapuis, G., Courcinous, S., Danesi, C., Daquo, A.-L., Deldossi, M., Guillemin-Lanne, S., Seizou, M., & Suignard, P. (2013).

Spontaneous speech and opinion detection: Mining call-centre transcripts. *Language Resources and Evaluation*, *47*(4), 1089–1125. https://doi.org/10.1007/s10579-013-9224-5

Cogo, A. (2012) English as a Lingua Franca: concepts, use, and implications. *ELT Journal*, *66*(1), 97–105. https://doi.org/10.1093/elt/ccr069

Cogo, A., & House, J. (2018). The pragmatics of ELF. In: W. Baker, M. Dewey, & J. Jenkins (Eds). *The Routledge Handbook of English as a Lingua Franca* (pp. 210-223). Routledge. https://doi.org/10.4324/9781315717173

Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English. *System*. *27*(1) 49-64. https://doi.org/10.1016/S0346-251X(98)00049-9

Corl, E. (2019). *How Startups Drive the Economy.* https://medium.com/@ericcorl/how-startups-drive-the-economy-69b73cfbae1

Council of Europe (2001). *Common European Framework for Languages: Learning, teaching, and assessment.* https://rm.coe.int/16802fc1bf

Council of Europe (2020). *Common European Framework for Languages: Learning, teaching, and assessment.* https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

Coviello, N. (2006, July). The network dynamics of new international ventures, *Journal of International Business*, 5 (37), 713-731. https://www.jstor.org/stable/4540377

Crystal, D. (2005). *English as a global language*. Cambridge University Press.

Cushing, I. (2023). Policy Mechanisms of the Standard Language Ideology in England's Education System. *Journal of Language, Identity & Education*, *22*(3), 279–293. https://doi.org/10.1080/15348458.2021.1877542

D'Angelo, J. (2012). Curriculum and world Englishes. In E. Low & A. Hashim (Eds.), *English in South East Asia* (pp.289-306). John Benjamins Publishing Company. https://doi.org/10.1075/veaw.g42.22dan

D'Angelo, J. (2016). *A Broader Concept of World Englishes for Educational Contexts: Applying the "WE Enterprise" to Japanese Higher Education Curricula.* [Doctoral thesis, North West University]. http://hdl.handle.net/10394/17014

Dealroom (2023). *Spain Ecosystem Report.* https://dealroom.co/uploaded/2023/04/Dealroom-Spain-startup-Report-2023-v4.pdf

De Bernardi, P. D., & Azucar, D. (2020). *Startups and knowledge sharing in ecosystems: incumbents and new ventures. Innovation in Food Ecosystems*. 161-188. Springer Link. https://link.springer.com/chapter/10.1007/978-3-030-33502-1_6

Deterding, D. (2013). *Misunderstandings in English as a Lingua Franca: An Analysis of ELF Interactions in South-East Asia*. De Gruyter Mouton. https://doi.org/10.1515/9783110288599

Deygers, B., Zeidler, B., Vilcu, D., & Carlsen, C. H. (2018). One framework to unite them all? Use of the CEFR in European university entrance policies. *Language Assessment Quarterly*, *15*(1), 3–15. https://doi.org/10.1080/15434303.2016.1261350

Derwing, T. M., Munro, M. J., & Carbonaro, M. (2000). Does Popular Speech Recognition Software Work with ESL Speech? *TESOL Quarterly*, 34(3), 592-603. https://doi.org/10.2307/3587748

Dong, Y., & Li, D. (2015). Automatic Speech Recognition: A Deep Learning Approach. Springer. https://tinyurl.com/4fvumvfr

El Hannani, A., Errattahi, R., Salmam, F.Z. et al. (2021). Evaluation of the effectiveness and efficiency of state-of-the-art features and models for automatic speech recognition error detection. *Journal of Big Data 8* (5). https://doi.org/10.1186/s40537-020-00391-w

Englis, P. D., Wakkee, I., & Sijde, P. V. D. (2007). Knowledge and networks in the global startup process. *International Journal of Knowledge Management Studies,* 1(3), 497-514. https://doi.org/10.1504/IJKMS.2007.012538

Enisa (2021). *Spanish Tech Ecosystem.* https://www.enisa.es/foro/contenido/Dealroom-Spanish-tech-ecosystem-report-2021.pdf

Errattahi, R., El Hannani, A., & Ouahmane, H. (2018). Automatic Speech Recognition Errors Detection and Correction: A Review. *Procedia Computer Science*, *128*, 32–37. https://doi.org/10.1016/j.procs.2018.03.005

Ester, P. (2017). *Accelerators in Silicon Valley.* Amsterdam University Press. Amsterdam University Press. https://www.jstor.org/stable/j.ctt1zrvhk7.10

Fairbairn, M., Kish, Z., & Guthman, J. (2022). Pitching agri-food tech: Performativity and non-disruptive disruption in Silicon Valley. *Journal of Cultural Economy*, *15*(5), 652–670. https://doi.org/10.1080/17530350.2022.2085142

Fang, F., & Ren, W. (2018). Developing students' awareness of global Englishes. *ELT Journal*, *72*(4), 384–394. https://doi.org/10.1093/elt/ccy012.

Figueras, N. (2012) The impact of the CEFR, *ELT Journal*, *66*(4), 477–485. https://doi.org/10.1093/elt/ccs037

Fisher, J. H. (1996). *The Emergence of Standard English*. University Press of Kentucky. http://www.jstor.org/stable/j.ctt130jmk2

Forbes. (2023). *Remote Work Statistics and Tech Trends In 2023*. https://www.forbes.com/advisor/business/remote-work-statistics/

Gabrielsson, M., Pelkonen, T. (2008). Born internationals: Market expansion and business operation mode strategies in the digital media field. *Journal of International Entrepreneurship 6*, 49–71. https://doi.org/10.1007/s10843-008-0020-z

Galloway, N., & Rose, H. (2015). *Introducing global Englishes*. Routledge.

Galloway, N., & Rose, H. (2018). Incorporating Global Englishes into the ELT classroom. *ELT Journal*, *72*(1), 3–14. https://doi.org/10.1093/elt/ccx010

Galloway, R., & Numajiri, T. (2019). Global Englishes Language Teaching: Bottom-up Curriculum Implementation. *Tesol Quarterly 54*(1), 118-145. https://doi.org/10.1002/tesq.547

Gauthier, J., Penel, M. & Marmer, M. (2017). *Global Startup Ecosystem Report.* https://tinyurl.com/mrpebnsz

Google (n.d.). *Measure and improve speech accuracy.* https://cloud.google.com/speech-to-text/docs/speech-accuracy

Graddol, D. (1999). The decline of the native speaker. In G. Anderman & M. Rogers (Eds). *Translation Today: Trends and perspectives* (pp. 152-167). Multilingual Matters. https://tinyurl.com/55dpmdjw

Graddol, D. (2006). English Next. *Why Global English May Mean the End of "English as a Foreign Language*. British Council. https://www.teachingenglish.org.uk/sites/teacheng/files/pub_english_next.pdf

Graddol, D. (2019) Global English. The Open University. https://www.open.edu/openlearn/history-the-arts/culture/english-language/global-english

Halliday, M. A. K. (2003). Written language, standard language, global language. *World Englishes, 22*(4), 405-418.

Hernández, N. (2023). South Summit 2023: 'Neither rain nor politicians put a damper on the biggest startup party of the year'. *Disruptores e Innovadores.* https://tinyurl.com/mv7jvh94

Hickey, R. (2012). Standard English and standards of English. In R. Hickey (Ed.), Standards of English: Codified Varieties around the World (Studies in English Language, pp. 1-33). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139023832.002

Hinsvark, A., Delworth, N., Del Rio, M., McNamara, Q., Dong, J., Westerman, R., Huang, M., Palakapilly, J., Drexler, J., Pirkin, I., Bhandari, N., Jette, M. (2021). Accented Speech Recognition: A Survey. https://doi.org/10.48550/arXiv.2104.10747

Holliday, A. (2006) Native-speakerism. *ELT Journal, 60*(4), 385-387. https://doi.org/10.1093/elt/ccl030

House, J. (2002). Communicating in English as lingua franca. *EUROSLA Yearbook*, *2*(1), 243-261. https://doi.org/10.1075/eurosla.2.15hou

Irvine, J., & Gal, S. (2000). *Language Ideology and Linguistic Differentiation.* University of Stanford. https://web.stanford.edu/~eckert/PDF/IrvineGal2000.pdf

Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford University Press.

Jenkins, J. (2006). Current perspectives on teaching World Englishes and English as a Lingua Franca. *Tesol Quarterly, 40*(1), 157-181. https://doi.org/10.2307/40264515

Jenkins, J. (2014). *Global Englishes: A Resource Book for Students*. Routledge. https://doi.org/10.4324/9781315761596.

Jenkins, J. (2015). Repositioning English and multilingualism in English as a Lingua Franca. *Englishes in Practice, 2*(3), 49-85. https://doi.org/10.1515/eip-2015-0003

J.P.Morgan. (2023). *South Summit*. https://privatebank.jpmorgan.com/gl/en/of-interest/south-summit-2023

Jurafsky, D. & Martin, J. (2023) *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* (2nd ed). Pearson. https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the Outer Circle. In R. Quirk and H. Widdowson (Eds.). *English in the World: Teaching and Learning the Language and Literatures* (pp. 11–30). Cambridge University Press. https://doi.org/10.1017/S027226310000677X

Kachru, B. (1992). World Englishes: Approaches, issues and resources. *Language Teaching, 25*(1), 1-14. https://doi.org/10.1017/S0261444800006583

Kiczkowiak, M. (2019). Seven principles for writing materials for English as a lingua franca. *ELT Journal*, *74*(1), 1-9. https://doi.org/10.1093/elt/ccz042

Kiczkowiak, M. (2020). Researcher's Attitudes to Hiring 'Native' and 'Non-Native Speaker' Teachers: An International Survey. *The Electronic Journal for English as a Second Language, 24*(1), http://www.tesl-ej.org/pdf/ej93/a4.pdf

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D., & Goel, S. (2020). *Racial disparities in automated speech recognition. Proceedings of the National Academy of Sciences, 117*(14). https://doi.org/10.1073/pnas.1915768117

Kohl, J.R. (2008). *The Global English Style Guide: Writing Clear, Translatable Documentation for a Global Market*. SAS Institute. https://tinyurl.com/yc6vaenz

Lai, C., & Markl, N. (2021) Context-sensitive evaluation of automatic speech recognition: considering user experience & language variation. *Association for Computational Linguistics*, 34-40. https://aclanthology.org/2021.hcinlp-1.6

Lee, S., & Jeon, J. (2023). Addressing automatic speech recognition for ELT from a Global Englishes perspective. *ELT Journal*. https://doi.org/10.1093/elt/ccad038

Levis, J., M. & Suvorov., R. (2012). Automated Speech Recognition. In Chapelle, C. (Ed). *Encyclopedia of Applied Linguistics*. Blackwell. https://doi.org/10.1002/9781405198431.wbeal0066

Lilischkis, S., Te Velde, R. Korlaar, L. (2016). *Internationalisation of innovation in SMEs: case studies, exemplary support practices and policy implications.* Publications Office: European Commission, Directorate-General for Research and Innovation. https://data.europa.eu/doi/10.2777/440818

Lippi-Green, R. (1999). *English with an Accent: Language ideology, and discrimination in the United States*. Routledge. https://tinyurl.com/32e6s3c9

Little, D. (2006). The Common European Framework of Reference for Languages: Content, purpose, origin, reception, and impact. *Language Teaching, 39*(3), 167-190. https://doi.org/10.1017/S0261444806003557

Lounsbury, M., & Glynn, M. A. (2001). Cultural Entrepreneurship: Stories, legitimacy, and the acquisition of resources. *Strategic Management Journal, 22*(6-7), 545–564. https://doi.org/10.1002/smj.188

Lowenberg, P. H. (2000). Non-native varieties and the sociopolitics of English proficiency assessment. In J.K. Hall & W. G. Eggington (Eds.). *The sociopolitics of English language teaching* (pp.67-82). Multilingual Matters. https://tinyurl.com/yckzfpce

Li, J., Deng, L., Haeb-Umbach, R., & Gong, Y. (2015). *Robust automatic speech recognition: a bridge to practical applications.* https://tinyurl.com/yx7dpr5r

Mariño, C. (2011). Reflecting on the dichotomy native-non native speakers in an EFL context. *Anagramas-Rumbos y sentidos de la comunicación, 10*(19), 129-141. https://doi.org/10.22395/angr.v10n19a8

Markl, N. (June, 2022). Language variation and algorithmic bias: Understanding algorithmic bias in British English automatic speech recognition. *ACM Conference on Fairness, Accountability, and Transparency*, 521–534. https://doi.org/10.1145/3531146.3533117

Matsuda, A. (2003). Incorporating World Englishes in Teaching English as an International Language. *TESOL Quarterly, 37*(4), 719–729. https://doi.org/10.2307/3588220

Mair, C. (2003). *The Politics of English as a World Language*. Rodopi. https://tinyurl.com/yc6wkew9

Mauranen, A. (2012). *Exploring ELF: Academic English shaped by non-native speakers.* Cambridge University Press. https://tinyurl.com/mrxp765m

McKay, S. (2002). *Teaching English as an International Language*. Oxford: Oxford University Press.

McNamara, T., & Shohamy, E. (2016). Language testing and ELF: Making the connection. In M.-L. Pitzl & R. Osimk-Teasdale (Eds.), *English as a Lingua Franca: Perspectives and Prospects* (227–234). De Gruyter. https://doi.org/10.1515/9781501503177-030

Microsoft. (2023). Test accuracy of a Custom Speech model. https://tinyurl.com/2r7syj2w

Milroy, J., & Milroy, L. (2012) *Authority in language. Investigating Standard English* (4th ed.) Routledge. https://tinyurl.com/3tmn6fsp

Ministry of Economic Affairs and Digital Transformation (2022). *Digital Spain 2025.* https://espanadigital.gob.es/sites/agendadigital/files/2022-01/Digital-Spain-2025.pdf

Neubert, M. (2018). The Impact of Digitalisation on the Speed of Internationalisation of Lean Global Startups. *Technology Innovation Management Review, 8* (5). https://ssrn.com/abstract=3394507

Nevalainen, T., & Van Ostade, I. (2006). Standardisation. In R. Hogg & D. Denison (Eds.). *A History of the English Language* (pp. 271-311). Cambridge University Press. https://doi.org/10.1017/CBO9780511791154.006

North, B. (2021). The CEFR companion volume—What's new and what might it imply for teaching/learning and for assessment? *CEFR Journal: Research and Practice, 4*, 5–24. https://doi.org/10.37546/JALTSIG.CEFR

Oviatt, B. & McDougall P. (1995). Global start-ups: Entrepreneurs on a worldwide stage. *Academy of Management Executive*, *9*(2), 30-44. https://www.jstor.org/stable/40836140

Pasandi, H. B., & Pasandi, H. B. (2022). Evaluation of ASR Systems for Conversational Speech: A Linguistic Perspective. *Proceedings of the Twentieth ACM Conference on Embedded Networked Sensor Systems,* 962–965. https://doi.org/10.1145/3560905.3568297

Pennycook, A. (2008). Multilithic English(es) and language ideologies. *Language in Society, 37*(3), 435–444. https://doi.org/10.1017/S0047404508080573

Pérez Castillejo, S. (2021). Automatic speech recognition: Can you understand me? In T. Beaven & F. Rosell-Aguilar (Eds.), *Innovative language pedagogy report* (pp. 121–126). Research-publishing.net. https://doi.org/10.14705/rpnet.2021.50.1246

Prabjandee, D., & Savski, K. (2022). CEFR: A Global Framework for Global Englishes? *The Electronic Journal for English as a Second Language*,*26* (3). https://doi.org/10.55593/ej.26103a1

PWC. (2023). *Socio-economic impact of South Summit in Madrid*. https://www.southsummit.io/wp-content/uploads/2023/04/ENG-Impacto-Socioeconomico-South-Summit-2023.pdf

Rose, H., & Galloway, N. (2019). *Global Englishes for Language Teaching.* Cambridge University Press. https://doi.org/10.1017/9781316678343

Sadeghpour, M., & D'Angelo, J. (2022). World Englishes and 'Global Englishes': competing or complementary paradigms?, *Asian Englishes, 24*(2), 211-221. https://doi.org/10.1080/13488678.2022.2076368

Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics, 24*, 209-239. https://www.proquest.com/docview/85671035

Seoane, E., & Suárez-Gómez, C. (2016). *World Englishes : New theoretical and methodological considerations.* John Benjamins Publishing Company. https://ebookcentral.proquest.com/lib/bibliotecaupves-ebooks/reader.action?docID=4526412

Sharifian, F. (2009). *English as an International Language: Perspectives and Pedagogical Issues.* Multilingual Matters. https://doi.org/10.21832/9781847691231

Startup Law. (2022). *Digital Spain 2025. Axis 06. Digital transformation of business and digital entrepreneurship. Measure 29.* https://espanadigital.gob.es/sites/agendadigital/files/2022-02/E06M29_Startups_Law.pdf

Solmaz, O. (2023). Linguistic landscapes tasks in Global Englishes teacher education. ELT Journal. https://doi.org/10.1093/elt/ccad027

Statista (2023). *The most spoken languages worldwide in 2023.* https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/

Tatman, R. (2017). Gender and Dialect Bias in YouTube's Automatic Captions. *Proceedings of the First ACL Workshop on Ethics in Natural Language* Processing, 53–59. https://doi.org/10.18653/v1/W17-1606

Tamkin, A., Jurafsky, D., & Goodman, N. (2020). Language through a prism: A spectral approach for multiscale language representations. *Advances in Neural Information Processing Systems, 33*. https://doi.org/10.48550/arXiv.2011.04823

Tech.Eu. (2022). *Inspiring the Good Future at Valencia Digital Summit.* https://tech.eu/2022/09/28/inspiring-the-good-future-at-valencia-digital-summit/

Teker, D., Teker, S., & Teraman, Ö. (2015). Venture Capital Markets: A Cross Country Analysis. *Procedia Economics and Finance, 38*, 213–218. https://doi.org/10.1016/S2212-5671(16)30192-7

The English Effect. (2013). *The impact of English, what it's worth to the UK and why it matters to the world.* https://www.britishcouncil.org/sites/default/files/english-effect-report-v2.pdf

The Economist. (2019). *Where are the world's best English speakers?* https://www.economist.com/graphic-detail/2019/12/04/where-are-the-worlds-best-english-speakers

The Economist. (2023). *As it spreads around the world, who owns English?* https://www.economist.com/culture/2023/05/25/as-it-spreads-across-the-world-who-owns-english

Trudgill, P. (1999). Standard English: What it isn't. Lagb-education. https://lagb-education.org/wp-content/uploads/2016/01/SEtrudgill2011.pdf

Tsuda, Y. (2008). *English hegemony and English Divide*. Utah Valley University.

Turunen, H., & Nummela, N. (2017). Internationalisation at home: The internationalisation of location-bound service SMEs. *Journal of International Entrepreneurship, 15*(1), 36-54. https://doi.org/10.1007/S10843-016-0167-Y

Van Doremalen, J., Strik, H., Cucchiarini, C., (2009). *Utterance Verification in Language Learning Applications*. Radboud University. https://tinyurl.com/t82rd996

Vettorel, P. (2018). ELF and Communication Strategies: Are They Taken into Account in ELT Materials? *RELC Journal*, *49*(1), 58–73. https://doi.org/10.1177/0033688217746204

Wassink, A. B., Gansen, C., & Bartholomew, I. (2022). Uneven success: Automatic speech recognition and ethnicity-related dialects. *Speech Communication, 140*, 50–70. https://doi.org/10.1016/j.specom.2022.03.009

Walker, R. (2010). Teaching the Pronunciation of English as a Lingua Franca. Oxford University Press. https://doi.org/10.1002/9781118346952.ch24

Widdowson, H. G. (1994). The Ownership of English. *TESOL Quarterly, 28*(2), 377–389. https://doi.org/10.2307/3587438

Wolfram, W., & Schiling-Estes, N. (2005). *American English: Dialects and Variation*. https://tinyurl.com/5966zznm

Appendix I: Sample of questions from the questionnaire.

TFM: Speech recognition in the context of global English and Spanish startups

This research aims to analyse the use of Automated Speech Recognition software (ASR) with Spanish startup founders using English as a second language in the context of an online pitch.

This survey will only take 5 minutes of your time. All data will be treated confidentially and will only be used for the purposes of this research.

What is your current level of spoken English (approximately)? *

- ○ Beginner (A1)
- ○ Beginner (A2)
- ○ Lower-intermediate (B1)
- ○ Upper intermediate (B2)
- ○ Advanced (C1)
- ○ Proficient (C2)

Do you use English in your work? *

- ○ Yes
- ○ No

In which work situations do you use English? You can tick as many as are relevant. If you do not use English, select "I don't use English in my startup". *

- [ ] Talking to clients on the phone
- [ ] Talking to clients by videoconference
- [ ] Establishing professional contacts
- [ ] Giving presentations
- [ ] Speaking at meetings
- [ ] Seeking foreign investors
- [ ] Recruiting people
- [ ] Other
- [ ] I don't use English in my startup

How many hours a week do you speak English? If you don't speak English, write 'zero'. *

Short-answer text

How many hours a week do you speak English with native speakers? If you don't speak English with native speakers, write 'zero'. *

Short-answer text

Have you ever delivered a pitch online in English before? *

- ( ) Yes
- ( ) No

Have you ever practiced the pitch you are going to record for this experiment before? *

- ○ Yes
- ○ No

Have you ever presented your pitch in English in a context where the audience or participants included native Spanish speakers? For example, at technology events. You can tick as many as are relevant. *

- ☐ Yes, among the audience
- ☐ Yes, among participants (e.g. jury or investors)
- ☐ No, everyone in the audience was a native English speaker.
- ☐ No, all participants were native English speakers.
- ☐ I don't know
- ☐ I have never pitched in English before

Appendix II: Consent form

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Departamento de Lingüística Aplicada
Departament de Lingüística Aplicada

CONSENT FORM

Automated Speech Recognition software in the context of Global English.

Thank you for your interest in this project. Below you will find information about the investigation and a request for your consent to participate. If you agree, please sign this document and reply to me via email.

- **What is the study about?**
  The focus of this study is to analyse the use of speech recognition and machine translation software within the context of Global English, specifically Spanish startups pitching for investment.

- **What will my involvement be?**
  You will be required to present an investor pitch in English for a duration of five minutes. This will take place online using a videoconferencing platform and the video, audio and transcript will be recorded.

- **What will my information be used for?**
  The data will be collected and analysed for the purpose of a Masters' dissertation and research.

- **Will my information be anonymous?**
  Your participation will be anonymous - your personal and company details will be anonymised and will not be used in any reports or publications resulting from the study.

If you agree to participate in the research, please read the following statements and tick each box:

I have read and understood the project information. The project has been fully explained to me and I have been given the opportunity to ask questions.	
I agree to participate in the project, and I understand that this will include the recording of video, audio, and transcripts.	
I understand my personal details such as name, phone number, address, and email address as well as company information will be kept confidential and will be anonymised.	
I understand and agree that linguistic data may be quoted in publications, reports, web pages, and other research or publications but I will not be named in these outputs.	

Name of participant [Printed] Signature Date

Name of researcher [Printed] Signature Date

Contact details for further information:

Email address:
Universitat Politècnica de València: Department of Applied Linguistics