

Machine Translation of Polysemous Croatian Words in Various Text Genres

Tudor, Atena

Undergraduate thesis / Završni rad

2017

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka, Faculty of Humanities and Social Sciences / Sveučilište u Rijeci, Filozofski fakultet u Rijeci**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:186:662409>

Rights / Prava: [In copyright / Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-04-26**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Humanities and Social Sciences - FHSSRI Repository](#)



University of Rijeka
Faculty of Humanities and Social Sciences

Department of English Language and Literature

Atena Tudor

MACHINE TRANSLATIONS OF POLYSEMOUS CROATIAN WORDS IN VARIOUS TEXT GENRES

Supervisor: Prof. Dr. Sc. Marija Brala – Vukanović

Rijeka, September 2017

ABSTRACT

Machine translation is a powerful tool that is becoming ever more useful with the constant development of linguistics theory together with the practical applications of computational linguistics. Regardless of the highly advanced software, there are limitations that will always be present when a computer software renders a translation of a natural language. Mainly, these limitations are primarily related to *pragmatics*. A machine translator still cannot process problems of ambiguity, syntactic as well as semantic ones, and even when its corpus contains all the multiple translations of a certain word or phrase, it cannot decide on the proper equivalent in the given context. (Pritchard, 2002: 2) The phenomenon of *polysemy*, or multiple meanings, represents one of the problems that machine translators still fail to overcome, albeit, of course, it is not the only one. This work looks at examples of such machine translations of text examples from various text genres, namely texts from journalistic, medical, legal genres, and dictionary definitions.

TABLE OF CONTENTS

ABSTRACT	I
TABLE OF CONTENTS	II
1. INTRODUCTION.....	1
2. THE MOST IMPORTANT NOTIONS OF MACHINE TRANSLATION	2
2.1. Basic concepts of Machine Translation	2
2.2. The difficulties of Machine Translation	3
2.3. Polysemy as a linguistic phenomenon	4
2.4. Polysemy and MT	7
3. MACHINE TRANSLATIONS	8
3.1. Translations of the word <i>korijen</i>	19
3.2. Translations of the word <i>ključ</i>	20
3.3. Translations of the word <i>stolica</i>	21
3.4. Translations of the word <i>list</i>	21
3.5. Translations of the word <i>stanica</i>	22
3.6. Translations of the word <i>glava</i>	23
3.7. Translations of the word <i>lice</i>	24
4. FINDINGS	25
5. CONCLUSION	27
6. REFERENCES.....	30
7. TABLE OF FIGURES	31

1. INTRODUCTION

Each person that ever translated a text is familiar with the difficulties that can occur during that process. Some of the most common linguistic aspects of this difficulty are ambiguous words, idiomatic expressions, slang, and in general words that do not have literal meaning. Machine translation as a practice is considered rather helpful because it alleviates the translator's work. As a software, a machine translator operates by automatically translating word for word (or phrase for phrase) without human assistance from one language to another. In other words, the translator inputs a certain amount of text into the software, and the software renders the translation in seconds.

However, while a human translator can, for example, take a pragmatics – based approach when translating a text, a machine translator tool does not possess the capacity to “understand” the implied meaning of a word or a phrase within the text. Instead, a machine translator tool merely substitutes each word or phrase with the equivalent in the target language, which is statistically most commonly used when translating that specific word or phrase.

One of the linguistic phenomena that cannot be overcome by a machine translator is polysemy. Polysemy can be defined as a phenomenon where a linguistic unit has properties of multiple distinct, yet related meanings.

The purpose of this paper is to conduct an analysis of machine translations of specific texts from various text genres, with the focus on the following words in the Croatian Language that have a variety of meanings: *korijen*, *ključ*, *stolica*, *list*, *stanica*, *glava*, and *lice*.

2. THE MOST IMPORTANT NOTIONS OF MACHINE TRANSLATION

Translation (or *interlingual translation*) is most commonly defined as the process of producing a translation, i.e. interpretation of verbal signs by means of some other language, and has been the traditional focus of translation studies (Munday 2012: 9). With the advancement of technology, the concept of translating changed, as well as the role of the translator. When machine translation was introduced, most parts of the process of translating became automated. Regardless of its limitations, **machine translation** (hereafter MT) significantly improved the speed and the quality of translating.

From the very beginning, researchers dealt with scientific documents, which were less likely to contain a “culture – bound” vocabulary and contexts, which can be found in literature or legal texts. (Hutchins, 2010: 1) Although the field of Computational Linguistics, which studies natural language processing is rather new, since the 50s significant efforts have been made to improve the quality of translations, both in the expert field of studies, as well as in everyday language use.

Today, MT is most widely used on the Internet. Online translating services, such as Google Translate, are perhaps not known for their high quality translations, but they have a significant importance in giving immediate translations into users’ own languages for the purpose of simply conveying the information. (Hutchins, 2010: 17)

2.1. Basic concepts of Machine Translation

The development of artificial intelligence and advanced algorithms that could “mimic” the thought process of a translator while rendering a translation were crucial in

creating a model for MT. Another important studies that was of great importance for MT is Computational Linguistics, an interdisciplinary field that deals with processing of a natural language. MT can be considered as one of the most important components of Computational Linguistics. Computational Linguistics applies programming and math, especially statistics, in rendering language structures. This combination enables computer software to recognize natural language patterns, which has practical use in MT.

One of the most widely used methods of machine translation is *transfer – based machine translation*. Transfer – based MT “creates a translation from an intermediate representation that simulates the meaning of the original sentence”. (Okpor 2014: 162), Transfer MT breaks translation into three stages: the text firstly undergoes a morphological and syntactical **analysis** to determine its grammatical structure, then the resulting structure is **transferred** to a structure suitable for generating the translation in the target language, and lastly the translation is **generated**.¹

2.2. The difficulties of Machine Translation

A MT tool can potentially face multiple problems when rendering a text; there are problems with ambiguity, problems with multiple meanings, and problems that arise from structural and lexical differences between languages and units with multiple words, like idioms and collocations (Arnold 1994: 105).

Lexical and structural differences between languages occur because of the obvious differences in their lexis, the concepts that affect word choice for various expressions, and different structures that govern the composition of phrases and clauses (or grammar). This

¹ https://en.wikipedia.org/wiki/Transfer-based_machine_translation , accessed on June 24 2017

applies to situations when a word in the source language should be substituted with an entire phrase in the target language.

Idioms are specific expressions in a language that rarely have literal meaning, they are very often metaphorical in nature. Thus, idioms cannot be translated using translation procedures described above. An MT can only render such a word or phrase correctly if the translation was already incorporated in its corpus.

Ambiguity occurs when a word can have multiple meanings in a certain context. MT should theoretically be able to solve the problem by selecting the semantically correct equivalent in the target language, however MT does not have the capacity to “understand” the context of a given text.

When a word has multiple distinct, yet related meanings, it is an example of polysemy. Polysemy is one of the concepts that machine translators have difficulties with.

2.3. Polysemy as a linguistic phenomenon

Polysemy is the capacity of words to, apart from having the same form, share a related meaning. As one of the fundamental concepts in linguistics, polysemy does not represent a problem in communication for speakers; many words are polysemous as a result of economically using the existing vocabulary to give new meanings to the already available words. An example of polysemy would be the word *head*. The Oxford dictionary provides more than 10 distinct meanings of *head*, as the upper part of the body separated from the rest of the body by a neck, as a measure, as the front or upper part of something, a person in charge of something, and more.

It can be said that polysemy is one of the concepts that make communication, and life, much easier. Without polysemy and other linguistic concepts that enable word play and

abstract use of language, communication would be significantly restricted because speakers would have to use specific words for every individual concept. It is impossible to establish the number of words that would be needed to use each word for its own unique meaning.

Polysemy is not to be confused with ambiguity or homonymy. Polysemy occurs at the word level. When a speaker says that “the head of the department will be visiting our office the next day”, everyone will understand that the implied meaning of *head* is the person in charge of the department, not that a literal head will be flying around mid – air. We are capable to distinguish the logical equivalent from all the possible meanings that best suits the context of the conversation.

On the other hand, ambiguity, or simply vagueness of meaning, arises at the sentence level, and can be solved by considering the context. If someone says that they visited the apothecary earlier this morning, it cannot be clear whether they meant the pharmacy or the pharmacist. However, when a person says “I asked the apothecary to recommend me a good skin cream” it is clear that they are talking about a person, not the institution. Homonymy differs from polysemy and ambiguity in the way that it includes words with unrelated meanings even though they are the same in structure, and homonyms are usually words that belong to different word categories. An example of homonyms are the words *book* as a noun, meaning “a bind of papers containing a text to be read”, and *book* as a verb, meaning “to make a reservation”. Unlike polysemy where words have related meanings to a certain extent, homonyms have completely unrelated meanings.

In order to analyse polysemy as a linguistic concept, it is important to explain some important notions in linguistics relevant to the research of meaning, *the concept* and *the frame*. Cognitive linguistics studies language as a mental phenomenon. It is a modern approach to the study of meaning, and conceptual structure, and one of the most dominant

perspectives on the essence of language, and the mind. According to Brala – Vukanović, cognitive linguistics, and cognitive semantics in particular, view language, or meaning, in terms of the structure of concepts. Cognitive semantics claims that language, as a part of human cognition, is strongly connected to mental processing and conceptualization. In other words, language can be observed only in relation to the mind as a whole system. (Brala – Vukanović 2013: 84) One of the key notions in cognitive linguistics is *the concept*. A concept is an abstract notion, an idea, a mental image or a symbol.

In linguistics, the semantic concept is a basic idea represented symbolically in the mind, in contrast to a language unit to which it refers to. It is well said in the work *Concept understanding in cognitive linguistics* that “the word is not thought as some kind of objective reality existing “in and for itself”, it is shaped by our cognitive function including human perception, ability for categorization and conceptualization, interwoven with interpersonal experiences of human interaction”. (Manerko 2014: 474) Thus, the basis of the semantic concept behind the word “dog” is “animal, fur, paws, tail, pet, bark, fun”. Of course, every person might add something depending on their experience; if they have a dog they might add “happiness”, but the concept is that core of what all speakers have in common in their sensory image of the word “dog”. To simplify, a concept can be imagined as a circle, with the core notion as the centre of the structure, and all the notions added by culture, tradition and personal experiences being at the periphery. (Nemickienė 2011: 5)

In frame semantics, the term frame is defined as “a coherent structure of concepts that are related in such a way that without knowledge of all of them, one does not have complete knowledge of any of them either, and are in that sense types of gestalt.” (Brala – Vukanović 2013: 129) Observing polysemy in cognitive linguistics includes “viewing meaning as categorization, recognizing the importance of context for meaning and that

linguistic and encyclopaedic knowledge are hard to keep separate, and incorporating prototype theory into linguistics”. (Nemickiené 2011: 1)

2.4. Polysemy and MT

Polysemous words can represent a challenge for human translators. A translator needs to make further research within the source language and the target language in order to find the appropriate equivalent for their translation. Since a MT can only render a translation with the translation rules written in its code, and the language corpus that it already contains, it is interesting to analyse how a machine translator would manage to translate certain polysemous words from Croatian language to English.

Since machine translations follow the properties and respect the rules of natural language in their function of translating, the analysis of polysemous words and their translations can be made by analysing the concept of the words that were chosen as the output translation by the machine translator.

Observing natural language is important here because machine translation deals with the processing of the very same. Therefore, the quality of the translation of polysemous words can be studied by comparing the relation between the concepts behind each meaning of the polysemous word, and analysing whether and to what extent these words in Croatian and English are comparable.

3. MACHINE TRANSLATIONS

Like many other languages, Croatian Language is rich with polysemous words. Polysemy implies a metaphorical richness in meaning which shows how words go through semantic changes and attain new meanings while usually not losing the original essence of the meaning concept. Thus, a polysemous word usually needs to be put in a context in order to obtain the appropriate single meaning of all its multiple meanings. However, its meaning always contains a common semantic concept. For example, the meaning of “root” is always denoted by holding something grounded, or being the source or stem of something, being underneath the surface, whether it is a root of a plant, root of a tooth or the root of a problem.

For the purposes of this work, the Croatian words *korijen*, *ključ*, *stolica*, *list*, *stanica*, *glava* and *lice* in sentences from various texts were rendered through a simple machine translator, Google Translator in order to analyse polysemous words translated using an MT, how they are translated into the English Language, and with what level of accuracy. The original sentences were inserted into the MT tool, and the translations were automatically rendered. The sentences and their translations are shown in the following tables.

Table 1 Translations of the word **korijen**

Original sentence or phrase	Meaning and context, all definitions have been taken from Oxford Online Dictionary	Machine translation
korijen		
<p>Jedan od tri osnovna organa kopnenih biljaka je korijen.</p> <p>Source: http://e-skola.biol.pmf.unizg.hr/odgovori/odgovor425.htm</p>	<p>The part of a plant which attaches it to the ground or to a support, typically underground, conveying water and nourishment to the rest of the plant via numerous branches and fibres.</p> <p>Source: https://en.oxforddictionaries.com/definition/root</p>	<p>One of the three basic organs of land-based plants is the root.</p>
<p>Tvorbenica je osnova dio tvorenice koji čuva izraznu i sadržajnu vezu s osnovnom riječju. Najmanji zajednički sastavni element svih tvorbeno srodnih riječi, tj. cijele tvorbenice porodice zove se korijen.</p> <p>Source: http://www.enciklopedija.hr/natuknica.aspx?id=62863</p>	<p>A morpheme, not necessarily surviving as a word in itself, from which words have been made by the addition of prefixes or suffixes or by other modification.</p> <p>Source: https://en.oxforddictionaries.com/definition/root</p>	<p>The foundational basis is a part of the creature that maintains a meaningful and substantive connection to the basic word. The smallest common element of all the creature-related words, ie, the entire family of creation, is called the root.</p>
<p>Drugi ili kvadratni korijen iz pozitivnog broja a je pozitivan broj čiji je kvadrat jednak a.</p> <p>Source: https://bs.wikipedia.org/wiki/Kvadratni_korijen</p>	<p>A number or quantity that when multiplied by itself, typically a specified number of times, gives a specified number or quantity.</p> <p>Source: https://en.oxforddictionaries.com/definition/root</p>	<p>The second or square root of the positive number a is a positive number whose square is equal to.</p>
<p>Prilikom ekspanzije potrebno je uključiti što više zubi u sidrište i ako su korijeni mliječnih molara dosta resorbirani, bolje je pričekati nicanje premolara i onda ekspanzirati.</p>	<p>In stomatology, the part of a tooth within the socket.</p> <p>Source: https://www.merriam-webster.com/dictionary/root</p>	<p>When expanding it, it is necessary to include as many teeth as the anchorage, and if the roots of dairy molars are highly resorbable, it is better to wait for the premolar to start and then expanding.</p>

Source: https://bib.irb.hr/datoteka/591935.ORTODONTSKI_PRIRUNIK.pdf		
Korijen nosa / korijen nokta	The embedded part of a bodily organ or structure such as a hair, tooth, or nail. Source: https://en.oxforddictionaries.com/definition/root	Root of the nose / root of the nail
Papa o gospodarskoj i financijskoj krizi: Korijen problema je u etici, a rješenje u humanizmu. Source: http://www.index.hr/vijesti/clanak/papa-o-gospodarskoj-i-financijskoj-krizi-korijen-problema-je-u-etici-a-rjesenje-u-humanizmu-/593437.aspx	The basic cause, source, or origin of something. Source: https://en.oxforddictionaries.com/definition/root	The Pope on Economic and Financial Crisis: The root of the problem is in ethics, and the solution in humanism

Table 2 Translations of the word *ključ*

Original sentence or phrase	Meaning and context, all definitions have been taken from Oxford Online Dictionary	Machine translation
ključ		
<p>Sama pomisao na to da izgube ključeve automobila noćna je mora mnogih vozača.</p> <p>Source: http://www.oryx-asistencija.hr/pomoc-na-cesti/sto-kad-kljuc-nestane-6752</p>	<p>A small piece of shaped metal with incisions cut to fit the wards of a particular lock, which is inserted into a lock and turned to open or close it.</p> <p>Source: https://en.oxforddictionaries.com/definition/key</p>	<p>The very thought of losing the car keys is a nightmare for many drivers.</p>
<p>Francuski ključ Engleski ključ Matični ključ Imbus ključ</p>	<p>An adjustable tool like a spanner, used for gripping and turning nuts or bolts.</p> <p>Source: https://en.oxforddictionaries.com/definition/wrench</p>	<p>The French key Shifting wrench Socket spanner Allen wrench</p>
<p>Glazbeni ključ Violinski ključ Basov ključ</p>	<p>Any of several symbols placed at the left hand end of a stave, indicating the pitch of the notes written on it.</p> <p>Source: https://en.oxforddictionaries.com/definition/clef</p>	<p>Music key Treble clef Bass key</p>
<p>Spoznaja problema i poduzimanje koraka u njegovu rješavanju ključ je duhovnog i fizičkog ozdravljenja.</p> <p>Source: https://www.antenazadar.hr/clanak/2014/11/larisa-i-kristijan-mikulaj-spoznaja-problema-i-poduzimanje-koraka-u-njegovu-rjesavanju-kljuc-je-duhovnog-i-fizickog-ozdravljenja/</p>	<p>A thing that provides a means of achieving or understanding something.</p> <p>Source: https://en.oxforddictionaries.com/definition/key</p>	<p>Knowing the problem and taking the steps in solving it is the key to spiritual and physical healing.</p>

Ključna odluka	Of crucial importance. Source: https://en.oxforddictionaries.com/definition/key	Key decision
-----------------------	--	---------------------

Table 3 Translations of the word *stolica*

Original sentence or phrase	Meaning and context, all definitions have been taken from Oxford Online Dictionary	Machine translation
stolica		
Nudimo bogati izbor stolova i stolica .	A separate seat for one person, typically with a back and four legs. Source: https://en.oxforddictionaries.com/definition/chair	We offer a rich selection of tables and chairs .
Analiza uzorka stolice može otkriti mnogo o zdravlju čovjeka.	A piece of faeces. Source: https://en.oxforddictionaries.com/definition/stool	The analysis of the stool sample can reveal much about human health.
U međunarodnim odnosima I u međunarodnom pravu Sveta Stolica je zaseban subject međunarodnog prava, koji se ne smije miješati s Državom Vatikanskog Grada. Source: https://hr.wikipedia.org/wiki/Sveta_Stolica	The papacy or the papal court; those associated with the Pope in the government of the Roman Catholic Church at the Vatican. Source: https://en.oxforddictionaries.com/definition/holy_see	In international relations and in international law, the Holy See is a separate subject of international law, which must not interfere with the State of the Vatican City.

Table 4 Translations of the word *list*

Original sentence or phrase	Meaning and context, all definitions have been taken from Oxford Online Dictionary	Translation
list		
<p>Osim klorofilne fotosinteze, koja omogućuje mijenu tvari, list obavlja i razmjenu s atmosferom, pa biljka svojim staničjem diše i transpirira (ispušta vodenu paru).</p> <p>Source: http://bonsai-zen.hr/osnove_botanike/</p>	<p>A flattened structure of a higher plant, typically green and blade-like, that is attached to a stem directly or via a stalk. Leaves are the main organs of photosynthesis and transpiration.</p> <p>Source: https://en.oxforddictionaries.com/definition/leaf</p>	<p>In addition to chlorophyllic photosynthesis, which allows the exchange of matter, the leaf also performs the exchange with the atmosphere, and plants the plant with its breathing and transpires (releases the water vapor).</p>
<p>Napišite na list papira imena svih mjesta koja biste htjeli posjetiti prije nego što umrete.</p> <p>Source: https://www.facebook.com/citajknjigu/posts/815786221930677</p>	<p>A rectangular piece of paper, especially one of a standard size produced commercially and used for writing and printing on.</p> <p>Source: https://en.oxforddictionaries.com/definition/sheet</p>	<p>Write on the paper sheet the names of all the places you would like to visit before you die.</p>
<p>Iako list spada u manje mišićne skupine, smatra se najjačim skeletnim mišićem našeg tijela, koji nam omogućava da hodamo, trčimo, skačemo i stojimo na nogama jako dug vremenski period.</p> <p>http://body.ba/bodybuilding/trening-i-vjezbe/vjezbe-za-ogromne-listove/63</p>	<p>The fleshy part at the back of a person's leg below the knee.</p> <p>Source: https://en.oxforddictionaries.com/definition/calf</p>	<p>Although the sheet belongs to a lesser muscle group, it is considered to be the strongest skeletal muscle of our body, which allows us to walk, run, skate and stand for a very long period of time.</p>
<p>List je jedna od vrsta riba specifičnih za Jadran, te je se može naći na muljevitom i pjeskovitom dnu diljem obale.</p> <p>Source: https://www.gastronomija.hr/list/</p>	<p>A marine flatfish of almost worldwide distribution, important as a food fish.</p> <p>Source: https://en.oxforddictionaries.com/definition/sole</p>	<p>List is one of the species of fish specific to the Adriatic, and it can be found on the sandy and sandy bottom of the coast.</p>

Table 5 Translations of the word *stanica*

Original sentence or phrase	Meaning and context, all definitions have been taken from Oxford Online Dictionary	Translation
stanica		
<p>U biologiji riječ stanica dolazi od latinske riječi <i>cellula</i> (hrv. ćelija). Ime je odabrao Robert Hooke promatrajući tanke prereze pluta (stijenke mrtvih stanica).</p> <p>Source: https://hr.wikipedia.org/wiki/Stanica</p>	<p>The smallest structural and functional unit of an organism, which is typically microscopic and consists of cytoplasm and a nucleus enclosed in a membrane.</p> <p>Source: https://en.oxforddictionaries.com/definition/cellb</p>	<p>In biology, the word cell comes from the Latin word <i>cell</i> (cellular cell). The name was chosen by Robert Hooke, observing the thin cuttings of the cork (the walls of the dead cells).</p>
<p>Pčelinje gnijezdo se sastoji od saća na kojem se nalaze pčele i njihovo leglo. Saće je zapravo skup šesterokutnih (heksagonih) stanica ili ćelija koje su pravilno poredane jedna do druge, tako da između njih nema praznog prostora.</p> <p>Source: http://blog.dnevnik.hr/api/kultura/2013/03/1631524480/kako-pcele-grade-sace.html</p>	<p>A small compartment in a larger structure such as a honeycomb.</p> <p>Source: https://en.oxforddictionaries.com/definition/cell</p>	<p>The bee's nest consists of a honeycomb with bees and their litter. Honeycomb is actually a set of hexagonal (hexagon) cells or cells that are properly aligned to each other so that there is no blank space between them.</p>
<p>Berlinska glavna željeznička stanica sagrađena je kao centralna željeznička stanica Berlina u samom centru grada, na mjestu nekadašnje Lehrter stanice.</p> <p>Source: https://bs.wikipedia.org/wiki/Berlinska_glavna_%C5%BEeljezni%C4%8Dka_stanica</p>	<p>A place on a railway line where trains regularly stop so that passengers can get on or off.</p> <p>Source: https://en.oxforddictionaries.com/definition/station</p>	<p>Berlin's main railway station was built as the central Berlin railway station in the very center of the city, at the site of the former Lehrter station.</p>
<p>Vozač je dužan da, u naselju, omogućiti</p>	<p>A bus or coach station.</p>	<p>The driver is obliged to allow the bus to be included in the</p>

<p>uključivanje u promet autobusa koji se propisno uključuje (primjerice dao je lijevi žmigavac) u promet sa stanice autobusa.</p> <p>Source: http://www.propisi.hr/print.php?id=8177</p>	<p>Source: https://en.oxforddictionaries.com/definition/station</p>	<p>bus, which is properly included (for example, the left flame arrester) is in traffic from the bus station.</p>
---	---	--

Table 6 Translations of the word *glava*

Original sentence or phrase	Meaning and context, all definitions have been taken from Oxford Online Dictionary	Translation
glava		
Heraklo je stigao do močvare lernejskog jezera gdje je živjela Hidra. Zatim se on suočio s njom. Svaki put kad bi odsjekao neku od njezinih glava , opet bi narasla. Source: https://hr.wikipedia.org/wiki/Heraklo	The upper part of the human body, or the front or upper part of the body of an animal, typically separated from the rest of the body by a neck, and containing the brain, mouth, and sense organs. Source: https://en.oxforddictionaries.com/definition/head	Heracle came to the swamp of Lerne Lake where the Hydra lived. Then he faced her. Every time he cut off one of her heads , she would grow again.
Glava obitelji	A person in charge of something; a director or leader. Source: https://en.oxforddictionaries.com/definition/head	Head of the family
Knjiga je podijeljena u sedam glava .	A main division of a book, typically with a number or title. Source: https://en.oxforddictionaries.com/definition/chapter	The book is divided into seven heads .
Glava kupusa	A thing resembling a head either in form or in relation to a whole. Source: https://en.oxforddictionaries.com/definition/head	Head of cabbage

Table 7 Translations of the word *lice*

Original sentence or phrase	Meaning and context, all definitions have been taken from Oxford Online Dictionary	Translation
lice		
<p>Ruža nije lijepa ako, uz cvjetove njene nema umilnog lica drage, voljene žene.</p> <p>Source: http://blog.dnevnik.hr/kira1102/2007/03/1622267188/zena.2.html</p>	<p>The front part of a person's head from the forehead to the chin, or the corresponding part in an animal.</p> <p>Source: https://en.oxforddictionaries.com/definition/face</p>	<p>Rose is not beautiful if, along with its flowers There is no humble face of a dear, beloved woman.</p>
<p>Dramski tekst dijeli se na popis lica ili osoba, didaskalije, dijaloge i monologe.</p> <p>Source: https://hr.wikipedia.org/wiki/Drama</p>	<p>A person in a novel, play, or film. A part played by an actor.</p> <p>Source: https://en.oxforddictionaries.com/definition/character</p>	<p>Dramatic text is divided into a list of faces or persons, dialects, dialogues, and monologues.</p>
<p>Kod prvog lica množine, kao i kod drugog lica jednine i množine, imperativ se tvori se od prezentske osnove i nastavaka.</p> <p>Source: https://hr.wikipedia.org/wiki/Imperativ</p>	<p>A grammatical classification of words that consists typically of singular and plural, and, in Greek and certain other languages, dual.</p> <p>Source: https://en.oxforddictionaries.com/definition/number</p>	<p>In the first plural face as well as in the other face of the singular and plural, the imperative is formed from the presentational basis and sequences.</p>
<p>Sav hrvatski kovani novac kovan je po pravilima medaljerskog kovanja, što znači da se vrtnjom kovanice oko njezine okomite osi, lice i naličje vide u uspravnom položaju.</p> <p>Source: http://www.kunalipa.com/katalog/kovanice/</p>	<p>The side of a coin or medal bearing the head or principal design.</p> <p>Source: https://en.oxforddictionaries.com/definition/obverse</p>	<p>All Croatian coins are coined by the rules of medallary (German) coins, which means that by turning the coins around its vertical axis, the face and the reverse are seen upright.</p>

3.1. Translations of the word *korijen*

Korijen, meaning “root”, is one of Croatian words that is a typical example of a polysemous word. Its basic meaning is “The part of a plant which attaches it to the ground or to a support, typically underground, conveying water and nourishment to the rest of the plant via numerous branches and fibres”. Except in biology, it can also represent an anatomical term, as “The embedded part of a bodily organ or structure such as a hair, tooth, or nail”. In general, speakers use the word *korijen* when talking about a basic source, a beginning or the origin of something, like in the phrase “the root of all evil”, or “the root of the problem”. The word *korijen* can be found in other fields of study, such as mathematics and linguistics. In mathematics, square root (*drugi korijen*) of a number *a* is a number such that when multiplied by itself, gives the number *a*. In linguistics, *korijen riječi*, meaning “the stem of the word” or the root of the word is “a morpheme, not necessarily surviving as a word in itself, from which words have been made by the addition of prefixes or suffixes or by other modification”.

The machine translations of the word *korijen* all had very satisfactory results. Thus, the sentence “Jedan od tri osnovna organa kopnenih biljaka je korijen.” was translated as “One of the three basic organs of land – based plants is the **root**”. The anatomical terms “korijen nosa” and “korijen nokta” also resulted with the correct English equivalents “**root** of the nose” and “root of the nail”. Taken from a stomatology textbook, the sentence “(...) potrebno je uključiti što više zubi u sidrište i ako su korijeni mliječnih molara dosta resorbirani, (...)” resulted with “(...) it is necessary to include as many teeth as the anchorage, and if the **roots** of dairy molars are highly resorbable, (...)”. In a more general context, a quote from the newspapers “Korijen problema je u etici, a rješenje u humanizmu” was translated as “The **root** of the problem is in ethics, and the solution in humanism”.

3.2. Translations of the word *ključ*

In a similar manner to the word *korijen*, the word *ključ* meaning “key”, has a wide range of meanings in Croatian Language. Apart from the basic meaning as “the object which is used to unlock doors”, *ključ* is used in mechanics for various types of wrenches: *francuski ključ* (monkey wrench), *engleski ključ* (shifting wrench), *matični ključ* (socket spanner), and *imbus ključ* (Allen wrench). In music, *ključ*, or clef, is “a group of pitches, or scale upon which a music composition is created”. More generally, *ključ* is used when speaking of problems and decisions, *ključ problema* being “the key of a problem”, meaning something that can resolve the problem, and *ključna odluka*, “a key decision”, that is a very important decision. Another interesting usage of the word *ključ* is the phrase “*ključ u ruke*” which is a kind of contract known as “turnkey” contract, where an investment is agreed upon, until the final degree when the investor can “turn the key” and start the contracted investment.

In the following examples, “Sama pomisao na to da izgube ključeve automobile noćna je mora mnogih vozača” was translated “The very thought of losing the car **keys** is a nightmare for many drivers”. The following terms *francuski ključ*, *engleski ključ*, *matični ključ*, *imbus ključ* were translated correctly as mentioned above, apart from ‘francuski ključ’ which was translated as ‘the French **key**’. *Glazbeni ključ* and *basov ključ* had a less satisfactory translation as ‘music **key**’ and ‘bass **key**’, with the word key instead of clef, however *violinski ključ* was correctly translated as the treble **clef**. The sentence „Spoznajati problema i poduzimanje koraka u njegovu rješavanju ključ je duhovnog i fizičkog ozdravljanje” was translated as „Knowing the problem and taking the steps in solving it is the **key** to spiritual and physical healing“. Finally, the phrase „ključna odluka“ was translated as „**key** decision“.

3.3. Translations of the word *stolica*

Stolica, meaning a “chair”, has a basic meaning as “a separate seat for one person, typically with a back and four legs”. Apart from its principal meaning, it is used in the phrase *Sveta Stolica*, or Holy See, as the name of the supreme authority of the Catholic Church, the Pope. In general, *stolica* is used when speaking of a part of university, which is translated into the English Language as “cathedra”, deriving from the Greek word for chair. *Stolica*, meaning stool, is the faeces of the human digestive system.

The machine translations of the following examples show that there were no problems in ascribing the correct equivalent of the word *stolica*, depending on the context. Thus, “Nudimo bogati izbor stolova i stolica” was translated as “We offer a rich selection of tables and **chairs**”, and “Analiza uzorka stolice može otkriti mnogo o zdravlju čovjeka” was also correctly translated as “The analysis of the **stool** sample can reveal much about human health”. The sentence regarding the *Sveta Stolica* as the name of the Pope, “U međunarodnim odnosima i međunarodnom pravu Sveta Stolica je zaseban subjekt međunarodnog prava, (...)” was translated as “In international relations and in international law, the **Holy See** is a separate subject of international law, (...)”.

3.4. Translations of the word *list*

List is one of the more interesting examples of polysemy in Croatian Language for its various meanings. As a term in biology, *list* meaning “leaf”, is “a flattened structure of a higher plant, typically green and blade-like, that is attached to a stem directly or via a stalk.”. In Croatian, *list* is also used to denote a sheet of paper. In anatomy, the calf muscle is called

list, because of its shape that reminds of a leaf. Lastly, *list* is a name for a fish species (sole fish) that is specific to the Adriatic Sea, which resembles to a leaf in its shape.

Some of the translations were rendered correctly. The sentence “Osim klorofilne fotosinteze, koja omogućuje mijenu tvari, list obavlja i razmjenu s atmosferom, (...)” was translated into “In addition to chlorophyllic photosynthesis, which allows the exchange of matter, the **leaf** also performs the exchange with the atmosphere, (...)”. The phrase „napisati na list papira“ resulted with an appropriate translation as „write on the paper **sheet**“. On the other hand, the *list* as a type of muscle, and as a kind of fish were not translated satisfactory, even though they were given within a very specific context to enable finding the right equivalent more easily.

The sentence „Iako list spada u manje mišićne skupine, smatra se najjačim skeletnim mišićem našeg tijela (...)“ was translated as „Although the **sheet** belongs to a lesser muscle group, it is considered to be the strongest skeletal muscle of our body (...)“, and the sentence „List je jedna od vrsta riba specifičnih za Jadran, te je se može naći na muljevitom i pjeskovitom dnu diljem obale“ was translated as „**List** is one of the species of fish specific to the Adriatic, and it can be found on the sandy and sandy bottom of the coast“.

3.5. Translations of the word *stanica*

The words *stanica* has multiple meanings in biology and transport. In biology, *stanica*, or the „cell“, means „the basic structural and functional unit of all living things“. It is used in the same form to denote a hexagonal cell that bees build with wax when creating honeycombs to store honey. In transport, *stanica* has rather related meanings; as a station – „a stop on the way from one point to another“; and as a bus station or train station, „a facility where trains or buses regularly stop to load or unload passengers and freight“.

Thus, the sentence „U biologiji riječ stanica dolazi od latinske riječi cellula. Ime je odabrao Robert Hooke promatrajući tanke prereze pluta (stijenke mrtvih stanica)“ was translated as „In biology, the word **cell** comes from the Latin word cellula. The name was chosen by Robert Hooke, observing the thin cuttings of the cork (the walls of the dead **cells**).“

In the sentence „ (...) Saća je zapravo skup šesterokutnih stanica koje su pravilno poredane jedna do druge, tako da između njih nema praznog prostora.“ *stanica* is properly replaced with the word cell: „ (...) Honeycomb is actually a set of hexagonal **cells** that are properly aligned to each other so that there is no blank space between them.“

The sentence „Berlinska glavna željeznička stanica sagrađena je kao centralna željeznička stanica Berlina u samom centru grada, na mjestu nekadašnje Lehrter stanice“ also had an adequate translation as „Berlin's main railway station was built as the central Berlin railway **station** in the very centre of the city, at the site of the former Lehrter **station**“, and the sentence „Vozač je dužan da omogući uključivanje u promet autobusa koji se uključuje u promet sa stajališta autobusa“ also was correctly translated as „The driver is obliged to allow the bus to be included in the traffic, which is properly included in the traffic from the bus **station**“, apart from some other minor grammatical and lexical mistakes in the translation.

3.6. Translations of the word *glava*

Glava, meaning “head”, is one of the words that has equally as many multiple meanings in Croatian and in English. In both languages, it is used as an anatomical term for a part of the body above the neck where the brain is located. In business, it denotes the individual that is in charge of the family, or a person that leads a company; it is used to describe a piece of

alimentary product shaped as a ball, such as a “head of cabbage”. In Croatian, *glava* is the term used for chapters in which books are usually divided into.

The words ‘glava’ meaning literally ‘head’, and ‘head of family’, had correct equivalents in the target language. Only the Croatian *glava* meaning ‘chapter’, as in the sentence “Knjiga je podijeljena u sedam glava”, was translated as “The book is divided into seven **heads**”, instead of using the term “chapters”.

3.7. Translations of the word *lice*

Lice has many various, although still in a way related meanings, and as such it is the most interesting example of all the polysemous words in this work. The most basic meaning of the word *lice* is the “face”, or the front side of the head from the top of the forehead to the beard. In drama, *lice* is a term for a character in a play. In linguistics, *lice* denotes the plural or singular form of nouns, pronouns and adjective and verb agreement, also known as the grammatical number. Finally, an interesting use of the word *lice* is in numismatics, where *lice* means a flat face of a coin, called face or obverse.

The translations of the term *lice* were the ones with most errors. Only the *lice* meaning the front of the head, was translated correctly as “face”. The sentence “Dramski tekst dijeli se na popis lica ili osoba, dijaloge i monologe” was translated as “Dramatic text is divided into a list of **faces** or persons, dialogues, and monologues”, using the word “face” instead of “character”. In the phrase ‘prvo lice množine’ the word ‘lice’ was translated as “first plural face” instead of “first person plural”. The translation of “lice kovanice” as the “**face** of the coin” was partially appropriate because it can both be translated as “the face of the coin” and the “obverse”, which is an official name for a flat side of the coin.

4. FINDINGS

Several things can be noted while observing the resulting translations of the words *korijen*, *ključ*, *stolica*, *list*, *stanica*, *glava*, and *lice*. The words *korijen*, *ključ*, and *stolica* resulted with translations of the most accuracy, while other words had mostly satisfactory results with few or more discrepancies from the expected translation, such as *stanica*.

In case of the word ‘korijen’ (root), its basic semantic concept is “to be underground (or invisible), to hold in place”. All the examples of the word in various contexts show that in both languages the multiple meanings of the said word share the same concept, and the result are translations with a 100% of accuracy for both words. The results of translations of the word ‘stolica’ show that despite the obvious differences in the English equivalents, who share nothing in common among each other, the statistical machine translated all words correctly probably because of how frequently these words appear in everyday language online.

When observing the word ‘glava’ (head), it can be agreed that the concept behind the word is “on top of something, very important, the centre”. Most of the instances of the word ‘glava’ in different contexts share the same basic concept; as a body part is on top of the neck, as a family leader ‘glava’ is on top of the family matter. As a piece of vegetable, the word ‘glava’ is used because of its round shape. The only wrong translation was of the word ‘glava’ as a book chapter. The Croatian term ‘glava’ is most likely derived from the more commonly used term ‘poglavlje’. Interestingly, the English equivalent for ‘glava’ (or ‘poglavlje’) is ‘chapter’, which derives from the Latin word for ‘head’.² However, since the word ‘chapter’ and ‘head’ in English don’t share the same semantic concept as the words

² <http://www.etymonline.com/index.php?term=chapter>, accessed on August 27 2017

‘glava’, meaning of a chapter, and ‘poglavlje’ share in Croatian, the result was a wrong translation.

In case of the words ‘ključ’ and ‘list’, the results show that in the more abstract examples where the basic concept of the word meaning is not semantically clearly related to one of its meanings as a polysemous word, the machine translator used the literal translation, which was a wrong translation as a result, in most cases. Thus, a type of wrench that is specifically named so in Croatian, ‘francuski ključ’, is translated as a key instead of a wrench. Regarding the word ‘list’, its meaning as the name of a muscle which in Croatian derives from the shape of the muscle that resembles a leaf. In English, there is no such occurrence. Thus, ‘list’ was translated as leaf instead of calf which would be the appropriate name for the muscle. Another interesting example is the word ‘list’ as a name of a fish. In Croatian, its name derives from its shape that resembles a leaf, while in English the fish is named sole, for its resemblance to a sandal, Latin *solea*.³ Because of this discrepancy, the machine translation tool rendered the translation in the way that it kept the original word ‘list’ without translating it at least to its literal equivalent ‘leaf’.

When observing the word ‘lice’, it is clear that in Croatian Language, ‘lice’ has multiple meanings that all share the principal concept of ‘a front, a person’. However, in English it is not the same, meaning that the English equivalents do not have any interrelations between their semantic concepts. Thus, the translations of the word ‘lice’ had the least satisfying results with only one correct translation out of four.

In conclusion, those word pairs (for example: head – glava) where both words are closer to the semantic concept in their own language, are more likely to result with an accurate translation using a simple statistical MT tool, such as Google Translate.

³ <http://www.etymonline.com/index.php?term=sole> accessed on August 27 2017

5. CONCLUSION

The goal of this work was to compare the given texts and their translations to English in regard to the level of accuracy by which the polysemous words in various contexts were translated into their English equivalents. By giving a brief overview of the field of machine translation and the notion of polysemy, we have found that the quality of translation could be analysed and explained through the analysis of the connections between the concepts of polysemous words in the Croatian and English Languages. Some translations showed a surprising accuracy, while others were, as expected, less than acceptable. The closer the basic concept of a given polysemous word conceptually is to its meaning (one of its meanings) in both languages, the more likely is the translation to be accurately found.

On the other hand, in cases where there are no links between the concepts of certain words in the source language and the target language, the only other component that can help to achieve a correct translation is to enable a text corpora extensive in such manner that there would be enough examples to statistically result with the correct translation every time. From everything that was explored in this work, it can be concluded that semantic notions, such as polysemy, are far too complex to be properly rendered by computer software without great efforts and many lines of coding.

Nirenburg (1993) states one of the problems with machine translations, which underlies the problem with polysemy in machine translation as well, and that is the fact that there are no complete one-to-one correspondences between lexical units in the source and target texts. Therefore, the machine translation cannot give an adequate translation, that is, the translation that we are satisfied with or one at the level of a human translator. However, as Nirenburg (1993) notes, this does not mean that machine translation is completely useless. For example, it is very helpful in processing large amounts of highly technical texts in a

short amount of time, when human translators simply would not be able to take on such large amounts of text. Another example of the usefulness of machine translation are the texts which were specifically made for machine translations e.g. texts containing nontextual symbols.

Linguistic research in the field of computational linguistics shows great potential in improving the accuracy of machine translations. There are different approaches that can be taken in this endeavour. Some studies take on the analysis of the text corpora of the source language and the target language, in order to improve the efficiency of parsers that represent the most important part of machine translating, which is analysing the string of words or phrases.

A research by Munteanu and Marcu from the University of Southern California showed that it is possible to improve machine translations by extracting parallel training data, in order to improve the performance of a statistical machine translation system. They developed a method for discovering parallel sentences in comparable, non – parallel corpora, by extracting parallel data from Chinese, Arabic, and English non – parallel newspaper corpora, and using the extracted data to improve the performance of the statistical machine translation system. (Munteanu, 2005) However, there are limitations to this method. In order to obtain useful results, the amount of parallel data needed to obtain good quality comparable corpora needs to be rather large, which is not applicable to languages such as Croatian, which are less represented online and generally less represented worldwide.

To conclude, machine translation is still not advanced enough as to face the challenge of polysemy. Semantic notions like polysemy can only be truly comprehended, and thus properly rendered, by a human who is capable of differentiating semantic and pragmatic meaning. However, machine translation is still extremely useful as an aid to translators in

many different situations, especially in rendering large amounts of text in a short amount of time. Therefore, machine translation should be used carefully with constant supervision to avoid such mistakes, at least until technology advances so as to understand pragmatic meaning.

6. REFERENCES

- Arnold, D., et al. (1994). *Machine Translation: An Introductory Guide*: Oxford: NCC and Oxford Blackwell.
- Munday, J. (2008). *Introducing Translation Studies*. Hoboken: Taylor and Francis.
- Anić, V. (2007). *Rječnik hrvatskoga jezika*: Novi Liber
- Brala – Vukanović, M. (2013). *Perspectives on Meaning*: Faculty of Humanities and Social Sciences of the University of Rijeka, English Department
- Pritchard, B. (2002). „Strojno prevođenje: izazovi i ograničenja“. *Primijenjena lingvistika – izazovi na početku XXI. Stoljeća*: zbornik.
- Hutchkins, W. J. (2010). *Machine translation: a concise history*. *Journal of Translation Studies*, vol.13, nos.1-2
- Manerko, L. (2014). *Concept understanding in cognitive linguistics and cognitive terminology science*. In G. Budin & V. Lušicky (eds.), *Languages for Special Purposes in a Multilingual, Transcultural World, Proceedings of the 19th European Symposium on Languages for Special Purposes, 8-10 July 2013, Vienna, Austria*. Vienna: University of Vienna, 471-483.
- Munteanu, D. and Marcu, D. (2005). *Improving Machine Translation Performance by Exploiting Non-Parallel Corpora*. *Computational Linguistics*, 31(4), pp.477-504.
- Nemickienė, Ž. *Concept in Modern Linguistics: the Component of the Concept “Good”* [Online] Available: <http://dspace.kauko.lt/handle/1/817Z>
- Nirenburg, S. (1993). *Progress in machine translation*. Amsterdam: IOS Press.
- https://en.wikipedia.org/wiki/Transfer-based_machine_translation , accessed on June 24, 2017
- <http://www.etymonline.com/index.php?term=chapter> , accessed on August 27, 2017
- <http://www.etymonline.com/index.php?term=sole> , accessed on August 27, 2017

7. TABLE OF FIGURES

Table 1 Translations of the word korijen	9
Table 2 Translations of the word ključ	111
Table 3 Translations of the word stolica	133
Table 4 Translations of the word list	144
Table 5 Translations of the word stanica	155
Table 6 Translations of the word glava	177
Table 7 Translations of the word lice	188