# Testing the efficiency of voice recognition software in translation

**Pernarčić, Marko**

J. J. Strossmayer University of Osijek
Faculty of Humanities and Social Sciences
Study Programme: Double Major MA Study Programme in English Language and Literature – English Translation and Interpreting Studies and Publishing

Marko Pernarčić

# Testing the efficiency of speech recognition software in translation

Master's Thesis

Supervisor: Dr. Marija Omazić, Full Professor

Osijek, 2019

J. J. Strossmayer University of Osijek
Faculty of Humanities and Social Sciences
Study Programme: Double Major MA Study Programme in English Language and Literature – English Translation and Interpreting Studies and Publishing

Marko Pernarčić

# Testing the efficiency of speech recognition software in translation

Master's Thesis

Supervisor: Dr. Marija Omazić, Full Professor

Osijek, 2019

Sveučilište J. J. Strossmayera u Osijeku
Filozofski fakultet
Studij: Dvopredmetni sveučilišni diplomski studij engleskog jezika i književnosti – prevoditeljski smjer i nakladništva

Marko Pernarčić

# Ispitivanje učinkovitosti alata za prepoznavanje govora u procesu prevođenja

Diplomski rad

Mentor: prof. dr. sc. Marija Omazić
Osijek, 2019.

Sveučilište J.J. Strossmayera u Osijeku
Filozofski fakultet Osijek
Odsjek za engleski jezik i književnost
Studij: Dvopredmetni sveučilišni diplomski studij engleskog jezika i književnosti – prevoditeljski smjer i nakladništva

Marko Pernarčić

# Ispitivanje učinkovitosti alata za prepoznavanje govora u procesu prevođenja

Diplomski rad

Znanstveno područje: humanističke znanosti
Znanstveno polje: filologija
Znanstvena grana: anglistika
Mentor: prof. dr. sc. Marija Omazić
Osijek, 2019.

**Abstract**

Today, the translation service industry and the speech technology industry are both in significant growth, showing no indication of stopping in the near future. Speech recognition technology is slowly becoming integrated in Computer Assisted Translation (CAT) tools, which has the potential to increase productivity of the translation process. The aim of this paper is to determine whether speech recognition is a more efficient input method than typing in the translation process—assessing primarily speed and accuracy. Additionally, the goal was to establish the level of adequacy of the integration of speech recognition technology with CAT tools in its current instance. The research consisted of tests through which the respondents were tested for their typing, dictating, and translating abilities, along with speed. Two independent evaluators assessed the translations, ultimately using the average of scores and grades. We have established a sizable number of metrics and used three different nonparametric tests in order to test the hypotheses for 8 research questions. Based on the results obtained, we have provided a foundation for improving the translation process.

**Key words**: translation industry, translation technology, translation efficiency, speech recognition, CAT tools

**Sažetak**

Prevoditeljska je industrija, uz industriju govornih tehnologija, trenutno u značajnom rastu i obje ne pokazuju naznake zaustavljanja u skoroj budućnosti. Tehnologija prepoznavanja govora polako se uvodi u alate za računalno potpomognuto prevođenje, što sadrži potencijal za unapređenje produktivnosti pri procesu prevođenja. Cilj ovog rada je ustanoviti je li je prepoznavanje govora efikasnija metoda unosa od korištenja tipkovnice kod prevođenja, pri čemu su se u obzir primarno uzimala brzina i točnost. Nadalje, nastojala se procijeniti adekvatnost integracije tehnologije prepoznavanja govora u alate za računalno potpomognuto prevođenje, u svojem trenutnom izdanju. Istraživanje se sastojalo od testova putem kojih se se provjeravala sposobnost i brzina tipkanja, diktiranja i prevođenja kod ispitanika. Dva neovisna ocjenjivača ocjenjivala su prijevode, te se koristio prosjek njihovih bodova i ocjena. Odredili smo veći broj parametara te koristili tri različita neparametrijska testa kako bi ispitali hipoteze kroz 8 istraživačkih pitanja. Prikupljeni rezultati poslužili su kao temelj za buduće unapređenje procesa prevođenja.

**Ključne riječi**: prevoditeljska industrija, prijevodne tehnologije, efikasnost prevođenja, prepoznavanje govora, alati za računalno potpomognuto prevođenje

**Table of contents**

## 1. Introduction

As we are approaching the second decade of the 21st century, it's a mere platitude to point out how the past 30 decades have witnessed the titanic rise of information and communication technology—to such an extent that modern discourse identifies it as the Third Industrial Revolution. The digital revolution has changed merely every field of human activity—and those changes show no sign of coming to a halt any time soon. Given the fact that the general premise behind digitalization is maximizing productivity and efficiency, many sceptics have voiced their concerns over the anticipated trend of increased automation and artificial intelligence to eventually cause a surplus of human workforce, leading to an ever-growing number of workers getting laid off.

How all that reflects on the translation industry is very palpable. An example that immediately springs up is Google Translate, the world's most famous free-to-use translation service. With the service continually growing more and more endowed with functionality and reliability overall, the question that underlies is whether there will be a decrease in demand for a number of roles carried out by professionally educated translators.

With the recent growing trend of using speech recognition in day-to-day life through voice assistants, which are slowly becoming capable of carrying out tasks seen in science fiction movies—from turning the lights on in a room to diagnosing diseases—it is only a matter of time until voice assistants become adept at speech translation. However, at this point, speech recognition has just recently started to integrate into Computer Assisted Translation (CAT) tools, thus the aim of our research is to determine the current level of efficiency of using speech recognition input in translation, compared to the traditional typing input. Additionally, our goal was to establish the level of adequacy of the integration of speech recognition with CAT tools in its current instance. Ultimately, we wanted to shed some light on the direction in which translators should move in order to maximize their productivity, and keep their heads above water in this highly competitive industry.

## 2. Current state of the translation service industry

Although translation has been an essential human process since time immemorial, the framework of the translation service industry remained generally stagnant until the emergence of translation technology during mid-20th century. The invention of the computer brought about revolutionary advancements of the translation process such as machine translation, computer-aided translation, and translation memory. Concurrently, the growing sphere of globalization through foreign trade and cultural transmission lead to a greater demand for translated content. With the introduction of the internet, along with continuing rapid technological advancements, the translation services industry essentially skyrocketed in the 21st century, and has been thriving since. Numbers show that the global market share of the language services industry has more than doubled over the last ten years (from 23.5 billion USD to 49.6 billion USD), and even estimate a leap to 56 billion USD in 2021 (Mazareanu, 2019).

But what exactly does the translation service industry encompass? In a very short time frame it has grown to be much more than merely companies that offer translations from source to target language and interpreting services. The industry is very fragmented and consists of an ever-diversifying number of agents across translation bureaus, localization companies, and technology vendors (Balkul, 2016:105). Further categorization could be extended to: "language service providers (LSPs), language technology and software developers, in-house localization/translation teams, linguists, research analysts, publications, and training institutes; and globalization and localization consultants" (GALA). Evidently, a growing number of newcomers with a background in the fields of information technology and marketing are entering the industry. Globalization allowed businesses to branch out easily to foreign countries—whereby, investing in their language services sector becomes imperative in order to connect with prospective local partners and consumers. The effect that has on a large scale is such that it bolsters competition in terms of the cost and price of their offered service, target market, technological infrastructure, speed of production, and so forth (Crowther and Aras, 2010).

How that reflects on the industry is that the volume of demanded translations is on the rise, while translation technology is finding ways to make the translation process quicker and more efficient. For instance, the Directorate-General for Translation (DGT), which is the European Commission's in-house translation service, translated about 2 255 000 pages in 2018—marking an 11% increase in output compared to 2017, in spite of being endowed with fewer internal resources (DGT, 2019).

From a business' perspective, having a multilingual internet presence and online publications correlate with revenue growth (Pangeanic). What is more, the translation services industry is one of few industries that are not affected by global recession, displaying a compound annual growth rate of 12% since 2008—a period of global economic decline colloquially dubbed 'The Great Recession' (Mellor, 2015). Nordictrans (2019) points out the following factors pertaining to the industry's robustness:

1. The continuous process of globalization
2. The number of business transactions across borders has increased
3. The lack of negative influence that could be brought by free web-based translation programs
4. The fact that many translation agencies embraced technology, determining more efficient translations and bringing faster deadlines

All that is to stress how the translation market is under-researched and that there is a lot to be improved. While many fear that future holds the total replacement of translators with robots, the dynamic role of translators is becoming more entrepreneurial—Sin-wai pragmatically asserts that it is up to them to utilize the technology to their advantage in order to maximize productivity and, in turn, their profits (2015:45). The traditional process of translating is gradually shifting from being the primary focus of a translator's profession. With the advancement of machine translation, translators are expected to strengthen their competence in post-editing, since humans are vastly superior with handling cultural aspects of translating, which is particularly important in the process.

## 3. Translation technology

Over the past three decades, the process of translating has become irreversibly symbiotic with technology. Translation technology is young and moves very quickly. Since its humble beginnings in the 1960s, it went through a period of rapid growth in the past decade, and currently is in the stage of expanding to developing countries. There are many ways of categorizing translation-focused computer systems, but the two fundamental categories of translation technology that fall within the scope of this paper are computer-assisted translation (CAT) tools and machine translation (MT).

### 3.1. Computer-assisted translation (CAT) tools

While CAT tools were considered a precious commodity as near as a decade ago, today many in-house translation positions require translators to be proficient in at least one of CAT software. Most professional translators use them—primarily those specializing in technical and legal fields, where an emphasis is placed on terminology consistency and formulaic writing, unlike in literary translation, for example.

A 'CAT tool' is an umbrella term used for software applications tailored to assist translators throughout the translation process, with the aim of maximizing productivity. They are based on four main functions (Candel-Mora, 2016:53):

1. Text search algorithm: facilitates referencing source and target languages by visually paralleling the two texts one to another. It proves to be convenient when it comes to specialized language and looking up equivalents in a certain register.
2. Project management statistics and analysis: provides analytics relevant to the translator's productivity based on the volume of text and tracked history of efficiency, e.g., estimated time for completing the translation.
3. Terminology management: provides automatic look-up in terminology databases and interaction with search results.
4. Segment alignment component: breaks texts into segments, allowing re-use of previously translated segments. It's the paramount feature representing productivity of CAT tools.

In order to fully utilize all the functions, a CAT tool focuses every translation project around translation memories (TM), which are databases consisted of previously translated segments used to recycle them for future use, and extract metrics relevant to translators' productivity. In fact, the TM's value is such that the more a translator translates, it leaves him with less to translate. Although the formats of TM differ among CAT tools, they can usually be converted for interchangeable use.

In addition to the aforementioned main functions, CAT tools have gained many more features that enhance productivity. Prima Lingua, one of the industry leaders in technical translation, names the following features characteristic of modern CAT tools:

1. Spell checkers, autocorrect—automatically highlight and fix spelling and grammar mistakes.
2. In-context review—display multimedia documents with images, text box sizes and layout in real time in both source and target language. Today CAT tools are not limited to recognizing text files only. They are able to process various kinds of multimedia files such as webpage files, PDF documents, Microsoft Office suite files, Java Properties files, etc.
3. Integrated machine translation (MT)—suggest translations for segments from a connected MT engine.
4. Adaptive machine translation—offer translations for segments from a connected MT engine as autosuggest pop-ups, learning from user input.
5. Electronic dictionaries—allow term search inside the tool and track usage statistics.
6. Automated Quality Assurance (AutoQA)—tools devised for scanning bilingual texts and detecting translation errors. Although they usually come built in within CAT tools, there are instances of standalone AutoQA software equipped with additional functionality.

CAT tools can be categorized in various ways. One of the first and most long-established ways to differentiate them is based on the extent to which the translation process is automated: Machine-Aided Human Translation (MAHT), characterized by human translation augmented with computer tools, and Human-Aided Machine Translation (HAMT), where the computer the translating agent, while the human assists with resolving translation-related problems (Sin-wai, 2017:167). However, a certain gray area lies within this distinction—sometimes it is difficult to determine the levels of cooperation between translators and computers in respect to the two categories. CAT

software can also be differentiated based on which operating systems they run on—some have cross-platform support, while others run exclusively on MS Windows, Linux, or OS X.

For industry practitioners, perhaps the most pertinent differentiation would be on free and paid CAT tools. Sin-wai's study numbered a total of 25 free CAT systems, some of which are accessible online and some downloadable from the internet (2017:168). The study found that—although are very accessible, easy to setup, and built with fairly intuitive user interfaces—their reliability overall is rather worrisome in regard to their versions, source codes, and supported operating systems. Some of the software were in fact copies of older versions of other software. Although limited features and functionality compared to paid CAT tools is undoubted, Sin-wai asserts that some of the free software are "fully functional and can adequately meet the needs of ad hoc users", despite their shortcomings (2017:169).

Among the industry leaders of paid CAT tools are: SDL Trados Studio, MemoQ, Wordfast, Déjà Vu and Across (PoliLingua). The general consensus in the translators' community it is up to everyone's individual preference to pick a personal favorite because they share the basic functionality—the rest depends on the translators' workflow and habits. That is to say that companies, given that they acquire discount deals for collaboration editions, usually provide in-house translators with software licenses, thus expecting prior experience with the tool. From the perspective of freelancer translators, who undoubtedly amount to the majority of the industry, it is not convenient to invest in a new license for a high-end industry standard CAT tools—as the MemoQ and SDL Trados' prices for individual licenses start at around €600, while annual new version upgrading costs around additional €200. However, Serbovets notes a few methods for acquiring the tools for a lower price: "keep track of discounts at the website of your CAT of choice (they are often somewhere in between 15% and 25%), register on Proz.com—they sometimes hold a so-called group buying there, which will give you a chance to get your CAT at wholesale price ( . . . ), find a local retailer—chances are it will offer your CAT a local price, which may be lower than the quotes on the website" (2016). Another problem identified in the industry lies in that certain agencies pay translators by hourly rates, or meaning that utilizing a CAT tool would cut down on translation time and, consequently, the salary—which ultimately defeats the purpose of the CAT tools' inherent productivity. Some agencies pay discounted rates for text segments matching previous translations in the translation memory, in some cases 60% for partial matches and even 0% for exact matches (Fox, 2013).

## 3.2. Machine translation

Having its roots back in the late 1940s, the first instances of machine translation (MT) marked the dawn of translation technology. The post-war period saw substantial investment in translation technology development and research. However, a major setback was faced in 1966, when the Automatic Language Processing Advisory Committee's (ALPAC) longitudinal research on the state of machine translation established that MT had no prospective utility, while being considerably costlier than human translation (Sin-wai, 2017:2). As a result, MT development faced a significant cut in funding, and was surpassed by CAT tools as the industry standard to this day. But today, with artificial intelligence and automation becoming the top buzzwords when it comes to current technology trends, MT is the most researched and anticipated field of translation technology. Nonetheless, its current model still remains heavily reliant on human assistance.

Machine translation generally refers to the automatized process of translating a source language to target language—MT software may or may not require human involvement in the process, although its principal is to fully automate the translation process. In terms of language processing, there are several different types of machine translation.

### 3.2.1. Rule-based Machine Translation (RBMT)

Rule-based Machine Translation (RBMT), also referred to as Knowledge-based Translation, is the first known commercial instance of MT, which bases word placement on the linguistic knowledge of source and target languages collected from dictionary and grammar databases. Those rules are curated by language experts and developers who establish connections between the target and source languages in order to allow the system to differentiate context of terms, while allowing users to manually improve translation through editing translation lexicons (Omniscien). Such system actually demonstrated the complexities of languages and their correspondence, which RBMT wasn't capable of processing in a more efficient manner, namely through required extensive post-editing, compared to human-driven translation processes.

### 3.2.2. Statistical Machine Translation (SMT)

Statistical Machine Translation (SMT) revolves around the statistical analysis of sizable volumes of bilingual parallel texts, in other words, existing human translations from source to target language. Compared to RMBT processing, which is predominantly word-based and requires manual definition of linguistic rules, modern SMT systems use a phrased-based approach, meaning that their focus is on exploring countless possible corresponding sequences of words, i.e., phrases, between the two languages (Sin-wai, 2015:201). Those phrases are not phrases in the linguistic common sense, but rather phrases derived statistically from the processed bilingual texts. Thus the statistical model notes the correspondences between translated phrases and estimates which would be the likeliest result. The system is flawed when it comes to taking context of the phrases in account, and tends to cause certain specific unpredictable errors. Until recently, the flaws could be observed by virtually anyone, since the world's most famous free-to-use MT service, Google Translate, operated on an SMT engine exclusively, before additionally implementing Neural Machine Translation (NMT) in 2016 (Berndsen, 2019).

### 3.2.3. Neural Machine Translation (NMT)

Currently Neural Machine Translation (NMT) stands out as the most successful MT algorithm to carry out translation, having not only Google, but also Microsoft and Amazon implementing it into their translation engines (Marr, 2018). It consists of an artificial neural network[1] of nodes capable of holding words, phrases, as well as longer sentences, as well as their correspondents in the target language. The algorithm relies on deep learning, which is a process of training the system to be capable of translation between any two languages in a manner similar to human understanding of language, by analyzing a very large volume of translated expressions, without the requirement of applying linguistic rules (Cheng, 2019). Compared to its predecessors, NMT is capable of delivering translation that achieves high scores on standardized evaluation systems at a considerably faster rate, while using up less technical resources and memory (Venkatesan, 2018:40). The pace at which NMT is advancing acts intimidating to the traditional translation industry. The number of research body has shown a tremendous increase over the last 3 years—

[1] An artificial neural network is „a computational learning system that uses a network of functions to understand and translate a data input of one form into a desired output, usually in another form" (DeepAI). Its model is anthropomorphic—meaning it is based on how human neurons react to sensory stimuli.

with estimated thousands of active researchers, the first half of the year 2018 counted 196 NMT-related publications, as opposed to 91 over the same period in 2017 (Diño, 2018). What is more, One Hour Translation CEO Ofer Shoshan, one of the global leaders in professional translation service, asserts that: "within one to three years, neural machine technology (NMT) translators will carry out more than 50% of the work handled by the $40 billion market" (Marr, 2018). That is to say that the current costs of NMT training and required human post-editing efforts still not render its superiority to human translation (with the assistance of CAT tools). It is also worth noting that the overwhelming majority of NMT systems are trained with most-spoken languages, thus raising a question of how long will it actually take for NMT systems to become more efficient at handling low-resource languages—for instance, English to Croatian translation, let alone from other source languages.

## 4. Speech technology

Prior to recent breakthroughs in technology, having conversations with computers was merely something associated with science fiction, or the dystopic methods of governments spying on its citizens. At this point, with the growing number of households equipped with surveillance cameras, automated lighting control systems, fingerprint scanners, and so on, voice controlled utilities are steadily becoming the norm of everyday life. Since the introduction of Siri in 2011, other voice assistants, such as Amazon's Alexa and Google Assistant, are already capable of maintaining basic conversations with humans, and fast on picking up habits and preferences of their users. Speech technology is a booming industry, smart speaker devices sales are on the rise, while some estimates claim that currently 1.5 billion virtual assistant devices are in use, and by 2023 that number is expected to come close to 8 billion—"that represents annual compound growth of more than 25 percent" (Sterling, 2019). Advances in the internet and cloud computing allow larger volumes of data to train speech recognition systems, endowing them with more capabilities at this very instant.

Although they still may be used interchangeably, speech recognition and voice recognition are becoming two separate terms in recent technological and academic discourse. Speech recognition focuses on converting the audio to word data, which establishes its evaluation criteria for accuracy and speed (Kikel, 2019). On the other hand, voice recognition aims to identify who is the speaker—locking on the speaker's speech patterns and vocal physiology. Therefore, speech recognition is used in transcription, hands-free technology, translation, and so on, while voice recognition is applied in voice assistants, speaker verification and identification systems, etc.

### 4.1. Automatic Speech Recognition (ASR)

Today, speech recognition technology is synonymized with Automatic Speech Recognition (ASR), as well as with voice-to-text, speech to text (STT), and automatic voice recognition (AVR), all of which basically refer to the process of digitalizing human voice using a microphone, or a telephone. It has been a field actively researched for over five decades, while over the past two

decades, progress has been made in significant reduction of error rates in ASR systems, allowing them widespread usage today.

On the surface, the process may seem quite straightforward. However, ARS systems consist of four main components—signal processing and feature extraction, acoustic model (AM), language model (LM), and hypothesis search (Yu and Deng, 2014:23). The recorded audio signal is handled by the signal processing and feature extraction component, which eliminate noises, equalize the volume, and extract vectors used by the following models. The AM implements acoustics and phonetics-related knowledge, while the LM breaks down the recording into phonemes, all of which are sequentially analyzed. During the analysis, by means of statistical probability, the ASR system deduces whole words and sentences later on (Yu and Deng, 2014:23). A contemporary optimal ASR system requires working with massive vocabulary, free-form conversations[2], and mixed languages—recently such hurdles have been overcome with the integration of aforementioned deep neural networks into the systems.

### 4.2. The advantages and disadvantages of speech input

Sin-wai identifies some advantages to using a speech input method instead of typing (2017:264). First and foremost, speech comes across as natural since it does not require any additional equipment, training or abilities, except for a recording device. Although newer generations have been brought up surrounded by computers, allowing to expect their mastery at typing, smartphones have come to become the device of choice for future generations—which may cause a downward trend in typing skills, since smartphones have essentially pioneered the voice assistant utility. The second advantage to speech input would be its convenience, as it enables communicating in a fast and pleasant manner, allowing up to 210 English words per minute (WPM), without putting a strain on the eyes and hands. To put it into perspective, the average typing speed is considered about 40 WPM, while individuals deemed as the world's fastest documented typers achieved over 200 WPM (Leonard, 2019). Finally, speech is universal, given that virtually anyone is able to speak. What is more, Rapp et al. have identified a difference in the way human brain processes talking and writing, which may surmise possible cognitive benefits to speech input (2015). A recent study on people who lost their voices has discovered that it is possible to record their neural

[2] A free-form, or free-style, conversation refers to dialogue that comes about naturally in social and professional human interaction (Ram et al., 2018:1).

activity and decode the information to speech, which implies future alternative input methods that could bypass speech (Anumanchipalli et al., 2019:493).

However, speech input method still faces some challenges impeding it from becoming the norm. The biggest problem is its inaccuracy of word recognition. Even in case of ideal speakers, with standardized accents, the software is still yet to decipher given context of language on the same level as humans, while error correction is a painstaking process. For instance, the most common error is the inability to discern between homophones, or homonyms—it often confuses 'right' with 'write', 'there' with 'their', and so on. In addition, the bigger the vocabulary is, the harder it is for the system to differentiate between words (Gayar and Suen, 2018:22). Despite ASR increasingly becoming robust to poor acoustic conditions, dictation still requires an environment with minimal background noise in order to process the recorded audio properly. Further issues arise with the time required for training the software to get acquainted with the user's speech patterns and voice quality. It may take the system a long time to progress to a desired level of efficiency, which could ultimately sacrifice productivity compared to using the keyboard all along. Using system commands and inserting punctuation mark is considerably faster when using keys, while dictation, while slower, it would take additional time to get accustomed to the practice. ASR systems in their current form entail adopting a rather specific approach in order to be used efficiently.

### 4.3. Application of speech technology

Speech technology is already well established in call centers and customer service departments, where users are able to browse menus by dictating numbers according to the instructions. With the ongoing technological advancements, speech technology is entering "the realms of finance, HR, marketing, and even public transportation with the goal of bringing down business costs, simplifying outdated processes, and increasing overall efficiency" (Van der Velde, 2019). The growth of the speech recognition industry, with its current worth at 55 billion USD, and estimated growth rate of 11% from 2016 to 2024, attests to that (De Jesus, 2019). What is specifically considered to further expand the overall speech technology market is Virtual Reality (VR)—in 2007 Facebook added speech recognition to its VR platform, Oculus Rift (Grand View Research, 2018).

Within healthcare, law enforcement and legal sectors, implementing speech technology through transcription applications is a growing trend. For example, Robin Healthcare is a young startup

that focuses on developing a speaker device capable of recording physicians' speech without specialized dictation, transcribe the dictated words, and produce formatted clinical notes that are directly recorded in the electronic health record system (De Jesus, 2019). Nuance, a leading company in conversational artificial intelligence, has developed the voice recognition tool Dragon Law Enforcement, already used by thousands of law enforcement officers in the US for report writing and critical note-taking, which can not only greatly help in lessening enormous quantities of paperwork, but also potentially prevent police ambushes and similar life-threatening situations (Condon, 2018).

### 4.3.1. Integration with CAT tools

Despite the concurrent boom of both translation industry and speech technology industry, tangible advances regarding speech recognition implementation in CAT tools are yet to be made. Among their other products, Nuance developed Dragon Home, which functions as a speech input method utility for Microsoft Windows and OS X systems, claimed to be three times faster than typing, with recognition accuracy of 99% (Nuance). Therefore, some translators use speech recognition software with the purpose of faster word input in a CAT tool. In 2017, a survey on Proz.com, home to the biggest translators' internet community, has shown that over 10% translators use CAT tools in conjunction with speech recognition software, the most preferred being Dragon suite software (Peleman, 2017). The survey identified MemoQ as the CAT tool most compatible with Dragon software, namely for word accuracy and convenient correcting. However, in 2018 MemoQ announced the 'Hey memoQ' feature, which added dictation support relying exclusively on pairing with an iPhone or iPad (MemoQ). Besides the addition of voice commands, its stand-out functionality is language support for more than 30 languages, including Croatian.

Besides MemoQ, the only other commercially available CAT tool that recently implemented an ASR system is MateCat, basing the feature on the Google Speech API—indicating that CAT tool developers still do not perceive developing an internal ASR engine worthwhile (MateCat). The current workaround method adopted is combining with Dragon software, although it proves problematic when it comes to the limited number of available languages, and the users perceive the integration with CAT tools ineffective (Teixeira et al., 2019). On top of that, Teixeira et al. identify the growth of popularity of the use of ASR for translation, while in its current form it does

not come across as helpful for a lot of users, adding that the inability of the ASR engine to properly recognize speech is what most often causes problems (2019).

## 5. Methodology

### 5.1. Rationale and research questions

The aim of this research was to determine whether speech recognition is a more efficient input method than typing in the translation process—assessing primarily speed and accuracy. Additionally, the goal was to establish the level of adequacy of the integration of ASR with CAT tools in its current instance. The CAT tool used was MemoQ, as it is the first industry standard software with speech recognition support—through pairing with an iPhone.

The research questions we aimed to answer were the following:

1. Are there statistically significant differences in the overall quality of dictated translation between students and professional translators?
2. Are there statistically significant differences in the typed WPM between students and professional translators?
3. Are there statistically significant differences in the overall quality of all respondents' typed and dictated translations?
4. Are there statistically significant differences between the total duration of typed and dictated translation?
5. Are there statistically significant differences between the post-editing duration of typed and dictated translation?
6. Does a higher typed WPM predict overall faster typed translation?
7. Does a higher dictated WPM predict overall faster dictated translation?

### 5.2. Research design

In this section we will present the research design in steps, our corpus, respondents and the statistical test used to analyze the data.

The corpus of the research consisted of a test carried out in 4 phases. 10 respondents, consisting of a group of 5 students, and 5 professional translators, were given the following tasks, all computer-based:

1. WPM typing test
2. WPM dictation test
3. Typed English to Croatian translation test
4. Dictated English to Croatian translation test

Test 1 was self-conducted through a free online WPM test. Among various WPM tests, 10fastfingers was chosen, as it had a version in the Croatian language, used the standardized measurement of WPM[3], and provided all relevant metrics upon completing every test run. The test lists out random words to be typed by the user, ending a minute after the first keystroke. Given the unrepresentative nature of using a non-preferred laptop keyboard to perform the test, respondents were allowed 3 runs, of which their best performance was marked down. The most relevant metric was accuracy—the ratio of correct and total words, while respondents were placed in the following categories according to the number of correct words typed:

1. Slow (0-25 WPM)
2. Average (25-34 WPM)
3. Fluent (45-60 WPM)
4. Fast (60-80 WPM)
5. Pro (80+ WPM)

[3] WPM, according to the international standard, is the quotient of the total number of typed characters divided by 5 (Typing Speed Test).

In order to measure the relevant metrics for dictation, Test 2 consisted of a Croatian sample text to be dictated in 1 minute, while being timed externally.

> Sve veća složenost i povezanost unutar i između društava postale su glavne odlike suvremenog svijeta. One utječu na dijalog s građanima i oblikuju alate za informiranje javnosti. Kako se moć sve više globalizira, država prestaje biti jedini dionik u sustavu, usprkos pokušajima povratka nacionalnim rješenjima. Dobivanje podrške javnosti u vremenu društvenih promjena zahtijeva jasno, koherentno i kritičko sagledavanje alata za aktivno uključivanje građana. Stoga je radikalan postupak transformacije koji je u tijeku nužno sagledati u odgovarajućem kontekstu. Brige građana rezultat su napetosti između suprotstavljenih polova slobode naprama sigurnosti s jedne strane te solidarnosti naprama izolacionizmu s druge. One se odnose na pitanja identiteta, državljanstva, granica, demokracije i dijaloga te zahtijevaju jasne i konkretne odgovore. Suradnja s građanima odnosi se na koncept zajednice, koja uključuje lokalne, regionalne, nacionalne i međunarodne kontekste u kojima oni žive, kako bi se stvorio zajednički javni prostor u kojem pojedinci mogu surađivati na temelju zajedničkih vrijednosti.

Figure 1. Dictation test text

Measuring the relevant metrics required a manual approach. The dictated texts were head-to-head compared with the original, while the correct and wrong entries were counted. As the ASR system in some cases displayed drastic inconsistency in picking up words, an additional metric of omitted words was added. The WPM was calculated by dividing the dictated character count with 5.

Test 3 was rather straightforward—respondents were required to translate a text from English to Croatian. The text was of intermediate difficulty, at standard translation page length.

In most parts of Europe, many people speak another language to a higher degree. For instance, people in Luxembourg or Switzerland are considered bilingual. Many other people across Europe speak English to a very high degree – about 47% of young people have a very good understanding of English. However, this is not the case in the UK—where only 9% of young people can speak a foreign language to a higher degree. This does not mean that all British people are monoglots—in fact, there is a lot of them who speak Urdu, Farsi, Welsh, or even Polish, at home. The problem lies with what should be the second language for British people. It is obvious that other people across Europe study English as a second language, as that is the most common language used as a lingua franca in international businesses. With about 1.2 million native speakers, Chinese language is the number one most spoken language in the world. In fact, China has become the world's largest source of overseas students and the third most popular destination for studying abroad, after the US and UK. It is believed that China hosts 8% of all the world's international students–which amounts to about 4.5 million students on the go, in a given moment. In 2016, there were 545,500 Chinese students studying abroad, which is an increase of 36.26 percent compared with the data in 2012. It seems that the global economy is shifting away from the English-speaking world. Since 1975, the English-speaking share of global GDP has fallen significantly and will continue to fall.

Figure 2. Typed translation text

They were allowed to use their preferred workflow, and research potential unknown terms prior to beginning of translation. Time measurement was self-conducted, while noting down the times for translating and post-editing was required. Upon completion, respondents saved the final version of the translation on the computer, which was evaluated later.

Test 4 also consisted of translating a text of intermediate difficulty, at standard translation page length from English to Croatian, however, this time with respondents dictating the translation into an iPhone paired with MemoQ.

European culture is as much part of our values as it is shaped by them. In particular, audiovisual content produced in the EU reflects our rich and diverse cultural and linguistic heritage. For video content, EU rules make sure TV broadcasters continue to diffuse European work, with an obligation to dedicate at least 50% airtime to European and national content. That way you can have access to a wide range of diverse content that still speaks to you. Not only traditional TV broadcasters but also video sharing platforms need to protect minors from harmful content, promote European works and adhere to advertising rules. Moreover, 30% of content in video-on-demand catalogues need to be European works. When travelling to another EU country, citizens can since 1 April 2018 access any online services they have paid for or subscribed to at home. This means that you can continue listening to music, playing games, watching sports and never miss an episode of your favourite show, wherever you are in the EU. Cultural heritage breathes a new life with digital technologies and the internet. The citizens have now unprecedented opportunities to access cultural material, while the institutions can reach out to broader audiences, engage new users and develop creative and accessible content for leisure and education. Digitised cultural archives gives access to 53 million items including image, text, sound, and video material from the collections of 3,700 libraries, archives, museums, galleries and audio-visual collections across

Figure 3. Dictated translation text

This time, to provide a comfortable setting and reduce background noise while translating, respondents were allowed to carry out the task in private. Again, they noted down the translation and post-editing times, and saved the final version for later evaluation.

The assessment criteria used for evaluating texts were based on the criteria applied for translation evaluation within EU institutions. In total, 10 assessment criteria were established—1, 2, or 4 points were deducted for minor mistakes, and a double amount for major mistakes, while the maximum score was capped at 80 points. The types of mistakes were as followed:

1. Meaning
    - Minor (-4 points): general understandability of the text not affected, lack of precision, errors of lexical and factual accuracy; minor distortion of meaning
    - Major (-8 points): original meaning changed; refers to mistranslations, nonsense, severe errors of interference or paraphrase, literal translations meaningless in the context, misunderstanding of a part of the text

2. Grammar
    - Minor (-2 points): general understandability not affected (i.e. overuse of passive, wrong preposition, subject-predicator agreement, case agreement, conjunction, word order)
    - Major (-4 points): tense or mood misuse, results in unintended interpretation, indicates inadequate command of the target language's grammatical structure

3. Terminology
    - Minor (-2 points): wrong usage or failure to use a well-established basic term
    - Major (-4 points): wrong usage or failure to use a well-established term from the particular specialist field entered in the glossary
4. Clarity, consistency and register
    - Minor (-2 points): clumsy translation, inappropriate register, inappropriate collocations, loss of idiom or metaphor in the target language; not affecting the readability of the text
    - Major (-4 points): clumsy translation, inappropriate register, severe lack of clarity; affecting the readability of the text
5. Addition
    - Minor (-2 points): superfluous addition; meaning of the original not seriously affected
    - Major (-4 points): meaning of the original altered
6. Punctuation and formatting
    - Minor (-2 points): minor infringement of rules of punctuation, orthography and capitalization
    - Major (-4 points): infringement of rules on punctuation resulting in interpretation other than intended
7. Spelling
    - Minor (-1 points): minor misspelling or typo
    - Major (-2 points): serious misspelling resulting in unintended interpretation
8. Omission
    - Minor (-4 points): each partially translated line of text
    - Major (-8 points): each full line of the original text not translated

The remaining two categories focused on deducting and adding extra points for overall quality of the text:

9. Extra points deducted for overall quality
    - Minor (-2 points): minor formatting error (font type, size, layout, images, references), occasional lack of coherence
    - Major (-4 points): inconsistent formatting through the text, text reads like a translation; substandard translation
10. Extra points added for overall quality
    - Minor (-2 points): exceptional handling of difficult words, phrases or sentences (maximum 10 extra points)
    - Major (-4 points): translation of the highest quality; maximum or publication standard

With the score achieved, the translations were graded according to the criteria for overall assessment, while the passing threshold was 40 points:

1. Unacceptable (0–39): totally inadequate; substandard
2. Inadequate (40–53): substandard translation
3. Good (54–63): adequate, student standard
4. Very good (64–73): almost completely successful; minimum professional standard
5. Excellent (74–80): successful, maximum or publication standard

In order to provide an unbiased evaluation, two evaluators assessed the translations, ultimately using the average of scores and grades. Prior to and after the testing, respondents were briefly surveyed.

The pre-testing survey aimed to gather relevant information on the respondents' experience in translation, along with their level of familiarization with speech recognition software and CAT tools. They were also asked to indicate whether they perceive themselves as fast typers or coherent speakers.

Testing the efficiency of voice recognition software in translation

Faculty of Humanities and Social Sciences, University of Osijek

Pre-testing survey

*This survey is part of the research for a Master's thesis at the Department of English at the University of Osijek.*

*It should take less than 10 minutes, and your answers are completely anonymous.*

Translator:

1. Occupation: Student of translation / Professional translator
    a. If student, indicate the current year of study: 1 / 2
2. Professional translation experience: Yes / No
    a. If yes, for how long: _____
3. Have you ever used voice recognition software? Yes / No
    a. If yes, have you used it professionally? Yes / No
        i. If yes, which software have you used? _____
    b. If yes, have you used it for private purposes? Yes / No
        i. If yes, which software have you used? _____
4. Do you use any CAT tools?
        i. If yes, which software do you use? _____
5. Would you consider yourself a fast typer? Yes / No
6. Would you consider yourself a coherent speaker? Yes / No

Figure 4. Pre-testing survey

In the post-testing survey, the respondents were questioned about their user experience with the MemoQ speech recognition feature. The first 3 questions consisted of an assessment scale for rating the respondents' impression, perceived user-friendliness, and comfort with using the feature. In question 4, they were asked to indicate which input method they perceive as more efficient in translation—currently, and in the future. In the remaining 2 open-ended questions, respondents were questioned about ideas on potential ways to improve the feature, and also about the problems and advantages they encountered while using the feature.

Testing the efficiency of voice recognition software in translation

Faculty of Humanities and Social Sciences, University of Osijek

Post-testing survey

1. How did you like working with MemoQ's voice recognition module?

| 1 – Hated it | 2 – Didn't like it | 3 – Neutral | 4 – Liked it | 5 – Loved it |
|---|---|---|---|---|
| | | | | |

2. Did you find the module easy to use?

| 1 – Not at all | 2 – No | 3 – Neutral | 4 – Yes | 5 – Extremely easy |
|---|---|---|---|---|
| | | | | |

3. Did you feel comfortable using the module?

| 1 – No | 2 – Neutral | 3 – Yes |
|---|---|---|
| | | |

4. Which method do you find more efficient in translation:
   a. Presently: Typing / Dictation
   b. In the future: Typing / Dictation

5. What do you think could be improved in the software? Any features that should be added?

6. What problems have you encountered? And what advantages?

Figure 5. Post-testing survey

As our sample size is small (m=n=5) and samples are random and independent, a nonparametric test is appropriate. To compare the results of students with the results of professionals, we used the Mann Whitney U Test, which is a "a nonparametric test that allows two groups or conditions or treatments to be compared without making the assumption that values are normally distributed" (Statistics Solutions). In order to calculate the scores, we used the Social Science Statistics online calculators. The developers assure their reliability: "The output of the calculators and tools

featured on this web site has been audited for accuracy against the output produced by a number of established statistics packages, including SPSS and Minitab" (Statistics Solutions).

To compare the compound results, whereby they correlate with the students and professional groups respectively, we used the Wilcoxon test. It is also a nonparametric test, used to measure differences between two treatments or conditions when their samples are in correlation—"in particular, it is suitable for evaluating the data from a repeated-measures design in a situation where the prerequisites for a dependent samples t-test are not met" (Social Science Statistics). We used it to compare the total duration and overall score of typed and dictated translation.

To calculate correlations, we used the Spearman's Rho Calculator—"a non-parametric test used to measure the strength of association between two variables, where the value r = 1 means a perfect positive correlation and the value r = -1 means a perfect negative correlation" (Social Science Statistics).

## 6. Data analysis

### 6.1. Metrics

We established a sizable set of metrics, considering the extent of the entire research. Apart from the pre-testing survey, some data was collected from the respondents' data sheets—where they indicated the total, correct, and wrong WPMs of the typing and dictation tests, along with the start/end time markers for typed and dictated translation and post-editing. Based on those metrics, the rest was calculated in a Microsoft Excel spreadsheet.

Pre-testing survey:

1. Professional translator/student status
2. Years of translation experience/study

In terms of the test phases:

1. WPM typing test
   - Typed WPM
   - Typed correct words
   - Typing accuracy
2. WPM dictation test
   - Dictated WPM
   - Dictated total words
   - Dictated correct words
   - Dictated omitted words
   - Dictating accuracy
3. Typed English to Croatian translation test
   - Typed translation total time
   - Typed translation post-editing time
   - Typed translation score (Evaluator 1)
   - Typed translation grade (Evaluator 1)
   - Typed translation score (Evaluator 2)
   - Typed translation grade (Evaluator 2)
   - Average (mean) typed translation score
   - Average (mean) typed translation grade

4. Dictated English to Croatian translation test
    - Dictated translation total time
    - Dictated translation post-editing time
    - Dictated translation score (Evaluator 1)
    - Dictated translation grade (Evaluator 1)
    - Dictated translation score (Evaluator 2)
    - Dictated translation grade (Evaluator 2)
    - Average dictated translation score
    - Average dictated translation grade
5. Overall
    - Average (mean) time of translation
    - Average (mean) time of post-editing
    - Average (mean) translation score

In addition, to compare the data of students with professionals, the average (mean) of all the metrics of the respective groups was measured.

### 6.2. Results

The hypotheses are given below, and the tests are run at the 5% level of significance (i.e., $\alpha=0.05$).

**Research question 1: Are there statistically significant differences in the overall quality of dictated translation between students and professional translators?**

Hypothesis: dictated student translations are of lower overall quality compared to the quality of dictated translation by professional translators.

| | STUDENT/ TRANSLATOR | EVAL 1 DICTATED SCORE | EVAL 1 DICTATED OVERALL | EVAL 2 DICTATED SCORE | EVAL 2 DICTATED OVERALL | AVG DICTATED SCORE | AVG DICTATED OVERALL |
|---|---|---|---|---|---|---|---|
| 2 | S | 57 | 3 | 59 | 3 | 58 | 3 |
| 4 | S | 47 | 2 | 51 | 3 | 49 | 3 |
| 5 | S | 53 | 3 | 49 | 2 | 51 | 3 |
| 7 | S | 48 | 2 | 42 | 2 | 45 | 2 |
| 8 | S | 48 | 2 | 52 | 3 | 50 | 3 |
| AVG (S) | | 50,6 | 2,4 | 50,6 | 2,6 | 50,6 | 2,5 |
| | | | | | | | |
| 1 | T | 67 | 4 | 66 | 4 | 66,5 | 4 |
| 3 | T | 78 | 5 | 78 | 5 | 78 | 5 |
| 9 | T | 72 | 5 | 70 | 5 | 71 | 5 |
| 10 | T | 63 | 4 | 61 | 4 | 62 | 4 |
| 16 | T | 60 | 4 | 64 | 4 | 62 | 4 |
| AVG (T) | | 68 | 4,4 | 67,8 | 4,4 | 67,9 | 4,4 |

Figure 6. An overview of the data spreadsheet

| Sample 1 | Sample 2 |
|---|---|
| 58 | 66.5 |
| 49 | 78 |
| 51 | 71 |
| 45 | 62 |
| 50 | 62 |

The *U*-value is 0. The critical value of *U* at $p < .05$ is 4. Therefore, the result is significant at $p < .05$.

The *z*-score is -2.50672. The *p*-value is .00604. The result is significant at $p < .05$.

Figure 7. Results of the Mann Whitney U Test for research question 1. Note the difference in the decimal separators between the Excel spreadsheet in Croatian and calculator in English

The test confirmed the hypothesis that students' translations was of significantly lower overall quality than of professional translators, which was expected since professionals have 20 years of actual translation experience in average, compared to the students' average 1.6 years of studies.

**Research question 2: Are there statistically significant differences in the typed WPM between students and professional translators?**

Hypothesis: Students have higher typed WPM compared to professional translators.

| STUDENT/ TRANSLATOR | TYPED WPM |
|---|---|
| S | 56 |
| S | 39 |
| S | 60 |
| S | 56 |
| S | 46 |
| | 51,4 |
| | |
| T | 102 |
| T | 70 |
| T | 76 |
| T | 37 |
| T | 39 |
| | 64,8 |

Figure 8. An overview of the data spreadsheet

| Sample 1 | Sample 2 |
|---|---|
| 56 | 102 |
| 39 | 70 |
| 60 | 76 |
| 56 | 37 |
| 46 | 39 |

The *U*-value is 9.5. The critical value of *U* at $p < .05$ is 4. Therefore, the result is *not* significant at $p < .05$.

The *z*-score is -0.52223. The *p*-value is .30153. The result is *not* significant at $p < .05$.

Figure 9. Results of the Mann Whitney U Test for research question 1

We have found that, contrary to our expectations, the average WPM of translators is higher than that of students, but the difference is not statistically significant, thus not confirming the hypothesis. Considering that the students are digital natives, they were expected to display higher proficiency in typing than professionals. However, we have to speculate some advantage for the professionals, since the keyboards used were provided by courtesy of the college, where 4 of them are full-time employees for a number of years, presumably using the same type of keyboard in everyday tasks. That is not to assert that the students use different keyboards—but, undoubtedly, the professionals have spent more time using them.

**Research question 3: Are there statistically significant differences in the overall quality of all respondents' typed and dictated translations?**

Hypothesis 1: There are no statistically significant differences between the overall quality of typed and dictated translation (i.e. two types of translation are equal).

Hypothesis 2: There are statistically significant differences between the overall quality of typed and dictated translation (i.e. two types of translation are not of equal duration; $\alpha$=0.05).

| AVG TYPED SCORE | AVG TYPED OVERALL | AVG DICTATED SCORE | AVG DICTATED OVERALL |
|---|---|---|---|
| 69,5 | 4 | 58 | 3 |
| 58 | 3 | 49 | 3 |
| 58 | 3 | 51 | 3 |
| 59 | 3 | 45 | 2 |
| 20 | 2 | 50 | 3 |
| **52,9** | **3** | **50,6** | **2,5** |
| | | | |
| 65,5 | 4 | 66,5 | 4 |
| 72 | 5 | 78 | 5 |
| 76,5 | 5 | 71 | 5 |
| 71,5 | 5 | 62 | 4 |
| 74,5 | 5 | 62 | 4 |
| **72** | **4,7** | **67,9** | **4,4** |
| | | | |
| 62,45 | 4 | 59,25 | 3 |

Figure 10. An overview of the data spreadsheet

## Wilcoxon Signed-Rank Test Calculator

**Success!**

*Explanation of results*

We have calculated both a *W*-value and *z*-value. If the size of *N* is at least 20 - see the Results Details box - then the distribution of the Wilcoxon *W* statistic tends to form a normal distribution. This means you can use the *z*-value to evaluate your hypothesis. If, on the other hand, the size of *N* is low, and particularly if it's below 10, you should use the *W*-value to evaluate your hypothesis.

You should also note that if a subject's difference score is zero - that is, if a subject has the same score in both treatment conditions - then the test discards the individual from the analysis and reduces the sample size. If you have a lot of ties, this procedure will undermine the reliability of the test (and also suggests that the requirement that the data is continuous has not been met).

| Treatment 1 | Treatment 2 | *Sign* | *Abs* | *R* | *Sign R* |
|---|---|---|---|---|---|
| 69.5 | 58 | 1 | 11.5 | 7 | 7 |
| 58 | 49 | 1 | 9 | 5 | 5 |
| 58 | 51 | 1 | 7 | 4 | 4 |
| 59 | 45 | 1 | 14 | 9 | 9 |
| 20 | 50 | -1 | 30 | 10 | -10 |
| 65.5 | 66.5 | -1 | 1 | 1 | -1 |
| 72 | 78 | -1 | 6 | 3 | -3 |
| 76.5 | 71 | 1 | 5.5 | 2 | 2 |
| 71.5 | 62 | 1 | 9.5 | 6 | 6 |
| 74.5 | 62 | 1 | 12.5 | 8 | 8 |

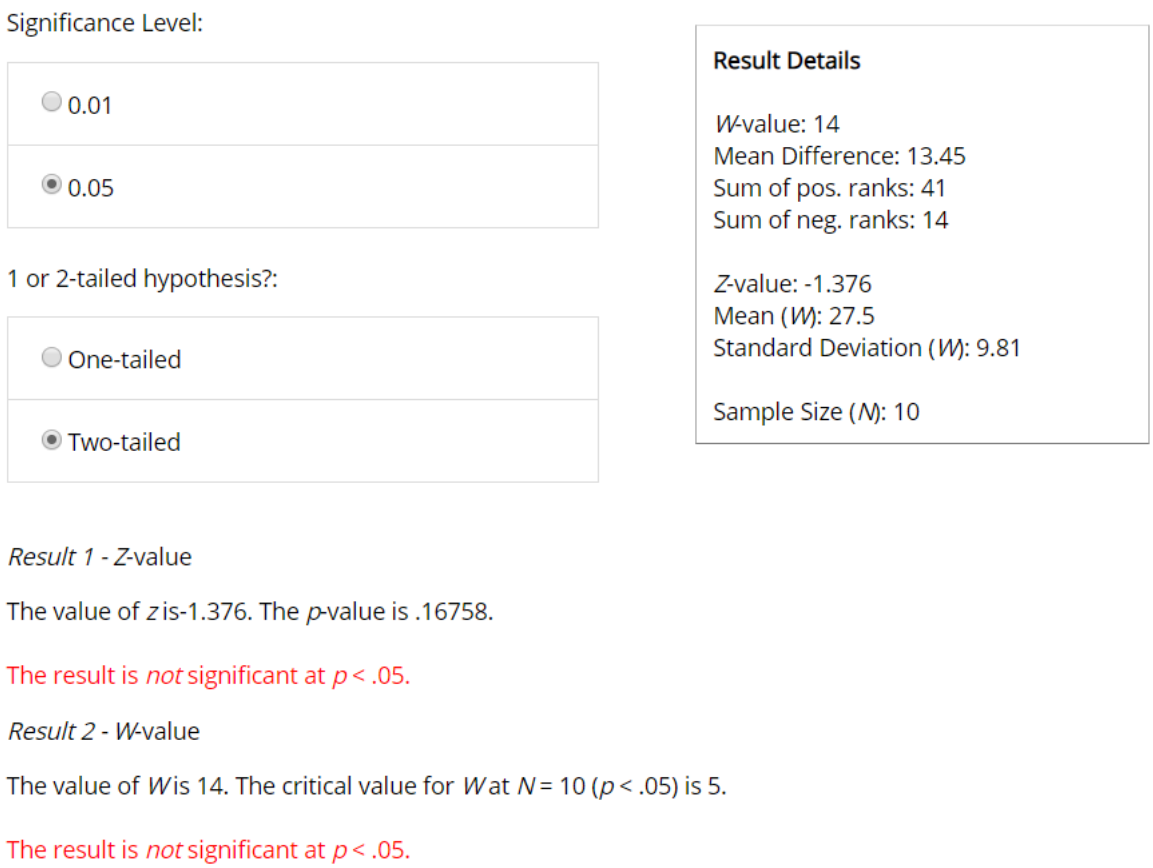


Figure 11. Results of the Wilcoxon test for research question 3

Due to the sample size of 10, which is considered low for this type of test, the W-value was used to evaluate the hypotheses. Although the average of typed translations is a grade higher than of dictated translations, the test found no significant difference in overall quality between the two of them, confirming hypothesis 1. A noticeably larger amount of formatting mistakes and typos was present in dictated translations, allowing to assume that the respondents using MemoQ for post-editing did not have automatic Croatian spell check enabled, causing such errors.

**Research question 4: Are there statistically significant differences between the total duration of typed and dictated translation?**

Hypothesis 1: There are no statistically significant differences between the total duration of typed and dictated translation (i.e. two types of translation are equal).

Hypothesis 2: There are statistically significant differences between the total duration of dictated and typed translation (i.e. two types of translation are not of equal duration; α=0.05).

| TYPED TRANSLATION TOTAL TIME | TYPED TRANSLATION POST-EDIT TIME | DICTATED TRANSLATION TOTAL TIME | DICTATED TRANSLATION POST-EDIT TIME |
|---|---|---|---|
| 18 | 8 | 20 | 10 |
| 20 | 4 | 24 | 9 |
| 18 | 0 | 21 | 19 |
| 8 | 1 | 13 | 7 |
| 26 | 6 | 29 | 6 |
| 18 | 3,8 | 21,4 | 10,2 |
| | | | |
| 15 | 8 | 17 | 12 |
| 31 | 10 | 57 | 57 |
| 33 | 9 | 46 | 37 |
| 24 | 6 | 25 | 13 |
| 28 | 10 | 22 | 12 |
| 26,2 | 8,6 | 33,4 | 26,2 |
| | | | |
| 22,1 | 6,2 | 27,4 | 18,2 |

Figure 12. An overview of the data spreadsheet

| Treatment 1 | Treatment 2 | *Sign* | *Abs* | *R* | *Sign R* |
|---|---|---|---|---|---|
| 18 | 20 | -1 | 2 | 2.5 | -2.5 |
| 20 | 24 | -1 | 4 | 6 | -6 |
| 18 | 21 | -1 | 3 | 4.5 | -4.5 |
| 8 | 13 | -1 | 5 | 7 | -7 |
| 26 | 29 | -1 | 3 | 4.5 | -4.5 |
| 15 | 17 | -1 | 2 | 2.5 | -2.5 |
| 31 | 57 | -1 | 26 | 10 | -10 |
| 33 | 46 | -1 | 13 | 9 | -9 |
| 24 | 25 | -1 | 1 | 1 | -1 |
| 28 | 22 | 1 | 6 | 8 | 8 |

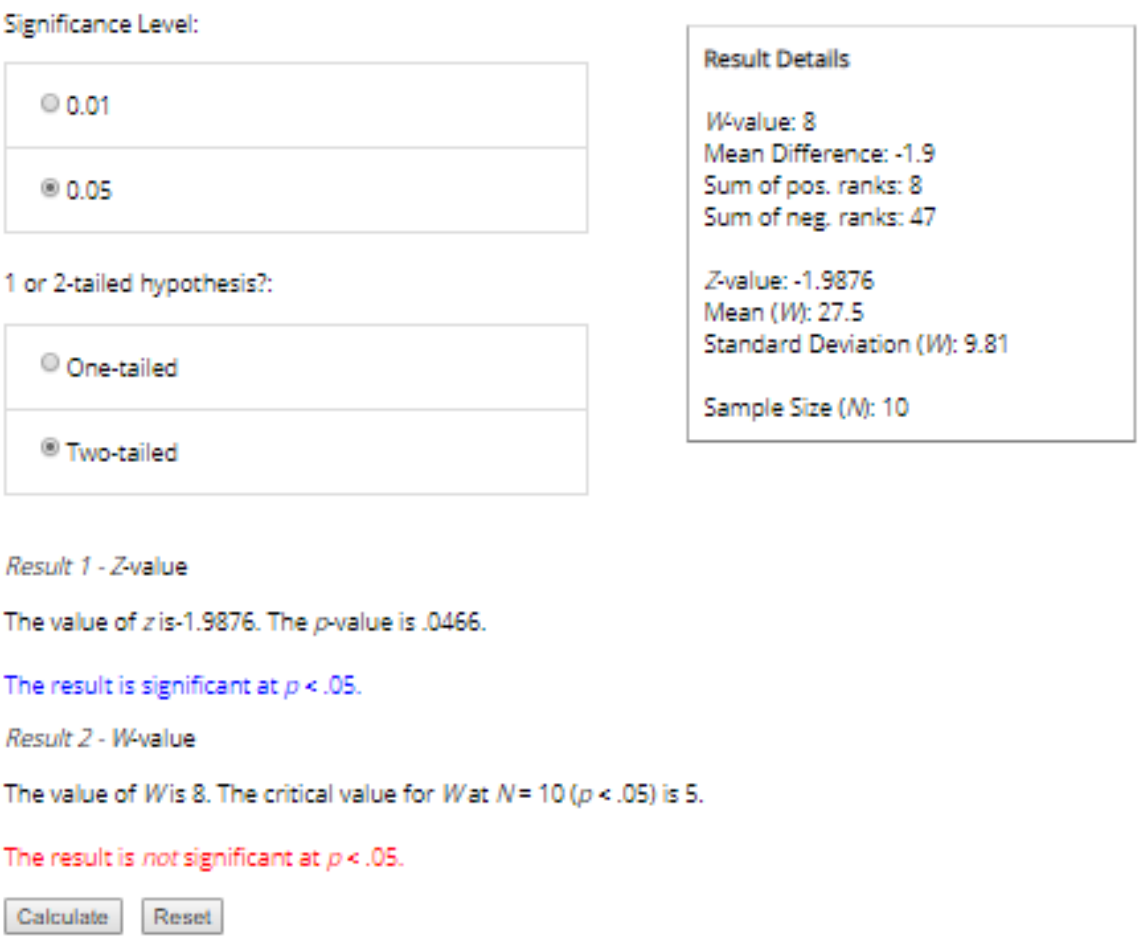


Figure 13. Results of the Wilcoxon test for research question 4

The test found no significant difference in the total duration for typed and dictated translation, confirming hypothesis 1. The result is rather unexpected, considering that the post-editing time for dictated translation was triple the amount of post-editing time for typed translation. It is also unexpected in regard to the WPM difference, with the average of 58.1 average typed WPM, compared to the average of 134.5 average dictated WPM. On the other hand, 8 of the respondents considered themselves as fast typers, which is confirmed by the overall average of 58.1 typed WPM (just 2 WPMs short of 60, which is the threshold for fast typers). A metric such as 'fast speaker' could not be established due to the ASR's inability to consistently pick up words in Croatian, which constrained the respondents' dictation speed—as a perceived method to facilitate word recognition—along with omitting an average of 8.5 words per dictated translation.

**Research question 5: Are there statistically significant differences between the post-editing duration of typed and dictated translation?**

Hypothesis 1: There are no statistically significant differences between the post-editing duration of typed and dictated translation (i.e. two types of post-editing are equal).

Hypothesis 2: There are statistically significant differences between the post-editing duration of dictated and typed translation (i.e. two types of post-editing are not of equal duration; α=0.05).

| TYPED TRANSLATION POST-EDIT TIME | DICTATED TRANSLATION POST-EDIT TIME |
|---|---|
| 8 | 10 |
| 4 | 9 |
| 0 | 19 |
| 1 | 7 |
| 6 | 6 |
| 3,8 | 10,2 |
| | |
| 8 | 12 |
| 10 | 57 |
| 9 | 37 |
| 6 | 13 |
| 10 | 12 |
| 8,6 | 26,2 |
| | |
| 6,2 | 18,2 |

Figure 14. An overview of the data spreadsheet

| Treatment 1 | Treatment 2 | *Sign* | *Abs* | *R* | *Sign R* |
|---|---|---|---|---|---|
| 8 | 10 | -1 | 2 | 1.5 | -1.5 |
| 4 | 9 | -1 | 5 | 4 | -4 |
| 0 | 19 | -1 | 19 | 7 | -7 |
| 1 | 7 | -1 | 6 | 5 | -5 |
| 6 | 6 | n/a | 0 | n/a | n/a |
| 8 | 12 | -1 | 4 | 3 | -3 |
| 10 | 57 | -1 | 47 | 9 | -9 |
| 9 | 37 | -1 | 28 | 8 | -8 |
| 6 | 13 | -1 | 7 | 6 | -6 |
| 10 | 12 | -1 | 2 | 1.5 | -1.5 |

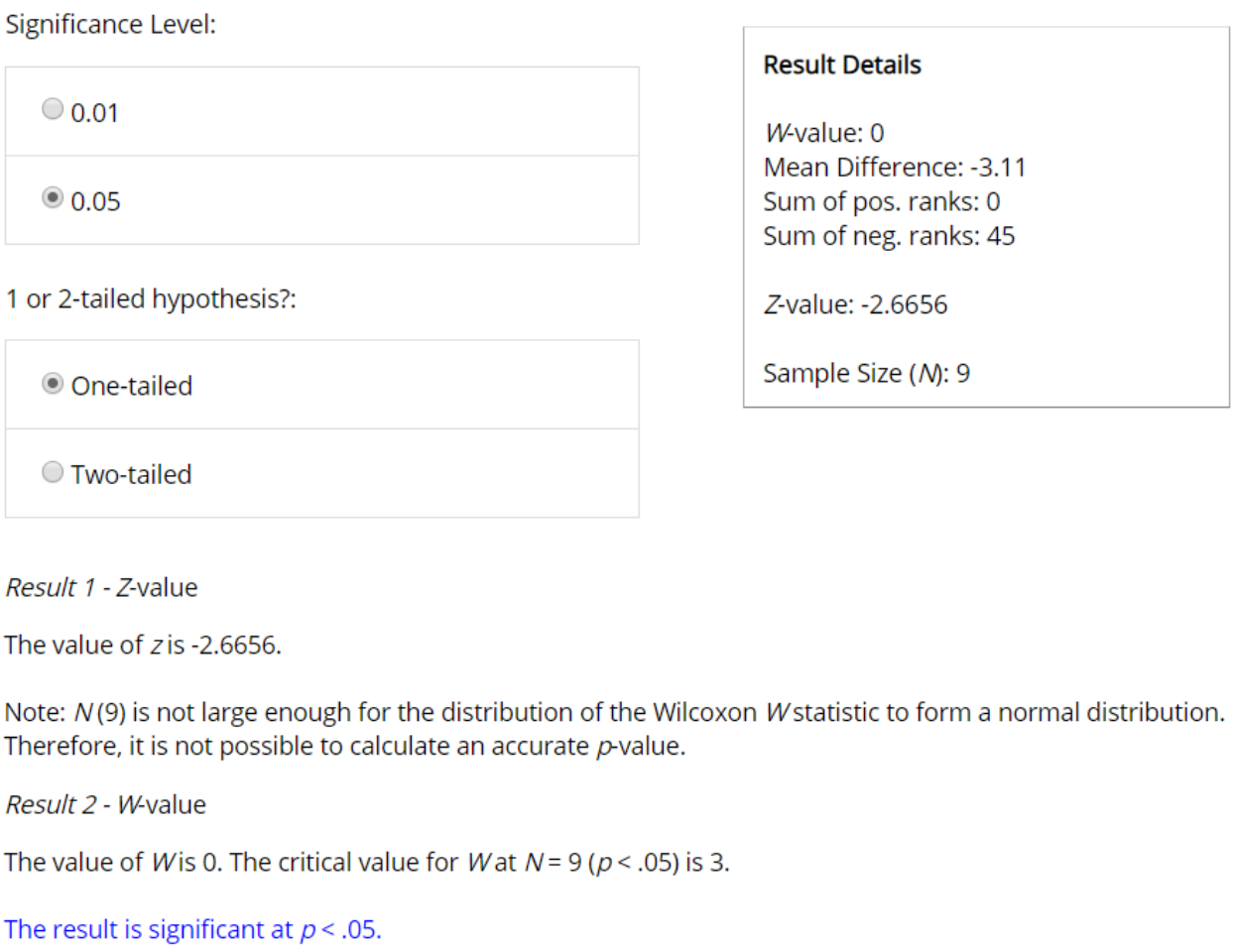


Figure 15. Results of the Wilcoxon test for research question 5

The test found significant a difference in the post-editing times, confirming the second hypothesis. It comes up as not much of a surprise—given that speech recognition, although faster, often picks up wrong words, or sometimes leaves them out. In particular, the rate of overall speech recognition accuracy (correct/total words ratio) was 70%, while typing accuracy was 91%, resulting in three times longer overall time for post-editing dictated texts than typed texts.

**Research question 6: does a higher typed WPM predict overall faster typed translation?**

| TYPED WPM | TYPED TRANSLATION TOTAL TIME |
|---|---|
| 56 | 18 |
| 39 | 20 |
| 60 | 18 |
| 56 | 8 |
| 46 | 26 |
| 51,4 | 18 |
| | |
| 102 | 15 |
| 70 | 31 |
| 76 | 33 |
| 37 | 24 |
| 39 | 28 |
| 64,8 | 26,2 |

Figure 16. An overview of the data spreadsheet

## Spearman's Rho Calculator

The value of $r_s$ is: -0.0581.

| X Values | Y Values |
|---|---|
| 56 | 18 |
| 39 | 20 |
| 60 | 18 |
| 56 | 8 |
| 46 | 26 |
| 102 | 15 |
| 70 | 31 |
| 76 | 33 |
| 37 | 24 |
| 39 | 28 |

$r_s$ = -0.0581, $p$ (2-tailed) = 0.87333.

By normal standards, the association between the two variables would not be considered statistically significant.

Figure 17. Results of the Spearman test for research question 6

Although we expected a higher typed WPM to be a predictor of faster translation typing input, the test found no significant correlation between higher typed WPM and overall lower typed translation time. It allows to deduce that the motoric ability does not outweigh the speed of translation as a mental activity.

**Research question 7: Does a higher dictated WPM predict overall faster dictated translation?**

| DICTATED WPM | DICTATED TRANSLATION TOTAL TIME |
|---|---|
| 87 | 20 |
| 116 | 24 |
| 143 | 21 |
| 159 | 13 |
| 165 | 29 |
| 134 | 21,4 |
| | |
| 140 | 17 |
| 148 | 57 |
| 141 | 46 |
| 136 | 25 |
| 110 | 22 |
| 135 | 33,4 |

Figure 18. An overview of the data spreadsheet

The value of $r_s$ is: 0.21212.

| X Values | Y Values |
|---|---|
| 87 | 20 |
| 116 | 24 |
| 143 | 21 |
| 159 | 13 |
| 165 | 29 |
| 140 | 17 |
| 148 | 57 |
| 141 | 46 |
| 136 | 25 |
| 110 | 22 |

$r_s$ = 0.21212, $p$ (2-tailed) = 0.55631.

By normal standards, the association between the two variables would not be considered statistically significant.

Figure 19. Results of the Spearman test for research question 7

As with typed WPM and translation, the test found no significant correlation between higher dictated WPM and overall lower dictated translation time. After the previous test, it does not come up as a surprise, especially when considering the ASR's inconsistencies.

**Research question 8: Does longer overall translation time predict translation of overall higher quality?**

| AVG TIME | AVG SCORE |
|---|---|
| 19 | 63,75 |
| 22 | 53,5 |
| 19,5 | 54,5 |
| 10,5 | 52 |
| 27,5 | 35 |
| 19,7 | 51,75 |
| | |
| 16 | 66 |
| 44 | 75 |
| 39,5 | 73,75 |
| 24,5 | 66,75 |
| 25 | 68,25 |
| 29,8 | 69,95 |

Figure 20. An overview of the data spreadsheet

The value of $r_s$ is: 0.53939.

| X Values | Y Values |
|---|---|
| 19 | 63.75 |
| 22 | 53.5 |
| 19.5 | 54.5 |
| 10.5 | 52 |
| 27.5 | 35 |
| 16 | 66 |
| 44 | 75 |
| 39.5 | 73.75 |
| 24.5 | 66.75 |
| 25 | 68.25 |

$r_s$ = 0.53939, $p$ (2-tailed) = 0.10759.

By normal standards, the association between the two variables would not be considered statistically significant.

Figure 21. Results of the Spearman test for research question 8

Despite the fact that the professionals spent additional 10 minutes on average, and achieved an almost 25% overall higher score compared to students, no significant correlation was found between longer overall translation time and translation of overall higher quality. The result is rather

surprising and incongruous with the numbers taken at face value, again suggesting the limitations regarding namely the small sample size, and broad generalization of scores and times perhaps.

In the post-testing survey, the respondents were given questions on their impression on using the MemoQ's speech recognition feature, and their attitude towards its efficiency. The respondents displayed an overall positive attitude toward using the feature, which was not fully anticipated, given that the feature is flawed with word recognition accuracy and requires a dedicated skillset for appropriate utilization. What was especially surprising was that none found the feature hard to use, while 2 respondents found it extremely easy.

The results are presented in the following graphs.

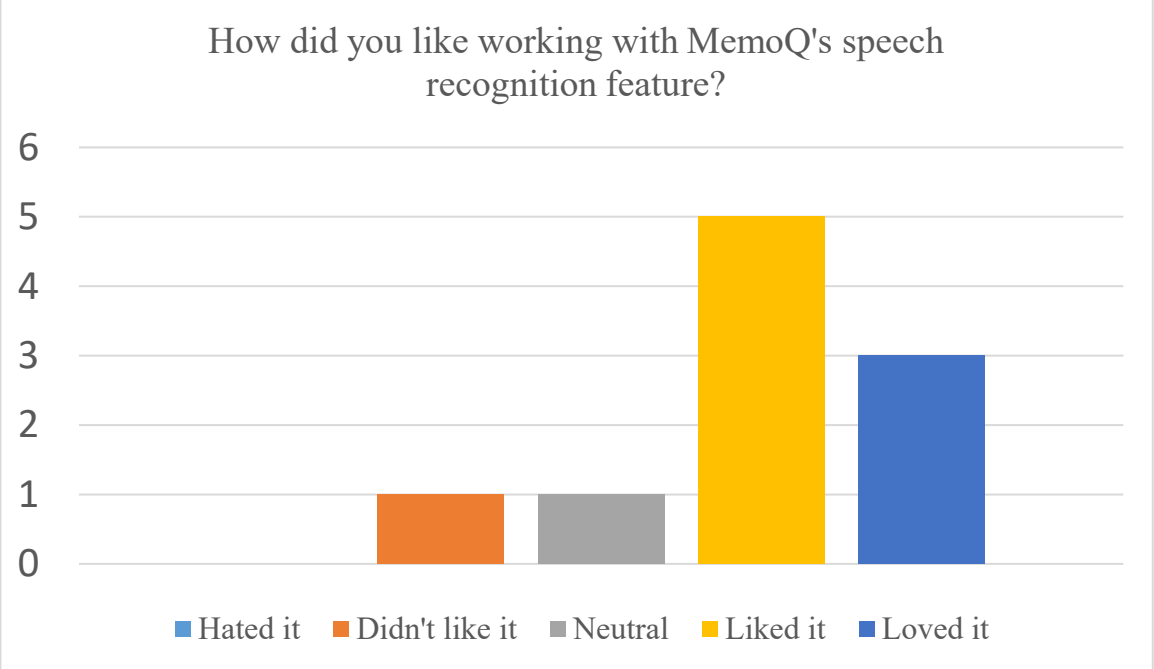


Figure 22. Post-survey question 1

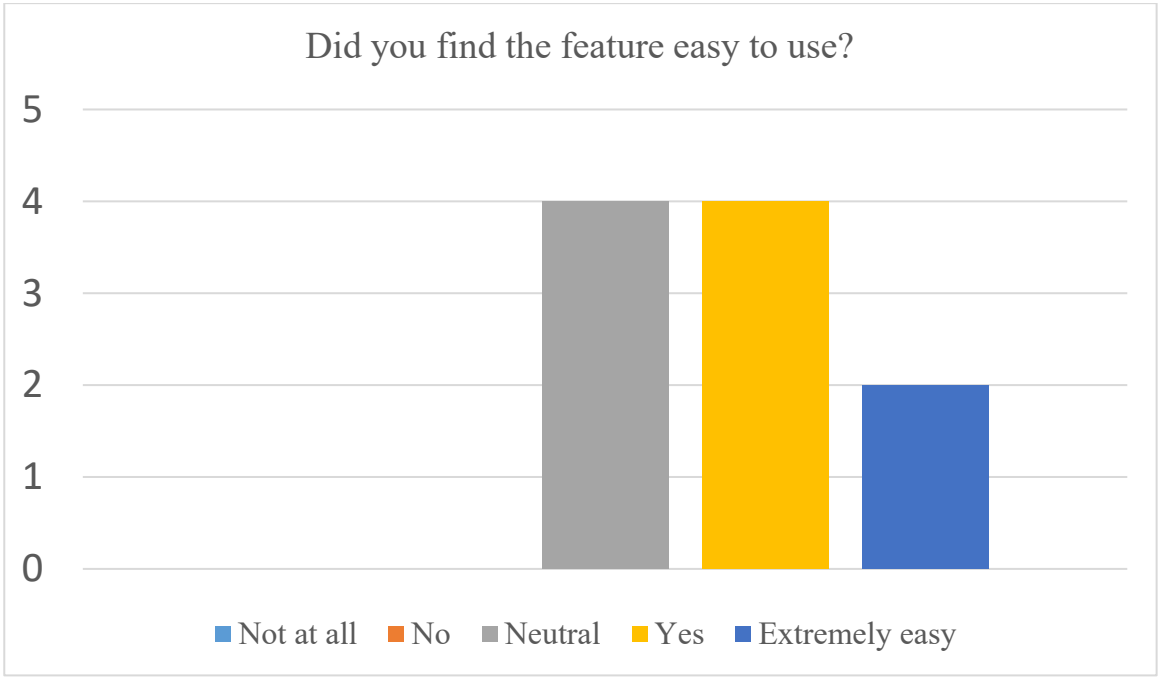


Figure 23. Post-survey question 2

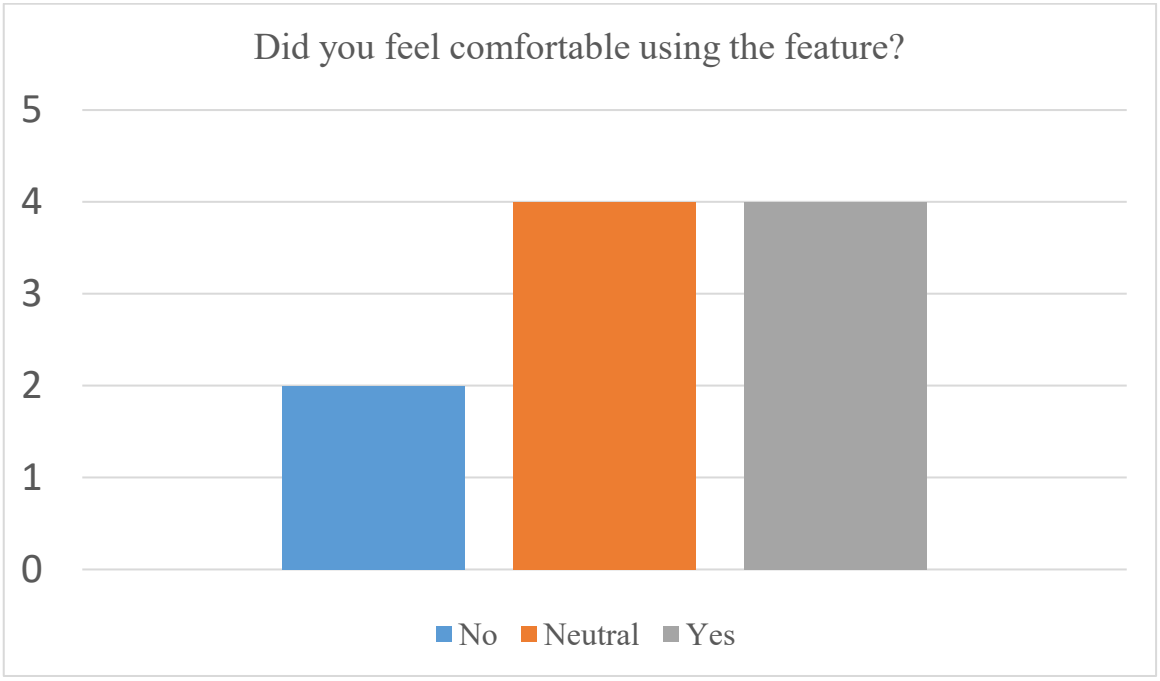


Figure 24. Post-survey question 4

Also, while acknowledging the typing input as the most efficient in translation currently, the overwhelming majority is aware of the implications of speech input disrupting the current translation process. It is apparent that this first instance of ASR-CAT tool integration earned the respondents' trust despite its flaws.

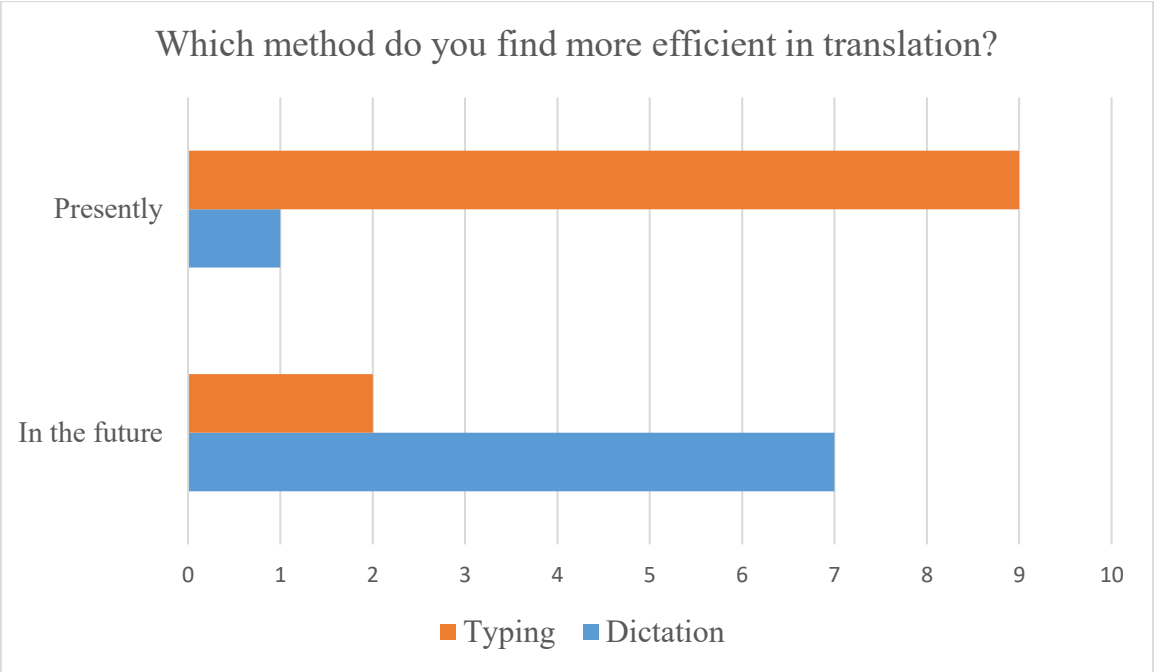


Figure 25. Post-survey question 5

In the next question the respondents were asked for ways to improve the features. Given its open-ended form, we received modest feedback—with only 7 respondents providing answers generally related to the need of improving speed and accuracy of ASR's functionality. On multiple occasions, respondents identified the necessity of ASR to adapt to its user, and to implement a functioning voice command system, in order to streamline navigating the document.

For the remaining question all respondents identified the problems and advantages encountered while using the feature. We can group them into the following, while an array of problems is apparent, compared to the advantages:

1. Problems
    - Inaccuracy
    - Issues related to the speed of word recognition
    - Lack of voice commands
    - Inconvenient device pairing
    - Longer post-editing times
    - Required adaptation to the user
    - Poor ability to distinguish Croatian cases
    - Overall bugs

2. Advantages
    - Faster input
    - Future potential
    - Appealing workflow

## 7. Conclusion

Generally, the research analysis found no statistically significant differences in efficiency when comparing typing and speech input method—with input speed and accuracy as the two main confounding factors of efficiency. Despite the fact that the respondents achieved higher scores and grades for typed translation, the test found no statistically significant differences in scores, compared to dictated translation. The similar happened when comparing the total duration of typed and dictated translation, with the numbers at face value implying typed translation as faster. However, despite the inconsistent accuracy of the speech recognition engine picking up words, additionally omitting an overall average of 8.5 words per translation—resulting in an overall three times longer duration of post-editing dictated translation than typed—no significant difference in the total duration was found. With the overall average of typed WPM amounting to less than half of the dictated WPM, it is apparent that the longer post-editing time for dictated translation evened out the score—while significant difference was found when comparing the duration of post-editing times. All in all, speech input is faster, but less accurate (70% versus 91% typed accuracy), thus resulting in a significantly longer post-editing duration—ultimately ending up only slightly slower compared to typing.

As for other findings, we expected a higher typed WPM to predict faster typed translation, as well as a higher dictated WPM to predict faster dictated translation—all of which has not been backed up by any statistically significant correlation. Also, no significant correlation was found with longer overall translation time and overall translation quality—in spite of the fact that professional translators on average spent 10 minutes more than students, achieving scores almost 25% higher than of students. An unexpected finding was that the average WPM of translators appeared higher than of students, although the difference was not statistically significant.

However, the limitations of this research—mainly revolving around the limited sample size of both respondent groups, and technical issues with speech recognition—imply the importance of looking past the numbers. In the post-testing survey, respondents identified a range of technical issues with the MemoQ's speech recognition feature: inaccuracy and unresponsiveness of word recognition, lack of voice commands, faulty pairing with iPhones, longer post-editing times, problems related to the Croatian language, and bugs overall. While the ability of modern-day ASR engines to pick up English language is still not perfect, it is even more unlikely to expect its perfection in recognizing Croatian in the near future—which was used as the input language in

this research. On the other hand, they displayed an overwhelmingly favorable attitude towards using the feature, naming some advantages to using it, such as faster input, its future potential, and its perceived appealing workflow. They mainly acknowledged speech input as a future disruptor of the translation industry, while looking forward to capitalize on the feature instead of condemning it. We have to consider the fact that the respondents have achieved a similar level of efficiency using ASR-CAT tool integration for the first time as using typing input ever since translating. Based on that, this this research provides a solid foundation for future work on the topic. Since using speech recognition still requires prior familiarizing, it may be possible to achieve a greater level in translation efficiency, in the case of arranging technical preparation of translators before using it.

## 8. References


Anumanchipalli, Gopala K., Josh Chartier, Edward F. Chang (2019). Speech synthesis from neural decoding of spoken sentences. *Nature.* 568: 493–498.

Balkul, Halil İbrahim (2016). Translation technologies: a dilemma between translation industry and academia. *International Journal of Language Academy.* 4: 100–108.

Berndsen, Juliette (2019). Google Translate: the end of language barriers? Available at: https://www.diggitmagazine.com/articles/google-translate-end-language-barriers (visited on 25th Aug 2019).

Candel-Mora, Miguel Angel (2016). Translator training and the integration of technology in the translator's workflow. María Luisa Carrió-Pastor, ed. *Technology Implementation in Second Language Teaching and Translation Studies: New Tools, New Approaches*. Singapore: Springer.

Cheng, Yong (2019). Robust Neural Machine Translation. Available at: https://ai.googleblog.com/2019/07/robust-neural-machine-translation.html (visited on 25th Aug 2019).

Condon, Stephanie (2018). How police are using voice recognition to make their jobs safer. Available at: https://www.zdnet.com/article/how-police-are-using-voice-recognition-to-make-their-jobs-safer (visited on 1st Sep 2019).

Crowther, David, Güler Aras (2010). Corporate Social Responsibility: Part II – Performance Evaluation, Globalisation and NFP's. Available at: https://www.academia.edu/2963633/Corporate_Social_Responsibility_CSR_Part_2 (visited on 20th Aug 2019). PDF

De Jesus, Ayn (2019). AI for Speech Recognition – Current Companies, Technology, and Trends. Available at: https://emerj.com/ai-sector-overviews/ai-for-speech-recognition (visited on 1st Sep 2019).

Diño, Gino (2018). Neural Machine Translation Research Output Ballooned in the First Half of 2018. Available at: https://slator.com/academia/neural-machine-translation-research-output-ballooned-in-the-first-half-of-2018/ (visited on 24th Aug 2019).

Fox, Jayne (2013). If I use a CAT tool will I get paid less for my translations? Available at: http://foxdocs.biz/BetweenTranslations/if-i-use-a-cat-tool-will-i-get-paid-less-for-my-translations/ (visited on 23rd Aug 2019).

Gayar, Neamat El, Suen Ching Yee (2018). *Computational Linguistics, Speech And Image Processing For Arabic Language.* Singapore: World Scientific.

Kikel, Chris (2019). Difference Between Voice Recognition and Speech Recognition. Available at: https://www.totalvoicetech.com/difference-between-voice-recognition-and-speech-recognition/ (visited on 25th Aug 2019).

Leonard, Kimberlee (2019). What Is a Good Typing Speed Per Minute? Available at: https://smallbusiness.chron.com/good-typing-speed-per-minute-71789.html (visited on 26th Aug 2019).

Marr, Bernard (2018). Will Machine Learning AI Make Human Translators An Endangered Species? Available at: https://www.forbes.com/sites/bernardmarr/2018/08/24/will-machine-learning-ai-make-human-translators-an-endangered-species/#5a9f8de03902 (visited on 25th Aug 2019).

Mazareanu, Elena Raluca (2019). Market size of the global language services industry from 2009 to 2021 (in billion U.S. dollars). Available at: https://www.statista.com/statistics/257656/size-of-the-global-language-services-market (visited on 19th Aug 2019).

Mellor, Joe (2015). Translation – A Recession-Proof Industry. Available at: https://www.thelondoneconomic.com/business-economics/business/translation-a-recession-proof-industry/20/08 (visited on 21st Aug 2019).

Peleman, Els (2017). Dragon NaturallySpeaking and CAT tools. Available at: https://epvertalingen.eu/de/dragon-naturallyspeaking-and-cat-tools (visited on 3rd Sep 2019).

Ram, Ashwin, et al. (2018). Conversational AI: The Science Behind the Alexa Prize. Available at: https://arxiv.org/pdf/1801.03604 (visited on 25th Aug 2019). PDF

Rapp, Brenda, Simon Fischer-Baum, Michele Miozzo (2015). Modality and morphology: what we write may not be what we say. *Psychol Sci.* 26(6): 892–902

Serbovets, Andrey (2016). Where can freelance translators get CAT tools at a more affordable price? Available at: https://www.quora.com/Where-can-freelance-translators-get-CAT-tools-at-a-more-affordable-price (visited on 23rd Aug 2019).

Sin-wai, Chan (2017). *The Future of Translation Technology*: *Towards a World without Babel.* London and New York: Routledge.

Sin-wai, Chan (2015). *The Routledge Encyclopedia of Translation Technology*. London and New York: Routledge.

Sterling, Greg (2019). Analyst: 8 billion voice assistants by 2023. Available at: https://searchengineland.com/analyst-8-billion-voice-assistants-by-2023-312035 (visited on 25 Aug 2019).

Teixeira, Carlos S. C., et al. (2019). Creating a multimodal translation tool and testing machine translation integration using touch and voice. *Informatics.* 6(1): 13.

Van der Velde, Naomi (2019). Innovative Uses of Speech Recognition Today. Available at: https://www.globalme.net/blog/new-technology-in-speech-recognition (visited on 1st Sep 2019).

Venkatesan, Hari (2018). Teaching translation in the age of Neural Machine Translation. *Selected Papers from the APLX2017*. 1:39–54

Yu, Deng, Li Deng (2014). *Automatic Speech Recognition: A Deep Learning Approach*. London: Springer.

DeepAI. What is a Neural Network? Available at: https://deepai.org/machine-learning-glossary-and-terms/neural-network (visited on 23rd Aug 2019).

DGT (2019). 2018 Annual Activity Report. Available at: https://ec.europa.eu/info/publications/annual-activity-report-2018-translation_en (visited on 20th Aug 2019). PDF

GALA. Language Industry Stakeholders. Available at: https://www.gala-global.org/language-industry-stakeholders (visited on 20th Aug 2019).

Grand View Research (2018). Voice and Speech Recognition Market Size, Share & Trends Analysis Report, By Function, By Technology (AI, Non-AI), By Vertical (Healthcare, BFSI, Automotive), And Segment Forecasts, 2018 – 2025. Available at: https://www.grandviewresearch.com/industry-analysis/voice-recognition-market (visited on 1st Sep 2019).

MateCat. Advanced options. Available at: https://www.matecat.com/phrase-based-vs-neural-mt-webinar-questions/advanced-options (visited on 3rd Sep 2019).

MemoQ. Hey memoQ — Dictation Support for memoQ Users. Available at: https://www.memoq.com/products/hey-memoq (visited on 3rd Sep 2019).

Nordictrans (2019). The translation industry victory against recession. Available at: https://www.nordictrans.com/blog/the-translation-industrys-victory-against-recession/ (visited on 21st Aug 2019).

Nuance. Get more done on your PC by voice. Available at: https://www.nuance.com/dragon/dragon-for-pc/home-edition.html (visited on 1st Sep 2019).

Omniscien. What is Rules Based Machine Translation (RBMT)? Available at: https://omniscien.com/rules-based-machine-translation/ (visited on 23rd Aug 2019).

Pangeanic. What is The Size of the Translation Industry? Available at: https://www.pangeanic.com/knowledge_center/size-of-the-translation-industry (visited on 21st Aug 2019).

PoliLingua (2018). The full list of CAT tools on the market: from translators to translators. Available at: https://www.polilingua.com/blog/post/cat_tools.htm (visited on 23rd Aug 2019).

Prima Lingua. Technology. Available at: https://primalingua.com/technology.html (visited on 23rd Aug 2019).

Social Science Statistics. Spearman's Rho Calculator. Available at: https://www.socscistatistics.com/tests/spearman/default.aspx (visited on 11th Sep 2019).

Social Science Statistics. The Wilcoxon Signed-Ranks Test Calculator. Available at: https://www.socscistatistics.com/tests/signedranks/default.aspx (visited on 11th Sep 2019).

Statistics Solutions. Mann-Whitney U Test Calculator. Available at: https://www.socscistatistics.com/tests/mannwhitney/ (visited on 11th Sep 2019).

Typing Speed Test. Available at: https://typing-speed-test.aoeu.eu (visited on 10th Sep 2019).