Polyglot Machine Translation

Luis A. Leiva * and Vicent Alabau **
Sciling, CPI UPV, 46022 Valencia (Spain)

Abstract. Machine Translation (MT) requires a large amount of linguistic resources, which leads current MT systems to leaving unknown words untranslated. This can be annoying for end users, as they might not understand at all such untranslated words. However, most language families share a common vocabulary, therefore this knowledge can be leveraged to produce more understandable translations, typically for "assimilation" or gisting use. Based on this observation, we propose a method that constructs polyglot translations tailored to a particular user language. Simply put, an unknown word is translated into a set of languages that relate to the user's language, and the translated word that is closest to the user's language is used as a replacement of the unknown word. Experimental results on language coverage over three language families indicate that our method may improve the usefulness of MT systems. As confirmed by a subsequent human evaluation, polyglot translations look indeed familiar to the users, and are perceived to be easier to read and understand than translations in their related natural languages.

Keywords: Minority Languages; Machine Translation; Linguistic Coverage; Vocabulary; Human factors

1. Introduction and Related Work

In an ideal world, the diversity of languages would not be an obstacle to the transmission of knowledge and culture. In order to enable communication between people separated by language barriers, computers are increasingly being used to automatically convert a *source* language into a *target* language, with machine translation (MT) technology. Maybe computers will never fully replace human translators, but MT is by far more scalable than manual translation for "assimilation" or gisting applications, since MT can automate and considerably speed up this task. Further, for many pairs of languages, even human translators do not exist [3].

However, only 10% of the current languages worldwide are currently covered by MT technologies [13]. The reason for such a low coverage is that MT systems adopt either rule-based or data-driven approaches (or a combination of both) to the translation task, which require fairly large collections of language resources.[1] This means that we can expect MT to work well for the more widely-spoken languages, while for other, less-spoken languages, the chances of successful implementation are more remote... Or can MT systems be adapted to support *any* language?

According to *Ethnologue* [14], around half of the 7,105 living languages worldwide have a developed writing system, all of them being considered minority languages or, from a natural language processing perspective, under-resourced or "noncentral" languages [20]. In theory, MT systems could be deployed for all of them, but in practice the lack of resources available for most of these languages would render any such system largely unusable, since much of the text would be left untranslated. What is more, resources vary greatly even for the 10% most popular languages; and, given their enormous rate of growth and state of continuous evolution [16], even the best-equipped languages cannot be covered in their entirety by MT systems.

At best, poor language coverage leads to what is known as the out-of-vocabulary (OOV) words problem. Current MT systems usually respond to this occurrence by leaving unknown words untranslated. This is rather problematic for two main reasons: firstly, untranslated words may be of paramount importance to the underlying meaning of a sentence or even a paragraph, so the message can be lost; secondly, when the source language is unrelated to the user's primary (reading) language, these untranslated words are often completely undecipherable. Consequently, in the extreme case of there being no resources available for a given source language, MT systems simply cannot be built and the automatic translation of these languages becomes a near impossible task.

*Both are corresponding authors (name@sciling.com).
**Work conducted while both authors were affiliated with the Universitat Politècnica de València.

To overcome resource scarcity and data sparseness, it is sometimes possible to use a better-equipped language as a *pivot* language [23], where the source language is translated to the pivot language, and then from the pivot language to the target language. Even multiple pivot languages can be used to derive translation hypotheses and later reach consensus between them [5,12]. Another option is transliteration (at the character level) to a target or pivot language that is similar enough to the source language [25], since transliteration is a rather small step toward delivering an intelligible text. Other approaches involve rephrasing the source text, searching for synonyms and paraphrases [17], aiming to find source sentences that the MT system can successfully translate. Other authors propose using a subset of the rules to generate phrase candidates [19,24]. In other cases, a profound knowledge of the source language and specific language tools might be required [6]. Unfortunately, these and other approaches to the resource scarcity problem developed along similar lines require explicit prior knowledge of the target and source languages in question. Their translation, therefore, remains problematic. Thus, making MT viable for any language worldwide would be quite a feat for MT technology. Our work is an early attempt to achieve this goal.

2. A Tale of Many Languages

We propose a novel approach that shows potential for overcoming the resource scarcity problem in MT: namely, the generation of translations using a combination of languages from the same family as the target language. After all, language contact is a fact of life. All languages are "mixed", to a greater or lesser extent influenced by other languages [28]. Similarly, many languages share a considerable amount of their vocabulary, whether down to geographical proximity [8] or cultural influence [10]. Good examples of this phenomenon of mixed languages can be found in multilingual countries, whose speakers frequently incorporate foreign words into their conversations. Aside from issues of word frequency and daily use, it has been suggested that they do so to compensate for a lack of language proficiency and, by using these foreign words, improve their chances of being understood [7].

Meanwhile, many artificial languages have been developed over the years with a view to facilitating human communication and overcoming traditional language barriers; Esperanto,[2] Ido,[3] or Interlingua,[4] for example. Nevertheless, despite their proven usefulness, artificial languages must still be learned. So, the question is, would it be possible to develop an MT system that is able to mimic this "mixedness" and leverage other languages to overcome gaps in languages resources? Would this system, by doing so, respond more helpfully to the untranslated word situation? Could this natural capacity for "mixedness" be exploited in MT to allow users to essentially understand a text written in a completely foreign language like they understand text written on their own language?

With these questions in mind, we have developed a statistical translation model aimed at tackling the resource scarcity problem head on and improving the usability of machine translations in resource-poor languages. Furthermore, the model allows language resources to be amassed over time until reaching levels whereby regular MT systems can be successfully adopted (Section 4).

On the other hand, contact languages have inspired some pioneering works where MT was envisaged to act as some kind of pidgin,[5] where the translation is made, not into a full language, but into a much more primitive though still comprehensible language, following a "word-for-word" procedure [15]. In this regard, the output of a polyglot MT system like the one we are proposing could be considered some kind of contact language, though our model allows for the production of more intelligible translations, as indicated below.

The crux of our method lies in leveraging translations available in languages related to the target language and replacing untranslated words (or groups of words, also known as *phrases*) by word candidates that are closest to the target language. The likelihood of these candidates is estimated by a normalized edit distance and a lexicon that can be obtained from as little as a list of words in the target language, which we consider the minimal amount of resources to define a language. Note that the automated system requires a reasonable level of knowledge regarding the languages related to the target language, for which alignment information can be derived, but only minimal language resources of the target language itself. The end users, meanwhile, do not need to be proficient in any of the related languages, since the words borrowed from these languages are selected on the basis of their similarity to the target language. Of course, the more the language resources the system has access to (e.g. bilingual dictionaries), the better the outcome. However, this is not a realistic assumption in the case of less-commonly spoken languages.

Tables 1 and 2 illustrate the translations that our polyglot MT model would produce when translating the sentence *"The game tells you a region and you must*

guess their capital." from Swedish to Spanish under two different scenarios. The model will be formulated and thoroughly described later, in Section 4.

Table 1

Worst-case scenario. A set of related languages are used to build SV→ES translations where no prior knowledge is available for a given target language; in this case, Spanish. Language codes are: SV: Swedish, ES: Spanish, PT: Portuguese, IT: Italian, FR: French.

SV	Spelet talar om en landsdel för dig och du måste gissa dess huvudstad.
ES	N/A
PT	O jogo indica-lhe uma divisão e você terá de adivinhar a sua capital.
IT	Il gioco ti dice una divisione e tu devi indovinare la sua capitale.
FR	Le jeu vous donne une division et vous devez deviner sa capitale.
	Le jogo ti dice una division e tu devez adivinhar la capital.

Table 2

Better-case scenario. A set of related languages are used to build better SV→ES translations where some words (in italic) cannot be translated into a target language; in this case, Spanish. Language codes are the same as in Table 1.

SV	Spelet talar om en landsdel för dig och du måste gissa dess huvudstad.
ES	El *Spelet* le muestra una división y *måste gissa* su *huvudstad.*
PT	O jogo indica-lhe uma divisão e você terá de adivinhar a sua capital.
IT	Il gioco ti dice una divisione e tu devi indovinare la sua capitale.
FR	Le jeu vous donne une division et vous devez deviner sa capitale.
	El *jogo* le muestra una división y *tu devez adivinhar* su *capital.*

3. Analyzing Language Coverage

Our central hypothesis is that by incorporating knowledge derived from a family of related languages, we can increase coverage of a language for which little to no language resources are available. To that end, we studied the extent to which translations into a given target language can be supplemented by data available for other languages related to it. Specifically, we analyzed the morphological similarities of the target language vocabulary with the vocabularies of its related languages.

3.1. Materials

To test our hypothesis, we used word lists that are publicly available,[6] in their turn compiled from the OpenSubtitles dataset,[7] containing vocabulary in 40 languages and sorted by frequency. From these languages, we selected three language families for which data were available (Table 3): the Western Romance language family (Spanish, Portuguese, French, and Italian); the Scandinavian language family (Icelandic, Norwegian, Danish, and Swedish); and the Slavic language family that uses the Cyrillic alphabet (Macedonian, Bulgarian, Serbian, Russian, and Ukrainian).

For each language family, we supplemented the language coverage of the first language using vocabulary from the other languages in its family. For example, for the Scandinavian language family, we took Icelandic as the target language and supplemented its coverage with vocabulary available for Norwegian (Bokmål), Danish, and Swedish.

We should point out that, only for evaluation purposes, Spanish was considered to be an under-resourced language. Actually, it is not at all less resourced than Portuguese or Italian as regards MT. However, Spanish is the authors' primary language and thus we could better interpret the results.

3.2. Procedure

We carried out two different experiments on each language family, aimed at exploring the worst- and better-case scenarios illustrated in Table 1 and Table 2. In the first experiment (worst-case scenario), we simulated the challenge of translating into a target language for which no prior knowledge is available (no MT system can be built, all words are left untranslated) and very few language resources are available to generate polyglot translations. To do so, we used a total vocabulary of 5,000 words from each of the related languages and analyzed coverage of the 5,000 most frequent words (+5k vocabulary) in the target language, which is considered to be a good estimate to cover a language [1].

In the second experiment (better-case scenario), we studied the case of translating into a language for which some prior knowledge is available (an MT system can be built, some words are left untranslated) and a reasonable amount of language resources are available to generate polyglot translations. Here, we used a total vocabulary of 50,000 words from each of the related languages and analyzed coverage of the 5,000 least frequent words (-5k vocabulary) in the target language, since the least frequent words are the ones that a MT system would leave untranslated. To avoid noisy data, we selected for analysis only those words that appeared at least 5 times in their respective word lists.

Both experiments illustrate ways in which the performance of an MT system can be improved: results from the worst-case scenario suggest that a workable MT system could be built without needing any prior training data with regard to the target language, while the second experiment (better-case) shows how the problem

Table 3

Word counts for each language family. Language codes are: ES: Spanish, PT: Portuguese, FR: French, IT: Italian; IS: Icelandic, NO: Norwegian, DA: Danish, SV: Swedish; MK: Macedonian, BG: Bulgarian, SR: Serbian, RU: Russian, UK: Ukrainian.

Family	**Western Romance**				**Scandinavian**				**Slavic Cyrillic**				
Language	ES	PT	FR	IT	IS	NO	DA	SV	MK	BG	SR	RU	UK
Words	106M	61M	58M	34M	3.2M	12M	27M	29M	5M	53M	48M	18M	591K
Unique	583K	392K	350K	366K	142K	248K	336K	377K	146K	509K	751K	450K	65K

of untranslated words can be successfully tackled using related-language vocabulary in a working MT system.

3.3. Method

To compute the similarity of the target vocabulary with that of its related languages, including with a mixed language that draws on all of said languages, we proceeded as follows: for each word w_t in the target vocabulary, we search for the most orthographically similar word w_r in the related-language vocabulary V_r, using the following decision rule (a *normalized* edit distance) with respect to V_r:

$$d_n(w_t, V_r) = \max_{w_r \in V_r} \left[1 - \frac{d(w_t, w_r)}{\max(|w_t|, |w_r|)} \right] \quad (1)$$

where $d(w_t, w_r)$ is the edit distance between a word in the target language w_t and a word in the related language w_r, and $|\cdot|$ denotes the length of each word. A value of $d_n = 1$ means that the target word can be found in a related-language vocabulary, i.e., similarity is maximum. On the other hand, a value of $d_n = 0.5$ would mean that the most similar word in the related language can be turned into the target word by changing 50% of its characters.

3.4. Results

Figure 1 provides an overview of language coverage for each of the three language families analyzed so far, including the coverage provided by each family's mixed language. As previously stated, the closer the similarity to 1, the better the language coverage. The band across each box shown in the y-axis represents the median and indicates the similarity with which half of the words in the target vocabulary can be covered by its related vocabularies. For example, if we take the top left boxplot as an example, we can see that half of the selected +5k Spanish vocabulary can be covered by French words with a similarity of 0.66 and above. This means that, by changing a maximum of 33% of the characters in French words, we can cover fully half of the +5k Spanish vocabulary. Figure 3 shows a graphical example of the results for this scenario, showing the relative coverage of the related languages.

In addition, we explored language coverage as a function of vocabulary similarity in the interval $[0, 1]$. This was performed for the three language families analyzed, together with the contributions of the mixed language to each language family. As observed in Figure 2, for some language families a small relaxation in the similarity threshold may lead to a vast increment in language coverage. For example, for the Slavic Cyrillic family and -5k vocabulary, using $d_n = 1$ coverage is 50% whereas for $d_n = 0.8$ coverage increases to 85%.

Some interesting observations can be made based on the results shown in these figures. First and foremost, the contribution of each natural language varies from language to language, and it is the mixed language in all cases that best supplements coverage of the target language. This is so by design, since the mixed language vocabulary contains the vocabulary of each related language. Meanwhile, the relationship between linguistic similarity and geographical proximity [8] is self-evident, with mutual intelligibility increasing with geographical proximity. For example, Iceland is physically isolated from the other countries in the Scandinavian language family (Denmark, Norway, and Sweden), so we can expect Icelandic to be quite different to the other languages in its family. This particular observation has already been made in the literature [2] and is empirically confirmed in both central boxplots of Figure 1. Indeed, even in the better-case scenario, no language in the Scandinavian family can totally supplement Icelandic coverage ($d_n < 1$). At most, the mixed language can successfully account for 75% of Icelandic vocabulary using a similarity of $d_n = 0.8$ (third quartile of the boxplots, 75th percentile). It is only with $d_n = 0.5$ when this becomes possible (Figure 2), however notice that $d_n \leq 0.5$ implies changing at a minimum 50% of the characters of a related word to match an Icelandic word. On the contrary, coverage of Spanish and Macedonian can be totally supplemented by some of the other languages in their respective families for

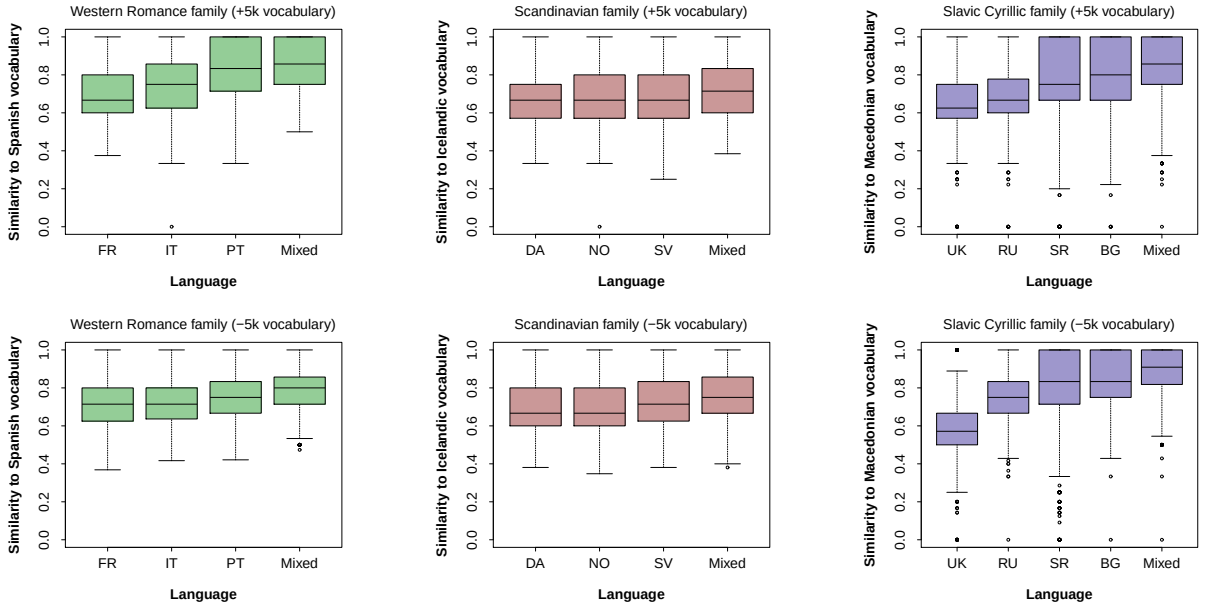

Fig. 1. Language coverage in terms of vocabulary similarity for the three language families analyzed. Language codes are: FR: French, PT: Portuguese, IT: Italian, DA: Danish, NO: Norwegian, SV: Swedish, UK: Ukrainian, RU: Russian, SR: Serbian, BG: Bulgarian.

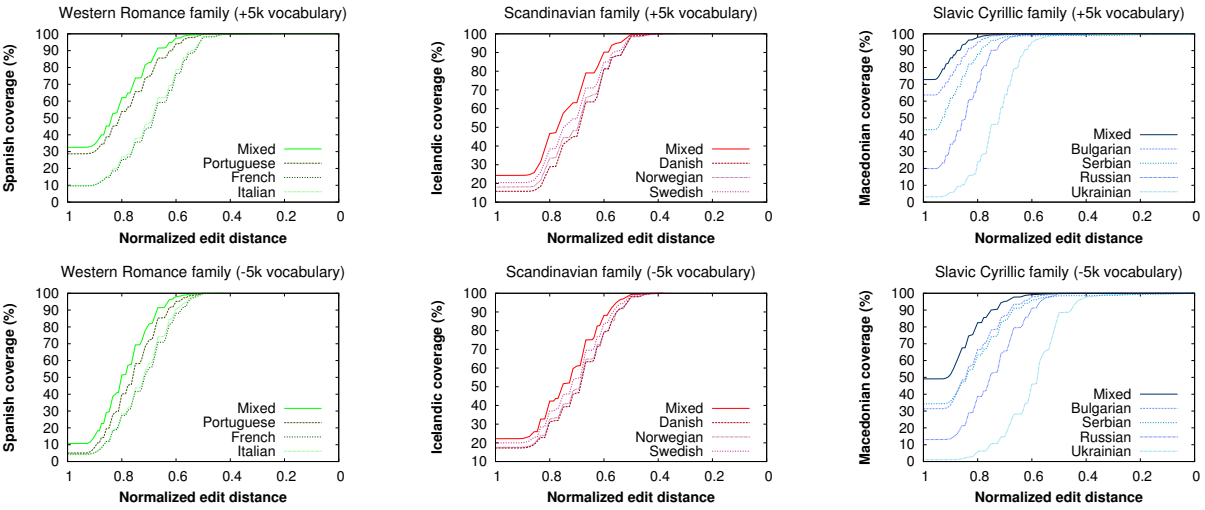

Fig. 2. Language coverage according to vocabulary similarity in the interval $[0, 1]$, where 1 is a perfect match to the target language, i.e., vocabulary similarity is maximum.

the +5k vocabulary. This was also observed for both -5k and +5k vocabularies in the Slavic Cyrillic family.

Another example of the relationship between linguistic similarity and geographic proximity can be found in the other two languages families. Specifically, in the Western Romance language family, Portuguese is far closer to Spanish than either French or Italian, yet both Portugal and France share a border with Spain. However, between France and Spain lies the natural barrier of the Pyrenees mountain range which, historically, has prevented the kinds of migratory fluxes that are more common between Portugal and Spain [26]. Finally, in the Slavic language family, Macedonian shares much more of its lexicon with Bulgarian and Serbian than with Ukrainian or Russian, which is somehow expected

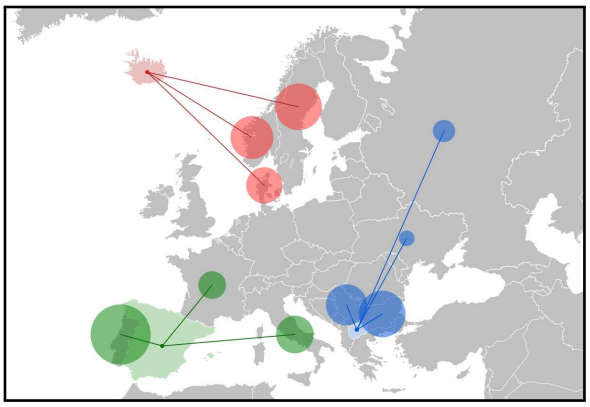

Fig. 3. Language coverage when supplemented by a set of related languages for each of the three language families analyzed. Circles represent the relative coverage of each related language.

since Bulgaria and Serbia are geographically very close, while Ukraine and Russia are much further away.

In light of these observations, we wondered to what extent each language, whether natural or mixed, would significantly better supplement coverage of the target language from a statistical point of view. To do so, we performed a one-way between-groups analysis of variance (ANOVA) test. Differences between language coverages were found to be statistically significant in all cases ($p < .0001$), meaning that there was always at least one language that performed significantly better than the rest in each family. Effect sizes suggest moderate practical significance ($0.25 \leq \eta^2 \leq 0.33$). Tukey's HSD post-hoc tests revealed that the mixed language gave significantly higher language coverage in comparison to any of the natural languages. All other comparisons were not significant.

4. A Model for Polyglot MT

The previous experiments suggest that, instead of leaving unknown words untranslated, MT systems should leverage language similarities between a target language and its related languages to provide a polyglot translation that is tailored to the user's own language. Doing so would significantly improve language coverage if the user were able to identify foreign words that are similar to her primary reading language. Based on this observation, we developed a model for polyglot MT, which we outline below.

From a statistical point of view, the best translation of a source sentence $\boldsymbol{s}$ into a target language can be computed using the fundamental MT equation:

$$\boldsymbol{y}^* = \arg\max_{\boldsymbol{y}} \Pr(\boldsymbol{y}|\boldsymbol{s}) \tag{2}$$

where $\Pr(\boldsymbol{y}|\boldsymbol{s})$ is the conditional probability that the target string $\boldsymbol{y}$ is the translation of the source string $\boldsymbol{s}$.

This equation can be considered a state-of-the-art statistical MT model. Let the target language be the user's primary reading language, usually denoted as $\mathcal{L}_1$. When $\mathcal{L}_1$ resources are scarce, typically $\boldsymbol{y}^*$ would contain untranslated words that are often indecipherable to the end user, especially when the source language is from a different language family. As already discussed, we can improve the understandability of $\boldsymbol{y}^*$ by leveraging a number N of translations $\boldsymbol{t}_n$ from languages related to $\mathcal{L}_1$ and a set of language resources $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_N\}$. So, following some mathematical transformations (detailed in the appendix), we obtain the following expression:

$$\boldsymbol{y}^* \approx \arg\max_{\boldsymbol{y}} \max_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_N} \Pr(\boldsymbol{y}|\boldsymbol{s},\boldsymbol{\theta},\boldsymbol{t}_1,\ldots,\boldsymbol{t}_N) \prod_{n=1}^{N} \Pr(\boldsymbol{t}_n|\boldsymbol{s},\theta_n) \tag{3}$$

where the first term selects the words closest to $\mathcal{L}_1$ from each word in the possible translations by using some similarity measure that can leverage the knowledge available at $\boldsymbol{\theta}$; and the second term is a pool of translations of the source string $\boldsymbol{s}$ into each language. The simplest resource that $\boldsymbol{\theta}$ can hold is a $\mathcal{L}_1$ vocabulary that can be obtained from as little as a list of words in the target language, as in our experiments, but there are many other ways to achieve this outcome; e.g., using glossaries or monolingual dictionaries. In the next subsections we outline a series of scenarios where our model can be further exploited to generate useful data for MT systems. Specifically, we propose a process in which polyglot translations are used to incrementally build up a set of language resources until reaching a state-of-the-art MT system. At this time, only Section 4.1 has been formally assessed. Nevertheless, the whole process summarizes our vision of how this work can be used in practice.

4.1. No Prior Knowledge about $\boldsymbol{t}_1$

In the worst-case scenario, we explored the application of our model where $\boldsymbol{\theta} = \theta_1 = V_1$ is the target

language vocabulary and the only language resources available for the source languages θ_n are the corpora used to build the translations $\boldsymbol{t}_n$. Indeed, this would be the worst-case scenario for an MT system, since V_1 can be regarded as the minimum amount of information required in order for the system to function [13]. It is here that we intervene, leveraging any available MT system to solve Eq. (3) in two steps:

1. We obtain the 1-best translation for each related language $1 < n \leq N$ using the MT systems available:

$$\boldsymbol{t}_n^* = \arg\max_{\boldsymbol{t}_n} \Pr(\boldsymbol{t}_n|\boldsymbol{s}) \tag{4}$$

2. Next, we mix these translations, tailoring the mix to $\mathcal{L}_1$. This mix is assumed to be independent from $\boldsymbol{s}$ and $\{\theta_2, \ldots, \theta_N\}$, since it depends on $\{\boldsymbol{t}_2^*, \ldots, \boldsymbol{t}_N^*\}$, which means that:

$$\boldsymbol{y}^* = \arg\max_{\boldsymbol{y}} \Pr(\boldsymbol{y}|V_1, \boldsymbol{t}_2^*, \ldots, \boldsymbol{t}_N^*) \tag{5}$$

where (5) is approximated by selecting each word $w \in \bigcup_{n=2}^{N} \boldsymbol{t}_n^*$ from the pool of translation candidates where $d_n(w, V_1)$ is maximum. More specifically, we first obtain the statistical alignments that result as a sub-product from (4). These alignments link the words in $\boldsymbol{t}_n^*$ to the words in $\boldsymbol{s}$, and hence, we can trace the alignments back to the other languages so that we can group them by phrases that cover the source spans. This creates a set of comparable phrases that joined sequentially compose an automaton. Then, each word of the automaton is associated with the word similarity score indicated in Eq. (1). Finally, (5) is approximated by a Viterbi-like traversal algorithm on the automaton, where each phrase score is normalized by the number of words involved. Note that Eq. (1) does not need to be converted into a probability since it would not change the maximum argument in Eq. (5).

Figure 4 provides a graphical example of (a fragment of) the sentence shown in Table 2, where unknown words and phrases in the MT output are replaced by occurrences in the languages that relate to the target language.

We should note that the resulting polyglot translation is unlikely to be syntactically correct according to the grammatical norms of the target language, either in terms of morphology or syntax. However, what it will be is understandable to users of the target language, since the particular mix of languages drawn upon is tailored to the target language in question—at the very least, the words will look familiar to them, and thus, they will likely be able to ease comprehension.

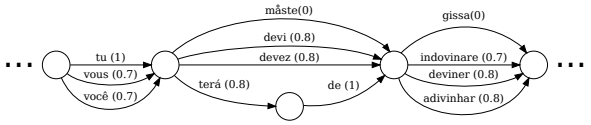

Fig. 4. Unknown words are assigned 0 probability (depicted in parentheses) of belonging to the target vocabulary (in this case, Spanish). Each node in the graph is transited according to the maximum probability of belonging to such a target vocabulary, depicted by Eq. (1). Branches are collapsed either when a word is common to two (or more) of the related languages, or when no compatible alignments between the related languages are found. In case of probability ties, alignment candidates are chosen at random.

4.2. Using Translation Dictionaries

Now consider that, in addition to the aforementioned vocabulary V_1, we have a simple bilingual dictionary D_1, meaning that $\theta_1 = \{D_1, V_1\}$. Then $\Pr(\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{\theta}, \boldsymbol{t}_1, \ldots, \boldsymbol{t}_N)$ in Eq. (3) can be limited to measuring the similarity d_n to the possible translations of each word in the source sentence. We would expect that the resulting translations be easier to understand by $\mathcal{L}_1$ users, since all text segments would include words coming exclusively from the target vocabulary.

Additionally, since in this scenario users are shown translations taken from a bilingual dictionary, it would be possible to perform transliteration at the character level: D_1 can be used to relate V_1 to V_n, so that words from V_n can be mapped to words in $\mathcal{L}_1$. This would generate parallel data that could be used to train an automatic transliteration system [18,25]. However, in this scenario, it would be preferable to present the users with a word taken from a related language that is known to be a correct translation of the source text and that will likely look familiar enough for the users to be able to infer its meaning. Besides, given that we assume that users have passing knowledge in some of the related languages, they would implicitly know some transliteration rules. Therefore, we believe it is more reliable to trust the user's knowledge rather than taking the risk of presenting them with a broken transliteration.

4.3. Allowing User-Defined Translation Rules

Now consider the scenario where users are allowed to post-edit our polyglot translations, and not just at the sentence level, but also sentence parts. This would result in a set of structural transfer rules formulated through the combined contributions of whole groups of users, which could then be fed into the MT system

as a valuable language resource. It would be particularly useful for many software localization tasks, where text is repeated over and again in, e.g., buttons, drop-down menus, technical manuals, short legal texts in disclaimers and certifications, etc. Furthermore, these user-generated resources would enable MT systems to be applied in other translation domains since, having undergone partial supervision, they would be suitable for use as ground truth data. In fact, this serves as a basis for the so-called online learning paradigm, where the MT system can build a translation model incrementally from scratch.

4.4. Filling Translation Gaps

At this point we have reached the current state-of-the-art in MT systems, for which parallel data are available for building usable translator workbenches. However, even in this scenario, the system would not be completely error-free, since untranslated words from the source language would still appear and, as discussed, would be left "as is" in the target text. Polyglot translations are still useful in these cases (c.f. Table 2), since a familiar-looking word from a related language should help the user to recognize the actual meaning of such word, basically by looking at its context in the sentence. We elaborate more on this scenario in Section 6.

5. Are Polyglot Translations Understandable?

Following on from the previous experiments (Section 3) and the proposed polyglot translation model (Section 4), we conducted a formal user evaluation over the Western Romance language family. Concretely, we tested the model under the scenario of a complete lack of data from $\mathcal{L}_1$. By doing so, we were able to extrapolate the results through to the better-case scenario, where state-of-the-art MT systems that already have enough resources for $\mathcal{L}_1$ would be enhanced by additional language coverage. In this section we re-analyze the data we gathered in previous work [13], aimed at providing more insights about the user evaluation.

Because Spanish is the authors' primary language, *only* for evaluation purposes we assumed that Spanish is an under-resourced language influenced by its neighboring countries: Portugal, Italy, and France. This way, we would have the necessary materials to perform the study: 1) Spanish belongs to the family of Western Romance languages; 2) we have publicly available parallel ground truth data for all of these languages; 3) we can easily recruit a representative user sample of native Spanish users; and 4) interpreting the results is effortless for us.

We recruited via email advertising and word-of-mouth communication 17 Spanish-only participants (11 male; 6 female) in their thirties. A requisite for taking part in the study was that participants should not have advanced knowledge in any of the 3 related Western Romance languages: Italian, French, and Portuguese. To verify this requisite, participants were told to score their general knowledge for these related languages. These results are shown in Table 4. All median scores are ≤ 2, which reveals that participants had actually little knowledge of these languages.

Table 4
General foreign languages knowledge as scored by our participants in a 1–5 scale, higher is better.

Language	Median	Mean	SD
Italian	2	1.9	0.6
French	2	2.2	1.0
Portuguese	1	1.7	0.8

The source language of the test sentences was Swedish, so that participants would not be able to understand the original sentences and had to rely on some form of MT, either polyglot or legitimate translations. As per the worst-case scenario described in previous sections, polyglot translations were produced by an MT system that had no prior knowledge of Spanish and used only data taken from closely related languages from its language family. The polyglot MT system was built with the ground truth translations from Swedish into each of the related languages.

5.1. Experimental Design

We formulated the following research hypotheses:

1. Polyglot translations look familiar to the user.
2. Unfamiliar words are very dissimilar from the user's (target) vocabulary.
3. Polyglot translations are more understandable than translations in the related languages.

To evaluate our first hypothesis, we tested if there were differences among all languages in terms of proportion of unknown words per sentence. To do so, we used a one-way ANOVA test and an α level of .05 to assess statistical significance. Participants were not told which was the language of the translations shown at any time.

Regarding our second hypothesis, we performed a correlation analysis of the words that were marked as unknown by each user and their similarity against the Spanish vocabulary. This would test whether words marked as unknown by the user are likely to have low similarity according to Eq. (1).

Our third hypothesis was evaluated on the basis of the following criteria:

1. *Fluency*: Is the polyglot translation readable?
2. *Comprehension*: Is the polyglot translation understandable?
3. *Adequacy*: Regarding the reference sentence (in Spanish), does the polyglot translation preserve meaning?

We carried out a two-step procedure to validate this hypothesis, both steps being evaluated on the same screen (Figure 6). In the first step we analyzed the polyglot translations in terms of the above mentioned criteria. In the second step we verified if any of the related languages would work better than the polyglot translations. Again, participants were not told which was the language of the translations shown.

It is important to note that, in the context of this study, it is difficult to apply classical evaluation tests to measure the quality and understandability of polyglot MT output, as translations are in a mixed language. For example, cloze tests [21] or gap-filling methods [19] have little application here. Moreover, classical reading comprehension tests and tests specifically tailored to measure language proficiency of MT such as the Interagency Language Roundtable [9] are questionnaires with multiple questions and multiple-item responses, which are best suited to assess full paragraphs and multi-line texts. Since we were interested in measuring single-line polyglot translations, we used single-question questionnaires that were answered in a simple 1–5 scale.

5.2. Materials

We used the KDE4 corpus, which comprises the localization files of popular software libraries, and is publicly available at the OPUS project [22]. This corpus has parallel text (source and translations) for 92 languages and 8.89 million sentence fragments, including their alignments. Therefore, we did not have to build a dedicated MT system for each of the related languages, as we already had the necessary ground truth data. Only the polyglot MT system had to be built, as described next.

We trained a polyglot MT system using Moses [11] with 167,000 sentences of each related language (2.2 million running words), and reserved 100 Swedish sentences for testing. All sentences were randomly selected. The target vocabulary used to feed our polyglot MT system was the `/usr/share/dict/spanish` file, which is simply a newline-delimited list of 86K Spanish words, and is available in all Unix systems.

A quick first look at the polyglot translations revealed that the contribution of each language, as assigned by our model in terms of vocabulary rates, was 25.2% Italian words, 12.6% French, 25.3% Portuguese, and 36.9% common.

The sentences reserved for testing were also available for each of the related languages, and two test partitions were selected for human evaluation. The first partition included 5 sentences, whereas the second one had 10 sentences.

5.3. Procedure

For the first study, the polyglot translations in the first test partition were shuffled together with their corresponding translations in Italian, French, and Portuguese. Each participant had to assess 5 sentences from each language, 20 translations in total, which were presented in random order. For each sentence shown, participants had to click on those words that were completely unknown to them. Participants did not know which was the language of the sentences shown at any time. A "Next" button allowed participants to load the next translation (Figure 5), which could be in Italian, French, Portuguese, or Polyglot.

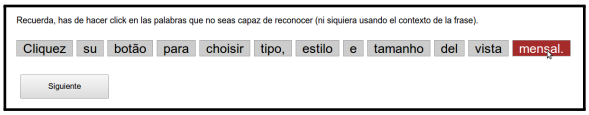

Fig. 5. Screenshot of the setup for Study 1. The indication given (in Spanish) is the following: *"Remember, you have to click on those words that you cannot understand at all (even using the sentence context)."*

The second study was performed using the data derived from the first study. For the third study, participants were sequentially presented with 10 Swedish-Polyglot translations, though participants actually were not told whether a translation was polyglot or legitimate. Each translation had to be assessed in a 1–5 Likert scale according to fluency, comprehension, and adequacy criteria. Then, participants had to rank all translations, including those in the related languages. Only to assess the adequacy criterion, participants were given the ground truth Spanish translations (i.e., the reference

translations of the Swedish sentences) as shown in Figure 6. Eventually we collected 17 users × 5 sentences × 4 languages = 340 samples for the first (and second) study, and 17 users × 10 sentences × 3 criteria = 510 samples for the third study.

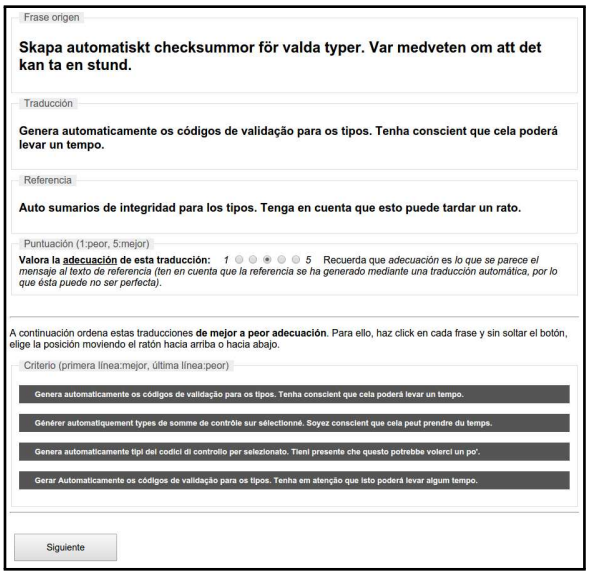

Fig. 6. Screenshot of the setup for Study 3. The indications given (in Spanish) are the following: *"Score the {fluency, comprehension, adequacy} of this translation (higher is better)"* and *"Now sort the following translations from higher to lower according to {fluency, comprehension, adequacy} (first result is higher). To do so, drag and drop each translation with your computer mouse."* The reference translation was only shown to the participant for completing the adequacy test, as in this screenshot.

Finally, participants filled out a questionnaire that measured their subjective appreciation toward the overall quality of polyglot translations. We decided to adapt (and translate into Spanish) the well-known System Usability Scale (SUS) questionnaire [4], since the texts we used in the user study come from localization files of user interfaces, and so it was considered a good strategy to collect user feedback. Participants could also complement the questionnaire with free-form comments and ideas.

5.4. Results

Regarding our first hypothesis, as anticipated, participants were able to recognize most of the words in polyglot translations. Specifically, users marked only a few unfamiliar words per sentence, and this was so also in the related languages (Table 5). Differences in percent rates were found to be statistically signif-

I feel I'm qualified to understand polyglot translations.
I found polyglot translations unnecessarily complex.
I thought polyglot translations were easy to understand.
I think that I would need support to learn polyglot translations.
I found polyglot translations to be a valuable resource.
I thought polyglot translations were too much inconsistent.
I would imagine that most people would learn polyglot translations quickly.
I found polyglot translations very cumbersome to understand.
I felt very confident reading polyglot translations.
I needed to learn languages before I could use polyglot translations.

Fig. 7. Adapted SUS questionnaire to assess polyglot translations. Each question was scored in a 1–5 Likert scale (1: strongly disagree, 5: strongly agree).

icant $[F(3, 327) = 11.7, p < .0001, \eta^2 = 0.11]$. Post-hoc comparisons using the Tukey HSD test indicated that the proportion of unknown words in polyglot translations was significantly lower in comparison to Italian ($p < .005$, Cohen's $d = -0.50$) and French ($p < .006, d = -0.16$). However, for Portuguese there were no statistically significant differences ($p = .056, d = 0.43$). These results indicate that participants were confident while reading sentences both in Polyglot and Portuguese. Considering that no participant was proficient in Portuguese, this may be explained in part because both Portuguese and polyglot translations were shorter overall than translations in French or Italian, as depicted in Table 5. This may also be explained because of the fact that Spanish and Portuguese have strong mutual influences, mostly due to geographic and cultural proximity (c.f. Figure 3). In fact this can be observed in the +5k vocabulary experiment (Figure 1), where common (most frequent) Portuguese vocabulary typically match up with common Spanish vocabulary much more than either French or Italian.

Table 5
Unknown words rate (word counts normalized by sentence length), lower is better.

	Unk. words (%)			Sentence length		
Language	**Median**	**Mean**	**SD**	**Median**	**Mean**	**SD**
Italian	12.9	14.9	9.3	15	15.7	2.2
French	11.5	11.8	9.7	15	17.4	4.7
Portuguese	7.7	6.8	7.3	13	15.1	2.8
Polyglot	8.3	10.3	8.7	12	12.8	1.1

Regarding our second hypothesis, a Pearson's correlation test between the words marked as unknown and their similarity against the Spanish vocabulary reported a statistically significant result $[\rho = -0.27, t(191) = -3.93, p < .001]$. A negative correlation means that

there is an inverse relationship between the number of unknown words marked by the user and their similarity to the user's vocabulary. Equivalently, this indicates that users' perceived word familiarity is related in a positive linear sense to that of measured by Eq. (1). This result validated our second hypothesis, which was somehow expected. What we did not expect, however, was that participants preferred the mixed language over the natural languages most of the time, as shown in Table 6. This can be explained by the fact that the KDE4 corpus was automatically generated by an MT system following partial post-editing; therefore some of the ground truth sentences are not error-free. Furthermore, according to the comments made by our participants, we have noticed that, besides the effort that must be invested to understand the message of a polyglot translation, users tend to tolerate less errors when reading a (presumably) legitimate translation of an official document or interface.

Table 6

Percentage of times a language was chosen in nth place by the participants. The best result is displayed in bold typeface.

Language	1st place	2nd	3rd	4th
Portuguese	28.5	47.6	0	23.8
French	5.9	5.9	44.6	43.4
Italian	0	17.8	49.4	32.7
Polyglot	**65.4**	28.5	5.9	0

Regarding our third hypothesis, we found fairly consistent results in terms of self-assessment scores of fluency, comprehension, and adequacy (Table 7). These validated our hypothesis, although, besides of these high scores, we did not find strong correlations ($0.1 < \rho < 0.3$). We suspect it is because such self-assessment scores may only be approximate indicators.

Table 7

User's self-assessment scores of polyglot translations in a 1–5 scale, higher is better.

Criterion	Median	Mean	SD
Fluency	4	4.1	0.7
Comprehension	4	4.2	0.1
Adequacy	5	4.2	0.9

Regarding the adapted SUS questionnaire, the average score was 67.65 (*SD*=12.1). Given that SUS scores are ranged between 0 and 100 (the higher the better), this result suggests that participants were satisfied with polyglot translations. We then inspected each SUS question individually and observed that, overall, they were scored as expected. For example, *"polyglot translations are easy to understand"* and *"polyglot translations are a valuable resource"* were notably ranked as positive; and, conversely, *"polyglot translations are unnecessarily complex'* or *"polyglot translations are very cumbersome to understand"* were ranked as negative.

On the whole, participants appreciated the polyglot approach and found these translations to be a valuable aid for "assimilation" or gisting use of MT systems. The viewpoint that participants agreed most was that *"the mixed translations aims to improve understanding"* and that *"polyglot sentences are both interesting and useful"*. One enthusiastic user stated that *"the automatically generated language has great possibilities, for example to complement or enhance those machine translation systems having many errors"*. One skeptical user reported that *"It surprised me! Polyglot translations were really helpful to convey meaning"*. Interestingly, some users remarked that *"polyglot translations were really easy to deal with, sometimes ever better than the reference sentences"*.

Finally, although a few people did not find polyglot translations very appealing (SUS < 50), Figure 8 shows that it worked quite successfully for them (criterion scores ≥ 4). This can be noticed in the first quadrant of the figure which clusters users that liked polyglot translations and indeed it worked for them. The opposite situation is summarized in the third quadrant of that figure, which clusters users that did not like polyglot translations and it did not work for them; actually no user fell in this quadrant. To conclude, not only are polyglot translations understandable and can, therefore, be usefully deployed in the absence of prior language knowledge, but their incorporation into existing MT systems can only enhance MT output.

6. Limitations and Future Work

First of all, a polyglot MT system requires translations available in languages related to the target language. This may not work for language isolates (language families with only one language) such as Albanian or Greek. Though most of the world's languages are known to be related to others [28], and so our method is expected to work for them.

Another limitation worth commenting is that of the so-called *false friends*, i.e., those words or phrases in two languages that look similar but differ significantly in meaning. For example, English 'embarrassed' is translated into Portuguese 'embaraçado', Italian 'im-

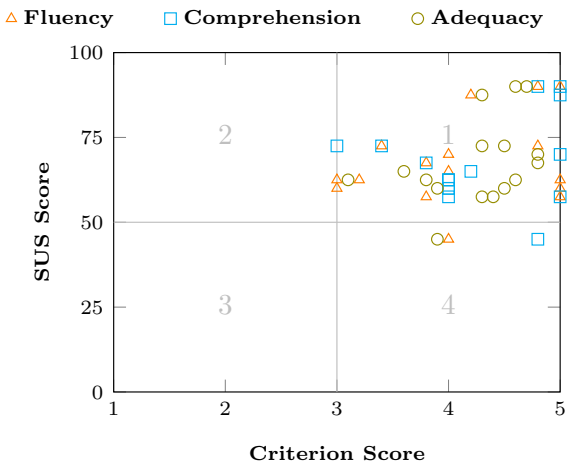

Fig. 8. SUS scores against fluency, comprehension, and adequacy. Quadrant #1: Users that liked polyglot translations and indeed they worked for them. Quadrant #2: Users that liked polyglot translations but they did not work for them. Quadrant #3: Users that did not like polyglot translations and they did not work for them. Quadrant #4: Users that did not like polyglot translations but they worked for them.

barazzato', and French 'embarrassé'. All these words are closer to Spanish 'embarazada' (English 'pregnant') than 'avergonzado' (the right translation of 'embarrassed'), so in this very particular case the chosen word would cause confusion since the original English meaning would not be preserved in any of the related languages. A plausible option to alleviate this corner case would be incorporating some form of semantic similarity, for example, using part-of-speech tagging or specialized databases such as WordNet or BabelNet. Currently, however, we believe the user should be able to recognize the actual meaning of such word by looking at its context. This notion in fact has been recently explored by others to improve MT in computer-mediated communication like messaging applications [27], where users were provided with two translations at once so they could better infer (by themselves) the meaning of the original sentence.

On the other hand, we should make the following observation. What happens when the source language is also from the family of languages used to build polyglot translations? One might think that leaving the source words untranslated would be better than replacing them. However, as our experiments suggest, it is better to provide the user with a more familiar word, provided that there is one candidate with higher similarity. Of course, if a source word is actually the better candidate, then the model would leave it "as is" in the polyglot translation.

Finally, a limitation of our current implementation is that it would not work with related languages with unrelated alphabet glyphs. For example, Romance languages have inherited many terms from Greek, however Greek uses a completely different alphabet set. This could be improved by modifying our normalized distance algorithm. In addition, many languages are agglutinative or even polysynthetic, and cannot therefore be covered by a simple vocabulary. Even the concept of "word" varies among languages and cultures, and it actually depends on the writing system. To overcome these issues, one could design low-cost edit operations, such as substituting a Portuguese "ç" by a Spanish "z" or a French "gn" by a Spanish "ñ", which would be better predictors of words' cognateness. Also, help from related languages not using the same alphabet could be made available by using simple transliteration rules. Even more, edit distances could be computed using different weights for different edit operations. This way, typical transformations between the target and the support languages could be considered as described above, which would lead to better choices.

7. Conclusion

Most language families share a common core vocabulary, so this information can be leveraged to enhance the usefulness of current MT systems. We have explored this concept with 13 languages in 3 families and have observed that, in general, polyglot translations can improve overall understanding since the words presented to the user will look the most familiar. We have focused on MT use for "assimilation" or gisting scenarios, though polyglot translations could be used for "dissemination" or post-editing scenarios by simply letting the users to amend the MT output.

By way of conclusion, we believe that polyglot MT is an important step toward overcoming resource scarcity and data sparseness problems. Our method can contribute significantly to more usable MT systems being deployed across more and more languages worldwide, allowing more of the world population to benefit from MT, irrespective of what languages do they or their applications speak.

Acknowledgments

We thank Rachel Spencer for editing earlier drafts of this article. This manuscript has been circulating for many years and so we also would like to thank any other referee who had reviewed it.

Endnotes

[1] By 'resources' we refer to *anything* that can be used to relate one language to another; e.g., parallel corpora, dictionaries, glossaries, spellcheckers, translation rules, etc.

[2] `http://www.uea.org/`

[3] `http://idolinguo.org.uk/`

[4] `http://www.interlingua.com/`

[5] A pidgin is a restricted language, with a very limited vocabulary and a simplified grammar.

[6] `http://invokeit.wordpress.com/frequency-word-lists/`

[7] `http://www.opensubtitles.org`

References

[1] S. Adolphs and N. Schmitt. Lexical coverage of spoken discourse. *Applied Linguistics*, 24(4), 2003.

[2] R. S. P. Beeke. *Comparative Indo-European Linguistics: An introduction*. John Benjamins Publishing Company, 2011.

[3] I. Boguslavsky, J. Cardeñosa, and C. Gallardo. A novel approach to creating disambiguated multilingual dictionaries. *Applied Linguistics*, 30(1), 2009.

[4] J. Brooke. SUS: A "quick and dirty" usability scale. In *Usability Evaluation in Industry*. Taylor and Francis, 1996.

[5] T. Cohn and M. Lapata. Machine translation by triangulation: Making effective use of multi-parallel corpora. In *Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, 2007.

[6] N. Habash. Four techniques for online handling of out-of-vocabulary words in Arabic-English statistical machine translation. In *Proc. Annual Meeting of the Association for Computational Linguistics on Human Language Technologies (ACL-HLT)*, 2008.

[7] R. Heredia and J. Altarriba. Bilingual code switching: Why do bilinguals code-switch? *Current Directions in Psychological Science*, 10(5), 2001.

[8] R. Hickey, editor. *The Handbook of Language Contact*. Wiley-Blackwell, 2010.

[9] D. Jones, W. Shen, and M. Herzog. Machine translation for government applications. *Lincoln Laboratory Journal*, 18(1), 2009.

[10] K. Kent. Language contact: Morphosyntactic analysis of Surzhyk spoken in Central Ukraine. In *LSO Working Papers in Linguistics. Proc. WIGL*, 2010.

[11] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. Moses: Open source toolkit for statistical machine translation. In *Proc. Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, 2007.

[12] P. Koehn, A. Birch, and R. Steinberger. 462 machine translation systems for Europe. In *Proc. MT Summit*, 2009.

[13] L. A. Leiva and V. Alabau. An automatically generated interlanguage tailored to speakers of minority but culturally influenced languages. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2012.

[14] M. P. Lewis, editor. *Ethnologue: Languages of the World*. SIL International, 17th edition, 2013.

[15] M. Masterman. *Machine Translation*, chapter Mechanical pidgin translation: An estimate of the research value of 'word-for-word' translation into a pidgin language, rather than into the full normal form of an output language. North-Holland Publishing Company, 1967.

[16] J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, The Google Books Team, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. L. Aiden. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 2010.

[17] S. Mirkin, L. Specia, N. Cancedda, I. Dagan, M. Dymetman, and I. Szpektor. Source-language entailment modeling for translating unknown terms. In *Proc. Joint Conf. Annual Meeting of the ACL and Intl. Conf. on Natural Language Processing of the AFNLP*, 2009.

[18] P. Nakov and J. Tiedemann. Combining word-level and character-level models for machine translation between closely-related languages. In *Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, 2012.

[19] J. O'Regan and M. L. Forcada. Peeking through the language barrier: the development of a free/open-source gisting system for Basque to English based on apertium.org. *Procesamiento del Lenguaje Natural*, 51, 2013.

[20] O. Streiter, K. P. Scannell, and M. Stuflesser. Implementing NLP projects for noncentral languages: instructions for funding bodies, strategies for developers. *Machine Translation*, 20(4), 2006.

[21] W. L. Taylor. "cloze procedure" a new tool for measuring readability. *Journalism Quarterly*, 30, 1953.

[22] J. Tiedemann. News from OPUS - A collection of multilingual parallel corpora with tools and interfaces. In *Proc. Recent Advances in Natural Language Processing*, 2009.

[23] J. Tiedemann. Character-based pivot translations for under-resourced languages and domains. In *Proc. European Chapter of the Association for Computational Linguistics (EACL)*, 2012.

[24] F. M. Tyers. Rule-based augmentation of training data in Breton-French statistical machine translation. In *Proc. Annual Conf. European Association of Machine Translation (EAMT)*, 2009.

[25] D. Vilar, J.-T. Peter, and H. Ney. Can we translate letters? In *Workshop on Statistical Machine Translation (WMT)*, 2007.

[26] J. Wagner. European languages, 1997. Available at `http://ielanguages.com/eurolang.html`.

[27] B. Xu, G. Gao, S. R. Fussell, and D. Cosley. Improving machine translation by showing two outputs. In *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2014.

[28] G. Zuckermann. Hybridity versus revivability. *Journal of Language Contact*, 2, 2009.

Appendix

Mathematical Notation

$$\begin{aligned}
\boldsymbol{s} &\rightarrow \text{source language sentence}\\
\boldsymbol{y} &\rightarrow \text{mixed language sentence}\\
\boldsymbol{t}_1 &\rightarrow \text{target language sentence}\\
\boldsymbol{t}_n &\rightarrow \text{sentence in related target language } n\\
\theta_1 &\rightarrow \text{knowledge of target language}\\
\theta_n &\rightarrow \text{knowledge of related target language } n\\
\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n, \ldots, \theta_N) &\rightarrow \text{full language knowledge}
\end{aligned}$$

Derivation of the Polyglot Machine Translation Model

Let $\boldsymbol{s}$ be a sentence in source language $\mathcal{L}_s$. We want to convey the message $\boldsymbol{s}$ to a user whose primary reading language is $\mathcal{L}_1$ using language resources θ_1. From a statistical point of view, the sentence that best conveys the original message can be obtained following the fundamental MT equation:

$$\boldsymbol{y}^* = \arg\max_{\boldsymbol{y}} \Pr(\boldsymbol{y}|\boldsymbol{s}, \theta_1) \tag{6}$$

where $\Pr(\boldsymbol{y}|\boldsymbol{s}, \theta_1)$ is the translation model.

In case of a system with enough resources to perform the translation, $\boldsymbol{y}^*$ would be a sentence with words wholly in $\mathcal{L}_1$. In this case, Eq. (6) can be approached as a state-of-the-art statistical translation model. However, when there are not enough resources to perform the translation, typically $\boldsymbol{y}^*$ would include words from $\mathcal{L}_s$, for which a translation is available, along with, most likely, grammatical structures from $\mathcal{L}_s$. This would not be a desirable outcome, since $\mathcal{L}_s$ may well be completely indecipherable to the user. We can improve the understandability of $\boldsymbol{y}^*$ by leveraging the resources of languages related to $\mathcal{L}_1$. So, let $\mathcal{L}_2, \ldots, \mathcal{L}_n, \ldots, \mathcal{L}_N$ be a set of N related languages, Eq. (6) can be rewritten, marginalizing over all possible translations, as

$$\begin{aligned}
\boldsymbol{y}^* &= \arg\max_{\boldsymbol{y}} \Pr(\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{\theta})\\
&= \arg\max_{\boldsymbol{y}} \sum_{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n, \ldots, \boldsymbol{t}_N} \Pr(\boldsymbol{y}, \boldsymbol{t}_1, \ldots, \boldsymbol{t}_n, \ldots, \boldsymbol{t}_N|\boldsymbol{s}, \boldsymbol{\theta})\\
&= \arg\max_{\boldsymbol{y}} \sum_{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n, \ldots, \boldsymbol{t}_N} \Pr(\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{\theta}, \boldsymbol{t}_1, \ldots, \boldsymbol{t}_n, \ldots, \boldsymbol{t}_N) \Pr(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n, \ldots, \boldsymbol{t}_N|\boldsymbol{s}, \boldsymbol{\theta})
\end{aligned} \tag{7}$$

Assuming that translations $\boldsymbol{t}_n$ are independent of each other, and that $\Pr(\boldsymbol{t}_n|\boldsymbol{s}, \boldsymbol{\theta})$ does not depend on any language resources other than θ_n, it follows that

$$\boldsymbol{y}^* \approx \arg\max_{\boldsymbol{y}} \sum_{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N} \Pr(\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{\theta}, \boldsymbol{t}_1, \ldots, \boldsymbol{t}_N) \prod_{n=1}^{N} \Pr(\boldsymbol{t}_n|\boldsymbol{s}, \theta_n) \tag{8}$$

Since calculating all possible translations is not computationally feasible, it is typically approximated by the maximum (as in our implementation), for which efficient algorithms can be developed (e.g., dynamic programming), yielding

$$\boldsymbol{y}^* \approx \arg\max_{\boldsymbol{y}} \left[\max_{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N} \Pr(\boldsymbol{y}|\boldsymbol{s}, \boldsymbol{\theta}, \boldsymbol{t}_1, \ldots, \boldsymbol{t}_N) \prod_{n=1}^{N} \Pr(\boldsymbol{t}_n|\boldsymbol{s}, \theta_n) \right] \tag{9}$$

which gives us our general model for (resource-tuned) polyglot machine translation, depicted in Eq. (3).