Technical Report CS2016-01

The Efficacy of Esperanto as a Pivot Language in Statistical Machine Translation

Jake Ballinger

Submitted to the Faculty of
The Department of Computer Science

Project Director: Dr. Janyl Jumadinova
Second Reader: Dr. Robert Roos

Allegheny College
2015

I hereby recognize and pledge to fulfill my responsibilities as defined in the Honor Code, and to maintain the integrity of both myself and the college community as a whole.

Jake Ballinger

Contents

List of Figures

Chapter 1

Introduction

The primary use of language is to communicate information. However, when humans do not share a common language, information must be translated. When there is a bilingual human present who is able to effectively translate and preserve the clarity and integrity of the information, the translation problem presents few difficulties. Unfortunately, highly capable bilinguals are not always present to provide translation services, nor are there always sufficient resources to compensate for their services. These human translators are often pruned by the quality of their translations; poor translations do much more harm than no translation.

1.1 The Importance of Translation

The importance of an accurate translation cannot be understated. Consider the problem of translating the Qur'an into English. In a paper by Dehlia Sabry and Ibrahim Saleh, the dangers of a problematic translation are clear: "...the libels levelled against Islam are deeply rooted in the misconceptions propagated by the first Latin Qur'an translations perverted on purpose out of fear that Islam would shake the established faith of Christians." [65] In other words, human translators purposely changed the translations of the Qur'an in order to subdue what they saw as a threat to Christianity. In doing so, they essentially fabricated the holy book and used that as a basis for the long history of tension between the Christians and Muslims of the Middle Ages, an enmity which has continued to the present. Sabry and Saleh comment on another poor translation concerning the treatment of women as laid out by the Qur'an, and thus challenge the harmful misconception that Islam degrades women because of their sex. Consider two translations of *aya* 34 of *sura* 4 from the Qur'an:

> "Men have charge of women because Allah has preferred the one above the other and because they spend their wealth on them." [10]
>
> "The men are supporters of the women, by what God has given one more than the other." [20]

The first translation perpetuates the idea that women are to be subservient to men; the second—the more accurate translation, according to Sabry and Saleh [65]—teaches men that women should be respected as free-thinking individuals and were not created to be treated as second-class citizens. How radically different these ideas are!

The accurate translation of literature is also extremely important in an increasingly global society. Straumanis tells the story of how a Latvian publisher became interested in acquiring the rights to a Latvian translation of a work—a novel called *High Tide* by Ethopian author Inga Âbele—that was in fact a translation from Amharic to English [68]. How many Latvian-Amharic translators are there? Probably not many, but there are many translators from Amharic to English, and many again from English to Latvian. Had someone never translated Âbele's book to English, it never would have reached its Latvian-speaking audience. Had no group of people ever bothered to explore the translation of literature and culture from one nation to another, not only would culture be lost, but economic growth would not be as efficient as it currently is because goods would not enter foreign markets so seamlessly. In a different case, Pulitzer-prize winning author Jhumpa Lahiri, who normally writes in English, decided to learn a foreign language (Italian) and write exclusively in it; without translation, the English-speaking market would never again read her works [45]. Perhaps more evident is the loss of the great classics—the works of Ovid, the writings of Moses, and the epics of India would be lost with their worlds. Without translation, the West never would have found Yoga, and the scientific and mathematical knowledge of the Greeks and the Arabs never would have made the European scientific revolution possible.

That said, there are some pitfalls that translation can avoid. For example, when Schweppes tried to introduce their tonic water to the Italian market, they found that "tonic water" translated to "toilet water"—as one can surmise, sales did not take off, and the brand lost money in the blunder [31]. Other examples of these international branding mishaps include the American Motors Matador, a car that was introduced to Puerto Rico. Unfortunately, in Spanish, the word *matator* translates to *killer*, which did not inspire confidence in the Puerto Rican market. Another blunder, also by a car manufacturer, was when Ford tried to introduce their car to the Belgian market with the slogan "Every car has a high-quality body." However, when translated, the slogan read, "Every car has a high-quality corpse"—not quite the best marketing campaign [13]. All of these could have been avoided by a simple translation check.

There are more benefits to translation—The Guardian suggests that translation can help preserve dying languages, and efforts to manually translate endangered languages like Quechua are already underway [18, 75]. Perhaps even more importantly, since "Millions of Latin Americans lack health, employment or education services because they do not speak Spanish but instead one of the hundreds of indigenous languages of the region"[18], translation helps governments provide basic services for their people who may not speak the national langage. Linguistic discrimination based on one's native language is a very real experience, and research shows that preserving one's native language leads to personal, economic, social, intellectual, and educational advantages [35]. Lahiri also notes the strange isolation of living without one's native language: "When you live in a country where your own language is considered foreign, you can feel a continuous sense of estrangement. You speak a secret, unknown language, lacking any correspondence to the environment. An absence that creates a distance within you." [45]

Not all languages enjoy equal status. For example, the politico-linguistic situation on the African continent is strikingly unbalanced. Africa is home to more than 2,000 tongues, but only 242 of those are used in the mass media, only 63 are used in the judicial system, and only 56 are used in public administration [40, 59]. Since there is such a disparity of language in Africa, and since there is no one language that unifies Africa, other avenues, such as machine translation, must be explored. In fact, according to a report by Kelly et al., published by the Common Sense Advisory, "Translators of African languages may benefit greatly from machine translation advances if they view it as an additional productivity tool to add to their arsenal." [40] The report concludes that "Translation Will Power Africa's Future Socioeconomic Development" [40]. The phenomenon is true even on the level of individual countries, for Spanish does not unite Spain, Mandarin does not unite China, Portuguese does not unite Brazil, and Hindi certainly does not unite India.

1.2 Machine Translation

In an effort to provide translation services to the general population, many machine translation systems have appeared, most notably Google Translate. Google Translate, like many other web-based services, uses a data-driven approach to decode language;

that is, Google Translate treats the input text as something that has probably been said before and combs the *parallel corpora* of the web—two texts in different languages, one of which is a translation of the other—to determine the best translation [7]. Most machine translation systems are like Google Translate—automatic translation tools, commonly found online, that use statistical methods of create mappings between the user-input texts and the most likely output translation. While Google Translate and its brethern improve every year, they are still known to make errors. And indeed, sometimes these errors are harmful, as Lavoipierre notes was the case when the Australian State Department tried using machine translation to translate tweets about ISIS into Arabic [46].

This brings up an important point: when is machine translation necessary? Opinions are vastly diverent. According to an article from the Australian Institute of International Affairs, "Google translate can also be used for informal or casual communication. . . Anything official within a company would risk miscommunication or other related issues."[64] Others are not so strict, such as LondonTranslations, a London-based translation firm, which recommends that "If comprehension—rather than 100% accuracy—is the end goal of an exercise, machine translation can provide an effective way of getting rapid results."[3] Meanwhile, the U.S. Department of Labor advises that "It is seldom, if ever, sufficient to use machine translation without having a human who is trained in translation available to review and correct the translation to ensure that it is conveying the intended message" [54] and that machine translation is only approrpriate in emergency services (with the help of a human translator), leaving one wondering why bother using the machine at all. Another translation company, the Comprehensive Language Center, states that "Machine translation is best when used for unofficial and informational purposes or when material is perishable but too voluminous to be translated by humans . . . Machine translation is also useful for technical materials that use very consistent terminology and writing styles."[37] They do not recommend machine translation for material containing "nuanced or sensitive information, such as legal agreements, marketing material, business correspondence, or hand-written text." [37] As such, opinions on the proper usage of machine translation vary widely, and there is no real consensus on whether or not its usage is appropriate or inappropriate, except that it is not appropriate to process sensitive information where accuracy is the end goal.

1.2.1 Ethical Concerns

A strong criticism of machine translation is that it will do to human translators what automated tools did to factory workers in the 19^{th} and 20^{th} centuries—replace them. Even writers at MIT have pondered this [63]. However, these fears are largely exaggerated; reporters from the Huffington Post [38], The Guardian [39], and The Economist [76] cast their doubts. These fears do not have to exclude the automatic translation of spoken text, but according to Philipp Koehn, the chair of the Machine Translation research group at the Johns Hopkins University in Baltimore, MD, "Automatic spoken translation is a particular problem because you're working with two imperfect technologies tied together—speech recognition and translation." Thus it doesn't seem as though interpreters will be out of work in the near future, either [76].

Written works, which can be fed into a computer and translated much more quickly than a human could hope to do, present their own set of concerns. In this case, I point to the current quality of translated materials, which exhibit a statistical phenomenon known as translationese. (See Section 2.1.5 for more information about translationese.) Computers simply do not have the ability to refactor beautiful prose in one language into nuanced, elegant writing in another like humans do.

Critics of machine translation point out that it encourages harmful social ideas that devalue less-common languages and perpetuate the misconception that English is the only language worth learning, since everything can be translated into it [38]. Although a moral and ethical argument supporting machine translation practices is well beyond the scope of this thesis, I point out that machine translation can be invaluable in removing human biases implicit in human translations of sensitive texts (such as in the previously-cited example of the Qur'an); additionally, the machine translation of less-common languages makes the culture of that language much more accessible to the greater world population, which in turn can encourage others to learn that language or to help inform international policies in regions where that language is spoken, for language informs culture and culture informs law.

These limitations notwithstanding, machine translation systems have shown rapid improvements from their introduction. Although machine translation was first introduced in the 1960s, the idea of a data-driven approach to machine translation, whereby the quality of the output is directly influenced by the quality of the input, was first proposed by Brown et al. in 1990 [14]. These inputs are in the form of parallel corpora, and this data-driven method to machine translation is known as *statistical machine translation*.

1.3 Pivot Translation

It is a principle of statistics that the more data one has, the more reliable statistical predictions are. This is one of the reasons that college admissions committees are more heavily weighing the high school GPA of applicants than their standardized test scores [33]—the GPA is influenced by several data points (final grades in a class), which in turn were composed of even more data points (the individual assignments in a class), but standardized test scores are incidental[1]. We can map this understanding to statistical machine translation: the more data the system has, the better the translation will be.

Consider the problem of translating language L_a into language L_b, where there exists little to no shared literature or parallel corpora already between them, such as the previously-cited example about the novel called *High Tide*, which was translated from Ethopian to English to Latvian. Using current statistical methods, the likelihood of obtaining a stellar translation is low [77]. Now consider that there exists some language with which both L_a and L_b share a large amount of parallel corpora, the pivot language L_p. Statistically speaking, it would make much more sense to first translate L_a into L_p and then from L_p into L_b [77]. The pivot model was first proposed by Wu and Wang in 2007 [77].

The trade-offs of using a pivot language have been well-discussed in the literature [61, 72, 77, 78]. A study by Dr. Michael Paul and his colleagues [61] determined that "the selection of the optimal pivot language largely depends on the SRC-PVT and PVT-TRG translation performance"; that is, if the pivot language performs well when translated from the source language and performs well when translated into the target language, it is a good pivot. Consider the problem of translating between Arabic and Basque, which is becoming more necessary as the Basque Country (a large region of Northern Spain) is experiencing an influx of Arabic-speaking immigrants [74]. In this case, it might make sense to use Spanish as a pivot language, given the geographic proximity of Spanish to both Basque and Arabic. However, in other instances, such as when translating between Berber and Vietnamese, parallel corpora are rarer, less extensive, and the choice of the pivot language lesss obvious.

[1]Interested readers should reference the important statistical theorm called the Central Limit Theorm, which essentially states that for any normal distribution (i.e., the graph of all the plot points is a bell curve, like IQ), as the sample size increases, so does the accuracy of one's predictions.

Another consideration when identifying a viable pivot language is the consolidation of resources. Consider the European Union, which has 24 official and working languages [21, 62]. It is not realistic to expect every delegate sitting on the EU to be fluent in all 24 of these languages. It would also be difficult—and overwhelming—to have multiple machines or humans concurrently translating documents between all of these languages [28], as often in the EU, fast conversations need to happen between speakers of different languages [23]. Unfortunately, the speech-to-text problem is nowhere near reliable enough to provide real-time translations, and thus combining that technology with machine translation may lead to disastrous miscommunications [76]. In fact, there would need to be $\binom{24}{2}$ or 276 connections and systems set up to do all of these translations. The problem can be simplified to $\binom{23}{1}$, or 23, connections through one larger system by choosing a common pivot language—usually English—which can serve as an intermediary to all of these languages.

Although English is the most commonly-used pivot language, another language called Esperanto, which was developed in 1887 as an international auxiliary language, may be better-suited to serve as a pivot language. Esperanto's grammar is highly regular and rarely ambiguous, two strengths it has over English. But while the linguistic case for Esperanto is strong, the computational case is lacking in evidence—one large concern is whether or not there exists enough data in the form of parallel corpora between Esperanto and other languages to produce quality translations. Yet there has been no study to determine Esperanto's efficacy.

This study determines whether Esperanto performs well as a pivot by using the Moses statistical machine translation system. For more information about Moses, see [42]. This system uses a Recurrent Neural Network Language Model (see Section 2.2.4) to create language models of Esperanto and the test languages. It then translates from a test language into Esperanto, and from Esperanto into a target language.

Translation equality is evaluated with a BLEU score [60], a METEOR score (designed to fix the weaknesses inherent in the BLEU metric) [4, 25], and through human appraisals of the translation quality. If Esperanto performs better as a general-case pivot language, we can expect pivot-based machine translation systems to improve.

Chapter 2

Related Work

2.1 Statistical Machine Translation

Brown, in his genesis of statistical machine translation, wrote that he only considered the problem of translating sentences. More importantly, he was not concerned with reaching the "pinnacles of the translator's art," or the artistic nuances that good translations often exhibit [14]. In the beginning, the expectation was only that the computer would render a comprehensible translation, but nothing of great quality.

But Brown framed the problem in a way that might seem counter-intuitive: He considered that a sentence in a language is a possible translation of any sentence in another. To say that *J'adore ma chatte* and *Last night we watched the fireworks from our neighbor's back porch* are translations of each other would raise some eyebrows, but in the strictest mathematical sense we can still assign a probability to this pair. Intuitively, we assign this pair the probability of T given S; that is, $\mathrm{P}(T|S)$, or the probability of producing sentence T in the target language when given sentence S in the source language.

In machine translation, this is flipped. Thus, instead of searching for $\mathrm{P}(T|S)$, we search for $\mathrm{P}(S|T)$, or the probability of producing sentence S in the source language when given sentence T in the target language. Imagined differently, we seek the sentence from which the translator produced the translation. Mathematically, we can model this conditional probability using Bayes' rule:

$$P(S|T) = \frac{P(T|S) \times P(S)}{P(T)}, \tag{2.1}$$

where we seek to maximize the numerator $P(T|S) \times P(S)$.

To frame the problem in this way, we need to create **language models** (See Section 2.2 to create those probability distributions. Brown calls the first factor, P($T|S$), the *translation probability of* T *given* S, and the second factor, P(S), the *language model probability of* S. Essentially, the translation probability suggests words from the source language that probably produced the words in the target language translation, while the source language model established the rules for arranging these words in a sentence.

2.1.1 Pivot Statistical Machine Translation

The use of a pivot language is described by Wu and Wang in 2007, where they note its helpfulness when there exist little parallel corpora between two languages (L_S and L_T) but where there also exists a large amount of parallel corpora between both the L_S and L_P (the source and pivot languages) pair and the L_P and L_T (the pivot and target languages) pair [77].

Wu and Wang outlined their translation steps as follows: (1) segment a sentence into phrases, (2) translate each of the phrases into the target language according to the phrase translation distributions, and (3) reorder the translated phrases according to some distortion model. They formalized the phrase "translation probability" as a generative probabilistic process.

Wang and Wu observed that using more than one pivot language to improve translation performance is a possibility and an easy next step, saying that the ambiguities present in one language are often accounted for in another, and thus the overall translation quality for the language pair improves [77]. The combination of pivot models is called *linear interpolation*. Interpolated models can even use the original source and target languages, regardless of how small the parallel corpora between them are. Perhaps most interestingly, Wang and Wu found that a pivot language in a different language family produced a better translation than a pivot language in the same or in a similar language family. This was corroborated by Snyder and Barzilay in 2010, when it was extended to general NLP (natural language processing) tasks [66]. Overall, Wu and Wang found that the pivot model outperforms the standard model training on a small bilingual corpus, and their results consistently showed that more data correlated with better translations

Wu and Wang revisited their pivot language approach for machine translation two years later in a 2009 paper [78]. Here, they proposed a hybrid pivot model using a rule-based machine translation (RBMT) system and a normal statistical machine translation (SMT) system and found that the hybrid outperformed a normal pivot model, making RBMT systems especially useful when the corpora are independent of each other [78]; i.e., when note using parallel corpora. The finding that hybrid models are useful was also confirmed by a study seeking to translate between two dialects of German [56].

Typically, English is used as a pivot language, and quite successfully [23]. What, then, would happen if a language other than English was used? In a study that solidified the findings of Wu and Wang (that different language families were complementary towards each other), researchers Paul et al. found that using SMT techniques to translate between twelve languages revealed that the translation quality for 61 out of 110—or slightly more than half—improved when a non-English pivot was used [61]. Thus there is a precedent for the use of a non-English pivot language.

2.1.2 Pivot Phrase-based Statistical Machine Translation

In statistical machine translation (SMT), phrase-based models outperform word-based and sentence-based models [43, 72]. A study by Koehn, Och, and Marcu found that the best SMT performance can be achieved by using heuristic learning of phrase-based translations from word-based alignments and lexical weighting of phrase translations [43]. In simpler terms, by following certain rules of thumb and designating some phrases as more likely to occur than others, the system will produce a better translation. However, Koehn et al. also noted that what constitutes the best heuristic varies between language pairings and the size of the training corpus.

Koehn et al. carried out experiments to compare the performance of three different methods to build phrase-translation probability tables. First, they learned phrase alignments[1] from a corpus that was word-aligned using Giza++, a word-alignment tool. Second, they used a word-aligned corpus annotated with parse trees, generated by statistical syntactic parsers[2]. Third, they directly learned from phrase-level alignments of the parallel corpora [43]. The results showed that learning all phrases consistent with the word alignment was the best method, even with small phrases of at most three words.

2.1.3 Pivot Sentence-based Statistical Machine Translation

While the phrase-based technique is concerned with translation individual phrases and treats sentence endings as delimiters between phrases, the sentence-based technique ignore the concept of "phrases" altogether and sees the entire sentence as a unit. Utiyama and Isahara found that phrase-based techniques significantly outperform

[1] Alignments create a correspondance between words in language A and words in language B. (This kind of mapping is not necessarily one-to-one or onto.) [15]

[2] A syntactic parser works out the grammatical structures of sentences in a given language. It is statistical because it is data-driven; i.e., there is no linguist or programmer who prescribes rules for it to follow [79]

sentence-based techniques [72], and thus I excluded them from my model.

2.1.4 n-gram-based Machine Translation

Researchers from the Universitat Politècnica de Catalunya (the Polytechnic University of Catalonia) present what they call the Tuple n-gram Model [51]. According to the authors—henceforth referred to as Mariño et al.—a **tuple** is bilingual unit—a sample of "bilinguage" [22]. Hence, researchers are able to model transition model probabilities at the sentence level using n-grams of tuples, described by the following:

$$p(T, S) \approx \prod_{k=1}^{K} p((t, s)_k | (t, s)_{k-1}, (t, s)_{k-2}, \ldots, (t, s)_{k-n+1}), \tag{2.2}$$

where t refers to the target, s to the source, and $(t, s)_k$ to the kth tuple of a given bilingual sentence pair. In plain English, this means that the probability of getting the tuple (T, S), where T and S are sentences, is approximately equal to the conditional probability of getting the bilingual tuple (t, s), where t and s are words in T and S respectively, given all of the previous bilingual tuples in the those sentences. For our earlier example of the French *J'adore ma chatte* $= F$ and the new English sentence *Mary eats fish* $= E$, we can say:

$$p(F, E) \approx \prod_{k=1}^{K} p((f, e)_k | (f, e)_{k-1}, (f, e)_{k-2}, \ldots, (f, e)_{k-n+1}), \tag{2.3}$$

can be mapped to (keeping in mind that f and e are words such that $f_i \in F$ and $e_i \in E$)

$$p(F, E) \approx p(chatte, fish) | p(ma, eats) \times p(J'adore, Mary) \tag{2.4}$$

the probability of which would be close to zero.

Mariño et al. extract tuples from word-to-word bilingual aligned corpora [51]. In their study, Mariño et al. found that the single best translation, as determined by BLEU scores, was n-gram-based; however, the majority of the best translations were from pivot-based systems. Thus, we can attribute the n-gram method a higher level of variability than the pivot paradigm.

Then it seems the trade-off can be condensed into the following consideration: Does the researcher value greater variability in translation quality, with a few very good translations and a few very bad ones, or would they rather opt for a more reliable and still high-performing system? In this case, given that the best n-gram translation did not significantly nor consistently outperform the best pivot translation, the greater consistency of the pivot model make it the more appealing choice.

The practice of using a pivot language as a way of rectifying the poor qualities of translation was developed several years later, in 2007 [77]. With this technique, the source language L_S is not directly translated into the target language L_T but instead passes through an intermediate, or pivot, language L_P. English is commonly used as a pivot language given its current status as the world *lingua franca*. However, there are some linguistic reasons why English is not ideally suited for this purpose; mainly, its grammar is replete with exceptions and irregularities, and English-speakers are often ambiguous in their discourse.

2.1.5 Translationese

A great many works have been published pointing out the artificiality of translated material, so much that it has been given its own moniker—*translationese* [29, 70, 69]. The two most notorious aspects of translationese are *explicitation* and *simplification* [2, 12, 11]. In fact, these differences are so striking that they can be automatically classified with alarmingly high accuracy [5, 36, 44]. This is an active area of research, and currently being improved upon by adapting translation models to translationese

[48]. This means that a good next step, after this study, would be to adapt the pivot model with Esperanto to account for the translationese effect.

Lembersky, in his PhD Thesis at the University of Haifa, noted that the two major factors that influence the quality—and therefore recognizability—of translationese are the language model and the translation model. Lembersky and Twitto both noted that accounting for the effect of translationese in the training data dramatically improves the quality of the machine translation system [48, 71].

After translation, there exists the problem of correcting the grammatical errors in machine translations that exist in the target language but not in the source [34]. At the very least, human translators will be needed to look at the source material and use that to correct and improve upon the machine translation.

Finally, although the purpose of language is to communicate information, it is important to remember that language is an art form. The machine is only concerned with making sure the nuts and bolts of the language are held together, but it has no concept of how to make musical or poetic phrases, nor does it take artistic liberties. Human translators will need to continue to be artists so that the great works of literature are not lost in machine translation.

2.2 Language Models

The language model, defined as the probability distribution over a sequence of words, estimates the *a priori* probability of a sentence in the target language [47]. (This is the same as Brown's equation of conditional probability introduced in Section 2.1.) Most language models today are n-grams that model language as a Markov chain[3] of order $n-1$; these language models were developed for the problem of speech recognition

[3] A Markov chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. In the case of n-grams, each word only cares about the n words that directly proceed it.

and were later adopted by the machine translation community [1, 14]. Recently, the pendulum has begun to swing in the favor of the neural network language model (NNLM). In this section, I give a cursory introduction of the neural network, and then I explore different neural network representations of language models.

2.2.1 Neural Networks

A neural network, more properly termed an *artificial neural network*, is an artificial, mathematical, and statistical represention of a neuron in a human brain. Human neurons have inputs receptors (called dendrites) which send information to an activation function (in the cell soma) which has some activiation threshold (usually $-55mV$) that much be transcended before the act of firing (the action potential). The action potential travels through the axon and the terminal buttons to the next neuron. In the simplest case, in biological neurons a neurotransmitter (a chemical) is released at the terminal buttons and activates receptors on the dendrites of the receiving neuron; this pattern continues until a neuron's threshold is not transcended (i.e., the action potential fades out) or the action potential reaches the final neuron [6].

Biological and psychological theories of learning mostly ascribe to what is called Hebbian learning—"cells that fire together wire together" [17]. The idea is that the most-commonly used connections become the most powerful connections. The artificial neural network represents this in terms of *weights* [67]. Thus the output of each neuron in the artifical neural network is given some weight, which determines its efficacy on the next neuron. In this way, artifical neural networks are very powerful and finely tunable.

We represent the input of the basic artificial neuron as follows:

$$weightedsum = \sum_{i=0}^{n} x_i w_i \tag{2.5}$$

where the *weighted sum* corresponds the the sum of all of the inputs x_i multiplied by their weights w_i.

2.2.2 Bengio's Neural Network Language Model

Let's briefly revisit the n-gram model. Recall from Section 2.1.4 that:

$$p(T, S) \approx \prod_{k=1}^{K} p((t, s)_k | (t, s)_{k-1}, (t, s)_{k-2}, \ldots, (t, s)_{k-n+1}), \tag{2.6}$$

which, when turned monolingual, we can rewrite as the monolingual n-gram equation:

$$P(w_t | w_1^{t-1}) = P(w_t | w_{t-n+1}^{t-1}) \tag{2.7}$$

where w_t is the t-th word and $w \in$ some sentence W and $w_i^j = (w_i, w_{i+1}, \ldots, w_{j-1}, w_j)$ [9]. Conceptually, the n-gram model is designed to take advantage of the fact that temporally closer words in a sentence or word sequence are statistically more dependent; in other words, the n-gram model constructs tables of conditional probabilities for the next word based on what Bengio calls the *context*—combinations of the last $n - 1$ words. But there is an obvious problem with this approach—what happens when the system encounters a new sequence of n words that wasn't in the training corpus? The language model should do more than determine the validity of a word based on its preceeding words—it should also help us generate phrases that might be similar based on parts of speech; i.e., it should help us generate other grammatical sentences. For example, the sentences *Last night my sister's best friend roasted marshmallows at the firepit* and *Tuesday my brother's favorite alpaca gnawed apples underwater* have the same linguistic structure (Time Phrase + Subject + Verb + Object + Preposition), and while one is ridiculous, it is still grammatical. Indeed, the pairs {(Last night, Tuesday), (my sister's best friend, my brother's favorite alpaca), (roasted, gnawed), (marshmallows, apples), (at the firepit, underwater)} all play similar semantic and grammatical roles, so there is no reason we cannot interchange them to create equally grammatical—if not equally nonsensical—sentences like *Tuesday my sister's best friend gnawed marshmallows underwater.* Bengio calls this problem the *curse of dimensionality* [9].

The use of neural networks to model high-dimensional discrete distributions (like natural language) to learn the joint probability of $Z_1 \ldots Z_n$, a set of random variables with possibly different natures, was already proposed by Bengio in an earlier paper and found to be decomposed as a product of conditional probabilities [8]. Bengio's approach had three tenets:

1. Associate each word in the vocabulary with a distributed word feature vector,
2. Express the joint probability function of word sequences in terms of the feature vectors of these words in the sequence, and
3. Simultaneously learn the word feature vectors and the parameters of that probability function.

Essentially, we associate each word with a point in a vector space. The probability function is a prouct of conditional probabilities of the next word given the previous ones (much like the n-gram).

For the general language model, let the training set be a sequence $w_1 \ldots w_T$ of words $w_t \in V$, in which the vocabulary V is a large but finite set. The objective is to learn a good model M: $f(w_t, \ldots, w_{t-n+1}) = P(w_t|w_1^{t-1})$ that gives a high out-of-sample likelihood, where the function f is a composition of two mappings ($V- > \mathbb{R}$ and from the neural network to some distribution of random variables $Z_1 \ldots Z_n$).

Bengio found that this model yielded a 10-20% difference in perplexity[4] when

[4] *Perplexity* is the geometric average of $\frac{1}{P(w_t|w_{t-1}^1)}$.

compared with the trigram (3-gram, the most commonly used n-gram).

2.2.3 Neural Network Language Model for Low-Resource Languages

Unfortunately, Bengio's model assumes an abundance of data—Esperanto does not have a billion-words worth of annotated corpora for this type of neural model. Gandhe et al. proposed a solution to this data problem [27]. By using a feed-forward neural network language model (ff-NNLM), and subsequently training the ff-NNLM differently, they are able to subvert the curse of dimensionality and minimize the training time. In this model, the probability of the current word depends on the word history of n words, computed as an n-gram:

$$p(w_j|history) = p(w_j|w_{j-1}, w_{j-2}, \ldots, w_{j-n+1}) \tag{2.8}$$

In this way, history is presented in the ff-NNLM by many sparse vectors at the input layer. Each vector is then linearly mapped to a continuous word projection by a $N \times P$ weight matrix (F)—this concatentaiton of continuous word vectors results in a *projection layer*. The second layer—the hidden layer—uses a hyperbolic tangent function[5] as a nonlinearity. The output layer has N units (where N is the size of the vocabulary of the model). In order to produce the posterior probabilities, Gandhe applied a softmax function to the activiation of each unit to ensure that the probabilites all sum to 1 [27].

[5] The hyperbolic tangent function can be used where a sigmoid function would normally be used in a neural network [58].

2.2.4 Recurrent Neural Network-Based Language Model

More recently, Mikolov proposed the most state-of-the-art NNLM—the *recurrent* neural network language model (RNNLM), the main difference in which is the representation of history [53]. The advantages of the RNNLM over the ff-NNLM are as follows:

1. History: For the ff-NNLM, history is the previous $n-1$ words, but for the RNNLM, history is contained in the hidden layer of the neural network and is therefore much more comprehensive.

2. Pattern Recognition: The RNNLM can represent more advanced patterns in sequential data than the ff-NNLM, meaning that patterns relying on words that could have occurred at variable positions in the history can be much more incoded (example: consider the problem of placing the word "only" in the sentence "he told him he was sorry"[6]).

3. Training Time: Following Mikolov's thesis [53], training the RNNLM can take less time than training the ff-NNLM, despite the larger hidden layer.

The details of Mikolov's RNNLM and backpropogation through time (BTTT) algorithms can be found in [53]. I used Mikolov's RNNLM toolkit—the same one he used in [53] to create his language models [52].

2.3 The Case for Esperanto

Esperanto is a constructed (planned) language invented by Leyzer Ludwik Zamenhof in 1887 [80]. This section outlines why Esperanto is a good candidate for a pivot language in an SMT system and addresses concerns about its viability.

[6] Although this is a problem only in colloquial English, because in proper, written English we place the word "only" close to the word that only "only" modifies, as opposed to only placing the world "only" next to the word that "only" modifies. Another example: an earlier version of this footnote opened as "Although this is only a problem in colloquial English".

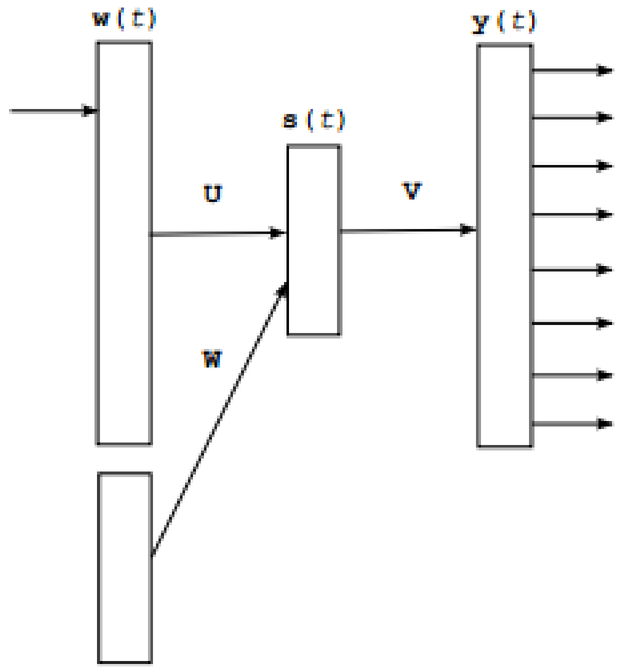

Figure 2.1: A simple recurrent neural network. Taken from Mikolov's PhD thesis [53].

As Wu and Wang noted, and as was later confirmed by Snyder abd Barzilay, the lexical ambiguities present in one language may be missing in another language, thus improving the translation quality [66, 77]. This finding was particularly strong when the pivot language was in neither the source nor target language family. Esperanto is a constructed language: It has no organic language family.

As a constructed language, Esperanto's classification is a little ambiguous: It is very lexically Romantic, but somewhat agglutinative (a feature of languages such as Turkish and Finnish, where suffixes and endings can be added to words and roots to create a string of characters representing a new word) and has 28 letters. Additionally, Esperanto contains some features not seen in Romance languages—for example, the accusative case (which is used in German). Words are derived by stringing together prefixes, roots, and suffixes, and compound words are formed as they are in English, with the modifier first. Regardless, the official classification of Esperanto is likely not important, for Snyder's observation that languages in different linguistic families are complimentary (in terms of machine translation) references the biological evolution of that language and less so the human classification it was given [66].

As Esperanto is a relatively new language, one might be concerned that it is lacking in resources for statistical machine translation. These concerns may include incidences of parallel corpora or whether the language lends itself well to NLP tasks. However, according to Dellert, "since the language has developed into a full replacement for natural languages in all situations, all the aspects of semantics and pragmatics that NLP (natural language processing) wants to address are present in Esperanto as much as in any natural language" [24]. Additionally, as van Cranenburgh and colleagues point out, "Although it was designed as an easy-to-learn language, with regular and transparent syntax and morphology, its semantic and pragmatic components have evolved naturally," [73] indicating that Esperanto is a viable candidate for NLP tasks.

The fact that Esperanto is now a possible language for Google Translate indicates that there are sufficient parallel corpora available to create a operable translation system with it and many other languages [16]. This suggests that Esperanto is indeed a viable candidate [49]. Additionally, since everything written in Esperanto was either first written in Esperanto and then translated into the author's native language or written in the author's native language and then translated into Esperanto, the phenomena of "translationese" should be greatly reduced [70].

2.3.1 Esperanto's Linguistic Properties

Like many languages, such as English, French, and Chinese, Esperanto most commonly uses an S-V-O (Subject-Verb-Object) word order. However, unlike most other languages, Esperanto includes an accusative case, where direct objects and their modifiers (i.e., adjectives) are marked with an -n suffix. All plural forms are marked with a -j suffix, and thus any plural direct object will end in -jn. Additionally, verbs have no conjugations—the verb for *to be*, *esti*, will never change except for to indicate a change in tense. Thus the complex rules of conjugations so many students struggle with—like the French verb for *to be*, *être*, whose present-tense conjugations are *je suis, tu es, il/elle est, nous sommes, vous êtes, ils/elles sont*—are nonexistent in Esperanto: in the present, the one drops the -i suffix in *esti* to form the root *est*, and add the -as suffix; thus, everyone simply *estas*. I am *mi estas*, you are *vi estas*, he/she/it is *li/ŝi/ĝi estas*, we are *ni estas*, and they are *ili estas*. There are no exceptions to these rules.

Transitive/intransitive verbs have their own markers: verbs with the *-ig* suffix are transitive; ie, they require a direct object (marked with the -n suffix). Likewise, verbs with the *-iĝ* suffix are intransitive, and can never take a direct object. This is another aspect of Esperanto's expressive power. For example, the work *weak* in Esperanto is *senforta*, a compound word coming from the prefix *sen*, or without, and the root word *forta*, or strong. (All adjectives in Esperanto end with an -a.) Without strong, (i.e., without strength) means weak. And much like in English, we can modify this word to create two new words which can also be expressed in English: to weaken (*senfortigas*) and to become weak (*senfortiĝas*). The reader can probably split these words into their constituent parts: *sen* (without), *fort* (strength/strong), *ig* (transitive marker), *as* (present tense verb marker). This way of building words contributes to a highly regular system of grammar.

Esperanto expands on this word-building through its correlative system, where 50 common question words that require specific answers (what, how, when, who, etc.) and their general answers (something, in no way, then, everyone, etc.) are constructed by combining a set of predefined prefixes and suffixes [32].

	Question (K)	Pointer (T)	Indefinite (I)	Universal (Ĉ)	Negative (N)
Individual (IU)	KIU who, which	TIU that one	IU some(one)	ĈIU every(one)	NENIU no one, none
Thing (IO)	KIO what	TIO that thing	IO something	ĈIO everything	NENIO nothing
Kind (IA)	KIA what kind of	TIA that kind of	IA some kind of	ĈIA every kind of	NENIA no kind of
Place (IE)	KIE where	TIE there	IE somewhere	ĈIE everywhere	NENIE nowhere
Motion (IEN)	KIEN where to	TIEN there	IEN somewhere	ĈIEN everywhere	NENIEN nowhere
Time (IAM)	KIAM when	TIAM then	IAM sometime	ĈIAM always	NENIAM never
Amount (IOM)	KIOM how much, how many	TIOM so much, so many	IOM some	ĈIOM all	NENIOM no amount
Manner (IEL)	KIEL how	TIEL so	IEL somehow	ĈIEL in every way	NENIEL in no way
Reason (IAL)	KIAL why	TIAL so	IAL for some reason	ĈIAL for every reason	NENIAL for no reason
Possession (IES)	KIES whose	TIES that one's	IES somebody's	ĈIES everybody's	NENIES nobody's

Table 2.1: Correlatives in Esperanto

2.3.2 Concerns for Esperanto as a Low-Resource Language

Nakov has shown that it is possible to improve statistical machine translation performance for a low-resource language by using related languages [55]; however, since Esperanto's classification is ambiguous, it has no generally-accepted closest languages. Additionally, given the techniques described by Mikolov and Gandhe, the amount of gathered data is enough that data concerns are not valid.

Esperanto data for this experiment was gathered from the Esperanto Corpus (the *Tekstaro en Esperanto*) [26], Project Gutenberg [30], and the Biblioteca Polyglotta from the University of Oslo [57]. Data from European languages was taken from Koehn's Europarl corpus [41], and the Christodoulopoulos's Bible Corpus was used as a supplement to the above [19]. Finally, the Penn Treebank corpus was used to train the English RNNLM [50].

Bibliography

[1] Lalit R. Bahl, Frederick Jelinek, and Robert L. Mercer. Readings in speech recognition. *Readings in speech recognition*, 1990.

[2] Mona Baker et al. Corpus linguistics and translation studies: Implications and applications. *Text and technology: In honour of John Sinclair*, 233:250, 1993.

[3] Jennifer Ball. Machines and translation: A help or a hindrance? `http://www.londontranslations.co.uk/guides-on-translation/machines-and-translation-a-help-or-a-hindrance/`. Accessed: 06 December 2015.

[4] Satanjeev Banerjee and Alon Lavie. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. `https://www.cs.cmu.edu/~alavie/papers/BanerjeeLavie2005-final.pdf`. 2005. Accessed: 10 December 2015.

[5] Marco Baroni and Silvia Bernardini. A new approach to the study of translationese: Machine-learning the difference between original and translated text. *Literary and Linguistic Computing*, 21(3):259–274, 2006.

[6] Mark F Bear, Barry W Connors, and Michael A Paradiso. *Neuroscience*, volume 2. Lippincott Williams & Wilkins, 2007.

[7] David Bellos. *Is that a fish in your ear?: Translation and the meaning of everything*. Macmillan, 2011.

[8] Yoshua Bengio and Samy Bengio. Modeling high-dimensional discrete data with multi-layer neural networks. In *NIPS*, volume 99, pages 400–406, 1999.

[9] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. A neural probabilistic language model. *The Journal of Machine Learning Research*, 3:1137–1155, 2003.

[10] Abdalhaqq Bewley and Aisha Abdurrahman Bewley. *The Noble Qur'an: A new rendering of its meaning in English*. Madinah Press, 1999.

[11] Shoshana Blum-Kulka. Shifts of cohesion and coherence in translation. *Interlingual and intercultural communication: Discourse and cognition in translation and second language acquisition studies*, 17:35, 1986.

[12] Shoshana Blum-Kulka and Eddie A Levenston. Universals of lexical simplification. *Strategies in interlanguage communication*, 1(983):119, 1983.

[13] Chad Brooks. Lost in Translation: 8 International Marketing Fails. http://www.businessnewsdaily.com/5241-international-marketing-fails.html. Accessed: 13 December 2015.

[14] Peter F Brown, John Cocke, Stephen A Della Pietra, Vincent J Della Pietra, Fredrick Jelinek, John D Lafferty, Robert L Mercer, and Paul S Roossin. A statistical approach to machine translation. *Computational linguistics*, 16(2):79–85, 1990.

[15] James Jonathan Jesse Brunning. *Alignment Models and Algorithms for Statistical Machine Translation*. PhD thesis, Cambridge University Engineering Department and Jesus College, 2010. Accessed: 15 December 2015.

[16] Sean Buckey. Google adds Esperanto as its 64th machine translatable language. http://www.engadget.com/2012/02/23/google-adds-esperanto-as-its-64th-machine-translatable-language/. Accessed: 2015-03-12.

[17] Neil R Carlson. *Physiology of Behavior, 11/E*. Allyn & Bacon, 2013.

[18] Julio Cesar Casma. Discriminados por hablar su idioma natal. http://www.bancomundial.org/es/news/feature/2014/04/16/discriminados-por-hablar-su-idioma-natal-peru-quechua. Accessed: 07 December 2015.

[19] Christos Christodoulopoulos. Bible Corpus. http://homepages.inf.ed.ac.uk/s0787820/bible/. Accessed: 28-03-2015.

[20] Thomas Cleary. *The Essential Koran*. Book Sales, 1998.

[21] European Commission. Official languages of the EU. http://ec.europa.eu/languages/policy/linguistic-diversity/official-languages-eu_en.htm, 2015. Accessed: 2015-03-12.

[22] Adrià de Gispert and José B Mariño. Using x-grams for speech-to-speech translation. In *INTERSPEECH*, 2002.

[23] Adrià De Gispert and Jose B Mariño. Catalan-English statistical machine translation without parallel corpus: bridging through Spanish. In *Proc. of 5th International Conference on Language Resources and Evaluation (LREC)*, pages 65–68. Citeseer, 2006.

[24] Johannes Dellert. *Lambda calculus on dependency structures for a wide-coverage grammar of Esperanto*. PhD thesis, BA Thesis, University of Tübingen, 2008.

[25] Michael Denkowski and Alon Lavie. Meteor universal: Language specific translation evaluation for any target language. In *Proceedings of the EACL 2014 Workshop on Statistical Machine Translation*, 2014.

[26] Esperantic Studies Foundation. Tekstaro de Esperanto. `http://tekstaro.com/`. Accessed: 27 October 2015.

[27] Ankur Gandhe, Florian Metze, and Ian Lane. Neural network language models for low resource languages. *Proceedings of INTERSPEECH*, 2014.

[28] Michele Gazzola. Managing multilingualism in the European Union: Language policy evaluation for the European Parliament. *Language policy*, 5(4):395–419, 2006.

[29] Martin Gellerstam. Translationese in Swedish novels translated from English. *Translation studies in Scandinavia*, pages 88–95, 1986.

[30] Project Gutenberg. Project Gutenberg. `http://www.gutenberg.org/wiki/Main_Page`. Accessed: 27 October 2015.

[31] Matt Haig. *Brand Failures: the truth about the 100 biggest branding mistakes of all time*. Kogan Page Publishers, 2005.

[32] Don Harlow. The Esperanto Correlatives. `http://donh.best.vwh.net/Esperanto/correlatives.html`. Accessed: 07 December 2015.

[33] William C Hiss and Valerie W Franks. Defining promise: Optional standardized testing policies in american college and university admissions. *Report of the National Association for College Admission Counseling (NACAC). http://www.nacacnet.org/research/research-data/nacac-research/Documents/DefiningPromise.pdf*, 2014.

[34] Qin Hongwu and Wang Kefei. Parallel Corpora-based Analysis of Translationese—The English "so...that" Structure and Its Chinese Equivalents in Focus. *Modern Foreign Languages*, 1:004, 2004.

[35] Intercultural Development Research Association (IDRA). Why is it Important to Maintain the Native Language? `http://www.idra.org/IDRA_Newsletter/January_2000_Bilingual_Education/Why_is_it_Important_to_Maintain_the_Native_Language?/`. Accessed: 07 December 2015.

[36] Iustina Ilisei, Diana Inkpen, Gloria Corpas Pastor, and Ruslan Mitkov. Identification of translationese: A machine learning approach. In *Computational linguistics and intelligent text processing*, pages 503–511. Springer, 2010.

[37] Comprehensive Language Center Inc. Machine Translation (MT). `http://www.comprehensivelc.com/language-services/machine-translation-mt.html`. Accessed: 06 December 2015.

[38] Samuel Johnson. Machine Translation: Babel or babble? `http://www.economist.com/blogs/johnson/2012/06/machine-translation`, 2012. The Economist. Accessed: 08 December 2015.

[39] Nataly Kelly. Why Machines Alone Cannot Solve the World's Translation Problem. `http://www.huffingtonpost.com/nataly-kelly/why-machines-alone-cannot-translation_b_4570018.html`, 2014. Huffington Post. Accessed: 08 December 2015.

[40] Nataly Kelly, Donald DePalma, and Vijayalaxmi Hegde. The Need for Translation in Africa: Addressing Information Inequality so that Africa May Prosper. `http://www.commonsenseadvisory.com/Portals/0/downloads/Africa.pdf`, 2012. Common Sense Advisory. Accessed: 08 December 2015.

[41] Philipp Koehn. Europarl: A parallel corpus for statistical machine translation. In *MT summit*, volume 5, pages 79–86, 2005.

[42] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, et al. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions*, pages 177–180. Association for Computational Linguistics, 2007.

[43] Philipp Koehn, Franz Josef Och, and Daniel Marcu. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 48–54. Association for Computational Linguistics, 2003.

[44] Moshe Koppel and Noam Ordan. Translationese and its dialects. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 1318–1326. Association for Computational Linguistics, 2011.

[45] Jhumpa Lahiri. Teach Yourself Italian. `http://www.newyorker.com/magazine/2015/12/07/teach-yourself-italian?Src=longreads`. Accessed: 07 December 2015.

[46] Angela Lavoipierre. Islamic State: ADF Twitter account used to fight propaganda sending nonsense Arabic tweets, expert says. `http://www.abc.net.au/news/2015-09-23/adf-sending-nonsense-arabic-tweets,-expert-says/6799874`. Accessed: 06 December 2015.

[47] Gennadi Lembersky. *The Effect of Translationese on Statistical Machine Translation*. PhD thesis, University of Haifa, 2013.

[48] Gennadi Lembersky, Noam Ordan, and Shuly Wintner. Improving statistical machine translation by adapting translation models to translationese. *Computational Linguistics*, 39(4):999–1023, 2013.

[49] Alan Reed Libert. What can Pragmaticists Learn from Studying Artificial Languages? In *Perspectives on Linguistic Pragmatics*, pages 397–432. Springer, 2013.

[50] Mitchell P. Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. Building a large annotated corpus of english: The penn treebank. *Comput. Linguist.*, 19(2):313–330, June 1993.

[51] José B Mariño, Rafael E Banchs, Josep M Crego, Adriàa de Gispert, Patrik Lambert, José AR Fonollosa, and Marta R Costa-Jussà. N-gram-based Machine Translation. *Computational Linguistics*, 32(4):527–549, 2006.

[52] Tomas Mikolov. RNNLM Toolkit. `http://rnnlm.org/`. Accessed: 12 December 2015.

[53] Tomas Mikolov. Statistical Language Models Based on Neural Networks. `http://www.fit.vutbr.cz/~imikolov/rnnlm/thesis.pdf`. Accessed: 12 December 2015.

[54] Michael Mulé. Machine Translation: Ensuring Meaningful Access for Limited English Proficient Individuals. `https://www.dol.gov/oasam/programs/crc/062414Machine_TranslationWebinar.pdf`. Accessed: 06 December 2015.

[55] Preslav Nakov and Hwee Tou Ng. Improved statistical machine translation for resource-poor languages using related resource-rich languages. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, EMNLP '09, pages 1358–1367, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[56] Friedrich Neubarth, Barry Haddow, Adolfo Hernández Huerta, and Harald Trost. A hybrid approach to statistical machine translation between standard and dialectal varieties. `http://www.ofai.at/~friedrich.neubarth/papers/ltc2013-neubarth.pdf`. Accessed: 15 December 2015.

[57] University of Oslo. Biblioteca Polyglotta. `https://www2.hf.uio.no/polyglotta/index.php`. Accessed: 28-03-2015.

[58] University of Stanford Unsupervised Feature Learning and Deep Learning. Neural Networks. `http://ufldl.stanford.edu/wiki/index.php/Neural_Networks`. Accessed: 14 December 2015.

[59] Adama Ouane and Christine Glanz. *Why and How Africa Should Invest in African Languages and Multilingual Education: An Evidence-and Practice-Based Policy Advocacy Brief.* ERIC, 2010.

[60] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics, 2002.

[61] Michael Paul, Hirofumi Yamamoto, Eiichiro Sumita, and Satoshi Nakamura. On the Importance of Pivot Language Selection for Statistical Machine Translation. In *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, page 221, 2009.

[62] Barbara Pozzo and Valentina Jacometti. *Multilingualism and the harmonisation of European law*. Kluwer Law International, 2006.

[63] David Rotman. How Technology Is Destroying Jobs. `http://www.technologyreview.com/featuredstory/515926/how-technology-is-destroying-jobs/`, 2013. Accessed: 15 December 2015.

[64] Mark Saba. HUMAN TRANSLATION VS. MACHINE TRANSLATION. `http://www.internationalaffairs.org.au/australian_outlook/human-translation-vs-machine-translation/`. Accessed: 06 December 2015.

[65] Dahlia Sabry and Ibrahim Saleh. The Role Played By Qur'an Translations In Steering Public Opinion Against Islam In Non-Muslim Communities. http://www.islamicwritings.org/quran/language/the-role-played-by-quran-translations-in-steering-public-opinion-against-islam-in-non-muslim-communities/. Accessed: 23 March 2015.

[66] Benjamin Snyder and Regina Barzilay. Climbing the Tower of Babel: Unsupervised Multilingual Learning. *In ICML*, 2010.

[67] Christos Stergiou and Dimitrios Siganos. Neural Networks. `https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html`. Accessed: 13 December 2015.

[68] Straumanis, Kaija. Why Literature in Translation is SUPER SUPER IMPORTANT. `http://www.rochester.edu/College/translation/threepercent/index.php?id=8872`. Accessed: 29 April 2015.

[69] Gideon Toury. In search of a theory of translation, 2006.

[70] Gideon Toury. *Descriptive Translation Studies and beyond: Revised edition*, volume 100. John Benjamins Publishing, 2012.

[71] Naama Twitto-Shmuel. *Improving Statistical Machine Translation by Automatic Identification of Translationese*. PhD thesis, University of Haifa, 2013.

[72] Masao Utiyama and Hitoshi Isahara. A Comparison of Pivot Methods for Phrase-Based Statistical Machine Translation., 2007.

[73] Andreas van Cranenburgh, Galit W Sassoon, and Raquel Fernández. Invented antonyms: Esperanto as a semantic lab. In *Proceedings of IATL*, volume 26, 2010.

[74] Eduardo Javier Ruiz Vieytez. *Pluralidades latentes: minorías religiosas en el País Vasco*. Icaria, 2010.

[75] Emily Wight. What can we learn from efforts to save an ancient South American language? `http://www.theguardian.com/education/2014/nov/18/endangered-language-quechua-columbia-chile`. Accessed: 07 December 2015.

[76] Martin Williams. Tech is removing language barriers—but will jobs be lost in translation? `http://www.theguardian.com/education/2014/sep/19/tech-removing-language-barriers-jobs-lost-translation`, 2014. Accessed: 06 December 2015.

[77] Hua Wu and Haifeng Wang. Pivot language approach for phrase-based statistical machine translation. *Machine Translation*, 21(3):165–181, 2007.

[78] Hua Wu and Haifeng Wang. Revisiting pivot language approach for machine translation. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, pages 154–162. Association for Computational Linguistics, 2009.

[79] Kenji Yamada and Kevin Knight. A Syntax-based Statistical Translation Model. *Proceedings of the 39th Annual Meeting on Association for Computational Lingusitics*, pages 523–530, 2001.

[80] Leyzer Ludwik Zamenhof. Unua Libro. *Warsaw, July*, 26, 1887.