Modeling the Music Genre Perception across Language-Bound Cultures

Elena V. Epure[1], **Guillaume Salha**[1,2], **Manuel Moussallam**[1], **Romain Hennequin**[1]

[1] Deezer Research, Paris, France
[2] LIX, École Polytechnique, Palaiseau, France
`research@deezer.com`

Abstract

The music genre perception expressed through human annotations of artists or albums varies significantly across language-bound cultures. These variations cannot be modeled as mere translations since we also need to account for cultural differences in the music genre perception. In this work, we study the feasibility of obtaining relevant cross-lingual, culture-specific music genre annotations based only on language-specific semantic representations, namely distributed concept embeddings and ontologies. Our study, focused on six languages, shows that unsupervised cross-lingual music genre annotation is feasible with high accuracy, especially when combining both types of representations. This approach of studying music genres is the most extensive to date and has many implications in musicology and music information retrieval. Besides, we introduce a new, domain-dependent cross-lingual corpus to benchmark state of the art multilingual pre-trained embedding models.

1 Introduction

A prevalent approach to culturally study music genres starts with a common set of music items, *e.g.* artists, albums, tracks, and assumes that the same music genres would be associated with the items in all cultures (Ferwerda and Schedl, 2016; Skowron et al., 2017). However, music genres are subjective. Cultures themselves and individual musicological backgrounds influence the music genre perception, which can differ among individuals (Sordo et al., 2008; Lee et al., 2013). For instance, a Westerner may relate *funk* to *soul* and *jazz*, while a Brazilian to *baile funk* that is a type of *rap* (Hennequin et al., 2018). Thus, accounting for cultural differences in music genres' perception could give a more grounded basis for such cultural studies. However, ensuring both a common set of music items and culture-sensitive annotations with broad coverage of music genres is strenuous (Bogdanov et al., 2019).

To address this challenge, we study the feasibility of cross-culturally annotating music items with music genres, without relying on a parallel corpus. In this work, culture is related to a community speaking the same language (Kramsch and Widdowson, 1998). The specific research question we build upon is: assuming consistent patterns of music genres association with music items within cultures, can a mapping between these patterns be learned by relying on language-specific semantic representations? It is worth noting that, since music genres fall within the class of Culture-Specific Items (Aixelá, 1996; Newmark, 1988), cross-lingual annotation, in this case, cannot be framed as standard translation, as one also needs to model the dissimilar perception of music genres across cultures.

Our work focuses on four language families, Germanic (English-**en** and Dutch-**nl**), Romance (Spanish-**es** and French-**fr**), Japonic (Japanese-**ja**), Slavic (Czech-**cs**), and on two types of language-specific semantic representations, ontologies and multi-word expression embeddings.

First, ontologies are often used to represent music genres, showing how they relate conceptually (Schreiber, 2016). We identify Wikipedia[1], the online multilingual encyclopedia, to be particularly relevant to our study. It extensively documents worldwide music genres relating them through a coherent set of relation types across languages (*e.g. derivative genre*, *sub-genre*). Though the relations types are the same per language, the actual music genres and the way they are related can differ. Indeed, Pfeil et al. (Pfeil et al., 2006) have shown that Wikipedia contributions expose cultural differences aligned with the ones in the physical world.

Second, music genres can be represented from a distributional semantics perspective. Word vector

---

[1]`https://en.wikipedia.org`

4765

spaces are generated from large corpora following the distributional hypothesis, *i.e.* words with similar contexts have akin meanings. As languages are passed on culturally, we assume that the language-specific corpora used to create these spaces are sufficient to convey concepts' cultural specificity into their vector representations. In our study, we focus on multiple recent multilingual pre-trained models to generate word or sentence[2] embeddings (Arora et al., 2017; Grave et al., 2018; Artetxe and Schwenk, 2019; Devlin et al., 2019; Lample and Conneau, 2019), to account for variances in the used corpora or model designs.

Lastly, we combine the semantic representations by retrofitting distributed music genre embeddings to music genre ontologies. Retrofitting (Faruqui et al., 2015) modifies each concept embedding such that the representation is still close to the distributed one, but also encodes ontology information. Initially, we retrofit music genres per language, using monolingual ontologies. Then, by partially aligning these ontologies, we apply retrofitting to learn multilingual embeddings from scratch.

The results show that we can model the cross-lingual music genre annotation with high accuracy by combining both types of language-specific semantic representations. When comparing the representations derived from multilingual pre-trained models, the smooth inverse frequency averaging (Arora et al., 2017) of aligned word embeddings outperforms the state of the art approaches. To our knowledge, this simple method has been rarely used to embed multilingual sentences (Vargas et al., 2019), and we hypothesize its potential as a strong baseline on other cross-lingual datasets and tasks too. Finally, embedding learning based on retrofitting leads to better multilingual music genre representations than when inferred with pre-trained embedding models. This opens the possibility to learn embeddings for rare music genres or languages when aligned music genre ontologies are available.

Summing up, our contributions are: 1) a study on how effective language-specific semantic representations of music genres are for modeling cross-lingual annotation, without relying on a parallel music item corpus; 2) an extensive evaluation of

multilingual pre-trained embedding models to derive representations for multi-word concepts in the music domain. Our study can enable complete musicological research, but also localized music information retrieval. This latter application is crucial for online music streaming platforms that leverage music genre annotations to provide worldwide, user-personalized music recommendations.

Our domain-specific study complements other works benchmarking general-language sentence representations (Conneau et al., 2018). Finally, we provide an in-depth formal analysis of the retrofitting part of our method. We prove the strict convexity of retrofitting and show that the ontology concepts' final embeddings converge to the same values despite the order in which concepts are iteratively updated, on condition that we know a single initial node embedding in each connected component of the ontology.

2 Related Work

Music genres are conceptual representations encompassing a set of conventions between the music industry, artists, and listeners about individual music styles (Lena, 2012). From a cultural perspective, it has been shown that there are differences in how people listen to music genres. (Ferwerda and Schedl, 2016; Skowron et al., 2017). Average listening habits in some countries span across many music genres and are less diverse in other countries (Ferwerda and Schedl, 2016). Also, cultural dimensions proved strong predictors for the popularity of specific music genres (Skowron et al., 2017).

Despite the apparent agreement on the music style for which the music genres stand, conveyed in the earlier definition and implied in the related works too, music genres are subjective concepts (Sordo et al., 2008; Lee et al., 2013). To address this subjectivity, Bogdanov et al. (2019) proposed a dataset of music items annotated with English music genres by different sources. In this line of work, we address the divergent perception of music genres. Still, we focus on multilingual, unsupervised music genre annotation without relying on content features, *i.e.* audio or lyrics. We also complement similar studies in other domains (art: Eleta and Golbeck, 2012) with another research method.

Then, in the literature, there are other works that benchmark pre-trained word and sentence embedding models. van der Heijden et al. (2019) compares multilingual contextual language models for

---

[2]Music genres can be multi-word expressions. Previous work (Shwartz and Dagan, 2019a) successfully embed phrases with sentence embedding models. In particular, contextualized language models result in more meaningful representations for diverse composition tasks.

named-entity recognition and part-of-speech tagging. Shwartz and Dagan (2019b) use multiple static and contextual word embeddings to represent multi-word expressions and assess their capacity to capture meaning shift and implicit meaning in compositionality. Conneau et al. (2018) formulate a new task to evaluate cross-lingual sentence representation centered on natural language inference.

Compared to these works, our benchmark is aimed at cross-lingual annotation; we target a specific domain, music, for which we try concept embedding adaptation with retrofitting; and we also test a multilingual sentence representation obtained with smooth inverse frequency averaging of multilingual word embeddings. As discussed in a recent survey on cross-lingual word embedding models (Ruder et al., 2019), there is a need to unlock domain-specific data to assess if general-language sentence representations are also accurate across domains. Our work builds towards this goal.

3 Cross-lingual Music Genre Annotation

Further, we formalize the cross-lingual annotation task and the strategy to evaluate it in Section 3.1. We describe the test corpus used in this work, together with its collection procedure in Section 3.2.

3.1 Problem Formalization

The cross-lingual music genre annotation consists of inferring, for music items, tags in a target language L_t, knowing tags in a source language L_s. For instance, knowing the English music genres of *Fatboy Slim* (*big beat*, *electronica*, *alternative rock*), the goal is to predict *rave* and *rock alternativo* in Spanish. As shown in the example, but also Section 1, the problem goes beyond translation and instead targets a model able to map concepts, potentially dissimilar, across languages and cultures.

Formally, given S a set of tags in language L_s, $\mathcal{P}$ the partitions of S and T a set of tags in language L_t, a mapping scoring function $f : \mathcal{P}(S) \rightarrow \mathbb{R}^{|T|}$ can attribute a prediction score to each target tag, relying on subsets of source tags drawn from S (Hennequin et al., 2018; Epure et al., 2019, 2020). The produced score incorporates the degree of relatedness of each particular input source tag to the target tag. A common approach to compute relatedness in distributional semantics relies on cosine similarity. Thus, for $\{s_1, ..., s_K\}$ source tags and

Language	nl	fr	es	cs	ja
en	12604	28252	32891	4772	14752
nl		7139	7689	1885	3426
fr			15616	3046	8622
es				3245	7644
cs					2065

Table 1: Number of music items for language pair.

Language	Number of unique music genres	
	Corpus *(Avg. per item)*	Ontology
en	558 *(2.12 ± 1.34)*	10748
nl	204 *(1.71 ± 1.06)*	1529
fr	364 *(1.75 ± 1.06)*	2905
es	525 *(2.11 ± 1.34)*	3988
cs	133 *(2.23 ± 1.34)*	1418
ja	192 *(1.51 ± 1.11)*	1609

Table 2: Number of unique music genres in the corpus (Section 3.2) and in the ontology (Section 4.1).

any target tag t, f can be defined as:

$$f_t(\{s_1, s_2, \ldots, s_K\}) = \frac{1}{K} \sum_{k=1}^{K} \frac{\mathbf{s}_k^T \mathbf{t}}{||\mathbf{s}||_2 ||\mathbf{t}||_2}, \quad (1)$$

where $|| \cdot ||_2$ is the Euclidean norm.

3.2 Test Corpus

Wikipedia records worldwide music artists and their discographies, with a frequent mentioning of their music genres. By manually checking the Wikipedia pages of miscellaneous music items, we observed that their music genres vary significantly across languages. For instance, *Knights of Cydonia*, a single by *Muse*, was annotated in Spanish as *progressive rock*, while in Dutch as *progressive metal* and *alternative rock*. In Figure 1, we show another example of different annotations in English, Spanish, and Japanese from Wikipedia infoboxes. As Wikipedia writing is localized, contributors' culture can lead to differences in the multilingual content on the same topic (Pfeil et al., 2006), particularly for subjective matters. Thus, Wikipedia was a suitable source for assembling the test corpus.

Using DBpedia (Auer et al., 2007) as a proxy to Wikipedia, we collected music items such as artists and albums, annotated with music genres in at least two of the six languages (**en**, **nl**, **fr**, **es**, **cs** and **ja**). We targeted *MusicalWork*, *MusicalArtist* and *Band* DBpedia resource types, and we only kept music items that were annotated with music genres which appeared at least 15 times in the corpus. Our final corpus includes 63246 music items. The number of annotations for each language pair is presented in Table 1. We also show in Table 2 the number of

Figure 1: Wikipedia infoboxes of Puerto Rican artist *Milton Cardona*, in English (**en**), Spanish (**es**) and Japanese (**ja**) languages. Some music genre annotations are culture-specific, such as *World*/ワールドミュージック which is present for **en** and **ja** but not for **es**, or 実験音楽 (*Experimental Music*) for **ja** only.

unique music genres per language in the corpus and the average number of tags for each music item.

The **en** and **es** languages use the most diverse tags. This can be because more annotations exist in these languages, in comparison to **cs**, which has the least annotations and least diverse tags. However, the mean number of tags per item appears relatively high for **cs**, while **ja** has the smallest mean number of tags per item.

4 Language-specific Semantic Representations for Music Genres

This work aims to assess the possibility of obtaining relevant cross-lingual music genre annotations, able to capture cultural differences too, by relying on language-specific semantic representations. Two types of semantic representations are investigated given their popularity: ontologies to represent music genre relations (presented in Section 4.1) and distributed embeddings to represent multi-word expressions in general (presented in Section 4.2). In contrast to this unsupervised approach, mapping patterns of associating music genres with music items across cultures could have also been enabled with a parallel corpus. However, gathering a corpus that includes all music genres for each pair of languages is challenging.

4.1 Music Genre Ontology

Conceptually, music genres are interconnected entities. For example, *rap west coast* is a sub-genre of *hiphop* or *música electrónica* is the origin of *synthpunk*. Academic and practitioner communities often use ontologies or knowledge graphs to represent music genre relations and enrich the music genre definitions (Schreiber, 2016; Lisena et al., 2018). As mentioned in Section 1, we use in this study Wikipedia-based music genre ontologies because the multilingual Wikipedia contributions on the same topic can differ and these differences have been proven aligned with the ones in the physical world (Pfeil et al., 2006).

We further describe how we crawl the Wikipedia-based music genres ontologies for the six languages by relying on DBpedia. For each language, first, we constitute the seed list using two sources: the DBpedia resources of type *MusicGenre* and their aliases linked through the *wikiPageRedirects* relation; the music genres discovered when collecting the test corpus (introduced in Section 3.1) and their aliases. Then, music genres are fetched by visiting the DBpedia resources linked to the seeds through the relations *wikiPageRedirects*, *musicSubgenre*, *stylisticOrigin*, *musicFusionGenre* and *derivative*[3]. The seed list is updated each time, allowing the crawling to continue until no new resource is found.

In DBpedia, resources are sometimes linked to their equivalents in other languages through the relation *sameAs*. For most experiments, we rely on monolingual music genres ontologies. How-

---

[3]We present these relations by their English names, which may be translated in the DBpedia versions in other languages.

ever, we also collect the cross-lingual links between music genres to include a translation baseline for cross-lingual annotation, *i.e.* for each music genre in a source language, we predict its equivalent in a target language using DBpedia. Besides, we try to learn aligned embeddings from scratch by relying on these partially aligned music genre ontologies, as will be discussed in Section 4.3.

The number of unique Wikipedia music genres discovered in each language is presented in Table 2. Let us note that the graph numbers are much larger than the test corpus numbers, emphasizing the challenge to constitute a parallel corpus that covers all language-specific music genres.

4.2 Music Genre Distributed Representations

As music genres are multi-word expressions, we make use of existing sentence representation models. We also inquire into word vector spaces and obtain sentence embeddings by hypothesizing that music genres are generally compositional, *i.e.* the sense of a multi-word expression is conveyed by the sense of each composing word (*e.g. West Coast rap*, *jazz blues*; there are also non-compositional examples like *hard rock*). We set our investigation scope to multilingual pre-trained embedding models, and we consider both static and contextual word/sentence representations as described next.

Multilingual Static Word Embeddings. The classical word embeddings we study are the multilingual fastText word vectors trained on Wikipedia and Common Crawl (Grave et al., 2018). The model is an extension of the Common Bag of Word Model (CBOW, Mikolov et al., 2013), which includes subword and word position information. The fastText word vectors are trained in distinct languages. Thus, we must ensure that the monolingual word vectors are projected in the same space for cross-lingual annotation. We perform the alignment with the method proposed by Joulin et al. (2018), which treats word translation as a retrieval task and introduces a new loss relying on a relaxed cross-domain similarity local scaling criterion.

Multilingual Contextual Word Embeddings. Contextual word embeddings (Peters et al., 2017; Devlin et al., 2019), in contrast to the classical ones, are dynamically inferred based on the given context sentence. This type of embedding can address polysemy as the word sense is disambiguated through the surrounding text. In our work, we include two recent contextualized language models compatible

with the multilingual scope: multilingual Bidirectional Encoder Representations from Transformers (BERT, Devlin et al., 2019) and Cross-lingual Language Model (XLM, Lample and Conneau, 2019).

BERT (Devlin et al., 2019) is trained to jointly predict a masked word in a sentence and whether sentences are successive text segments. Similar to fastText (Grave et al., 2018), subword and word position information is also used. An input sentence is tokenized against a limited token vocabulary with a modified version of the byte pair encoding algorithm (BPE, Sennrich et al., 2016). Multilingual BERT is trained as a single-language model, fed with 104 concatenated monolingual Wikipedias (Pires et al., 2019; Wu and Dredze, 2019).

XLM (Lample and Conneau, 2019) has a similar architecture to BERT. Also, it shares with BERT one training objective, the masked word prediction, and the tokenization using BPE, but applied on sentences differently sampled from each monolingual Common Crawl corpus. Compared to BERT, two other objectives are introduced, to predict a word from previous words and a masked word by leveraging two parallel sentences. Thus, to train XLM, several multilingual aligned corpora are used (Lample and Conneau, 2019).

Multilingual Sentence Embeddings. Contextualized language models can be exploited in multiple ways. First, as Lample and Conneau (2019) show, by training the transformers on multi-lingual data, cross-lingual word vectors are obtained in an unsupervised way. The word vectors can be accessed through the model lookup table. These embeddings are merely aligned but not contextual, thus directly comparable to fastText. For these three types of cross-lingual non-contextual word embeddings, fastText (**FT**), the multilingual BERT's lookup table (**mBERT**) and the XLM's lookup table (**XLM**), we compute the sentence embedding using the standard average (*avg*) or the smooth inverse frequency averaging (*sif*) introduced by Arora et al. (2017).

Formally, let c denote a music genre composed of multiple tokens $\{t_1, t_2, \ldots, t_M\}$, $\mathbf{t}_m$ the embedding of each token t_m initialized from a given pre-trained embedding model or to d-dimensional[4] null vector $\mathbf{0}_d$ if t_m is absent from the model vocabulary, and $\hat{\mathbf{q}}_i \in \mathbb{R}^d$ the representation of c which we want to infer. The *avg* strategy computes $\hat{\mathbf{q}}_i$ as

---

[4]In Appendix B, we report d for each embedding type.

$\frac{1}{M} \sum_{m=1}^{M} \mathbf{t}_m$. The *sif* strategy computes $\hat{\mathbf{q}}_i$ as:

$$\overline{\mathbf{q}}_i = \frac{1}{M} \sum_{m=1}^{M} \frac{a}{a + f_{t_m}} \mathbf{t}_m \tag{2}$$

$$\hat{\mathbf{q}}_i = \overline{\mathbf{q}}_i - \mathbf{u}\mathbf{u}^T \overline{\mathbf{q}}_i \tag{3}$$

where f_{t_m} is the frequency of t_m, a is a hyper-parameter usually fixed to 10^{-3} (Arora et al., 2017) and $\mathbf{u}$ is the first singular vector obtained through the singular value decomposition (Golub and Reinsch, 1971) of $\overline{\mathbf{Q}}$, the embedding matrix computed with the Equation 2 for all music genres. Vocabulary tokens of pre-trained embedding models are usually sorted by decreasing frequency in the training corpus, *i.e.* the higher the rank, the more frequent the token. Thus, based on the Zipf's law (Zipf, 1949), f_{w_m} can be approximated by $1/z_{t_m}$, z_{t_m} being the rank of t_m. The intuition of this simple sentence embedding method is that uncommon words are semantically more informative.

Second, contextualized language models can be used as feature extractors representing sentences from the contextual embeddings of the associated tokens. Multiple strategies exist to retrieve contextual token embeddings: to use the embeddings layer or the last hidden layer or to apply min or max pooling over time (Devlin et al., 2019). To infer a fixed-length representation of a multi-word music genre, we try max and mean pooling over token embeddings (Lample and Conneau, 2019; Reimers and Gurevych, 2019), obtained with the diverse strategies mentioned before. We denote these sentence embeddings $\mathbf{XLM}_{Ctxt}$ and $\mathbf{mBERT}_{Ctxt}$.

The contextualized language models can be further fine-tuned for particular downstream tasks, yielding better sentence representations (Eisenschlos et al., 2019; Lample and Conneau, 2019). Existing evaluations of cross-lingual sentence representations are centered on natural language inference (XNLI, Conneau et al., 2018) or classification (Eisenschlos et al., 2019). The cross-lingual music genre annotation would be closer to the XNLI task; hence we could fine-tune the pre-trained models on a parallel corpus of music genres translations or music genre annotations. However, our research investigates language-specific semantic representations. Also, using translated music genres would not model their different perception across cultures while obtaining an exhaustive corpus of cross-lingual annotations is challenging.

Last, we explore **LASER**, a universal language-agnostic sentence embedding model (Artetxe and Schwenk, 2019). The model is based on a BiLSTM encoder trained on corpora in 93 languages to learn multilingual fixed-length sentence embeddings. As in other models, sentences are tokenized against a fixed vocabulary, obtained with BPE from the concatenated multilingual corpora. **LASER** appears highly effective without requiring task-specific fine-tuning (Artetxe and Schwenk, 2019).

4.3 Retrofitting Music Genre Distributed Representations to Ontologies

Retrofitting (Faruqui et al., 2015) is a method to refine vector space word representations by considering the relations between words as defined in semantic lexicons such as WordNet (Miller, 1995). The intuition is to modify the distributed embeddings to become closer to the representations of the concepts to which they are related. Ever since the original work, many uses of retrofitting have been explored to semantically specialize word embeddings in relations such as synonyms or antonyms (Kiela et al., 2015; Kim et al., 2016), in other languages than a source one (Ponti et al., 2019) or in specific domains (Hangya et al., 2018).

Enhanced extensions of retrofitting exist, but they require supervision (Lengerich et al., 2018). The original method (Faruqui et al., 2015) is unsupervised and can simply yet effectively leverage distributed embeddings and ontologies for improved representations. Thus, we mainly rely on it, but we apply some changes as further described. Let $\Omega = (C, E)$ be an ontology including the concepts C and the semantic relations between these concepts $E \subseteq C \times C$. The retrofitting goal is to learn new concept embeddings, $\mathbf{Q} \in \mathbb{R}^{n \times d}$ with $n = |C|$ and d the embedding dimension. The learning starts with initializing each $\mathbf{q}_i \in \mathbb{R}^d$, the new embedding for concept $i \in C$, to $\hat{\mathbf{q}}_i$, the initial distributed embedding, and then iteratively updates $\mathbf{q}_i$ until convergence as follows:

$$\mathbf{q}_i \leftarrow \frac{\sum_{j:(i,j)\in E} (\beta_{ij} + \beta_{ji}) \mathbf{q}_j + \alpha_i \hat{\mathbf{q}}_i}{\sum_{j:(i,j)\in E} (\beta_{ij} + \beta_{ji}) + \alpha_i} \tag{4}$$

α and β are positive scalars weighting the importance of the initial, respectively, the related concept embeddings in computation. The formula was reached through the optimization of the retrofitting objective using the Jacobi method (Saad, 2003).

Equation 4 is a corrected version of the original work, as for a concept i, not only β_{ij} appears in it, but also β_{ji}. That is to say that when computing the

Pair	GTrans	DBpSameAs	mBERT$_{avg}$	FT$_{sif}$	XLM$_{Ctxt}$	LASER	Rfit$_{u\Omega}$FT$_{sif}$
en-nl	59.9 ± 0.3	72.2 ± 0.2	86.2 ± 0.2	86.5 ± 0.1	85.4 ± 0.3	80.9 ± 0.1	$\mathbf{90.0 \pm 0.1}$
en-fr	58.4 ± 0.1	70.0 ± 0.2	85.2 ± 0.3	87.4 ± 0.3	86.6 ± 0.2	82.0 ± 0.2	$\mathbf{90.8 \pm 0.2}$
en-es	56.9 ± 0.0	65.4 ± 0.2	83.8 ± 0.1	86.9 ± 0.2	85.4 ± 0.3	81.8 ± 0.1	$\mathbf{89.9 \pm 0.1}$
en-cs	60.6 ± 0.6	78.4 ± 0.6	88.2 ± 0.4	88.6 ± 0.4	89.0 ± 0.5	85.4 ± 0.3	$\mathbf{90.4 \pm 0.3}$
en-ja	60.9 ± 0.1	70.4 ± 0.2	72.4 ± 0.2	80.8 ± 0.3	74.2 ± 0.3	73.0 ± 0.1	$\mathbf{86.7 \pm 0.3}$
nl-en	53.5 ± 0.1	56.7 ± 0.3	74.6 ± 0.3	79.8 ± 0.4	74.2 ± 0.7	70.5 ± 0.1	$\mathbf{84.3 \pm 0.1}$
nl-fr	54.4 ± 0.2	60.0 ± 0.4	76.1 ± 0.3	79.3 ± 0.8	74.9 ± 1.0	72.6 ± 0.5	$\mathbf{81.5 \pm 0.7}$
nl-es	53.1 ± 0.2	56.8 ± 0.2	75.0 ± 0.4	77.7 ± 0.5	73.5 ± 0.3	71.0 ± 0.3	$\mathbf{80.5 \pm 0.4}$
nl-cs	57.8 ± 0.1	50.0 ± 0.0	80.8 ± 0.4	80.6 ± 0.2	79.0 ± 0.4	76.8 ± 0.2	$\mathbf{83.4 \pm 0.5}$
nl-ja	57.5 ± 0.4	62.7 ± 0.4	65.8 ± 2.5	74.9 ± 1.0	66.0 ± 0.6	68.5 ± 0.2	$\mathbf{80.0 \pm 0.7}$
fr-nl	58.6 ± 0.1	65.3 ± 0.4	79.4 ± 0.1	81.9 ± 0.4	78.6 ± 0.1	74.3 ± 0.4	$\mathbf{84.7 \pm 0.3}$
fr-en	55.3 ± 0.0	59.7 ± 0.2	77.2 ± 0.5	83.0 ± 0.2	78.8 ± 0.5	74.2 ± 0.4	$\mathbf{87.7 \pm 0.1}$
fr-es	54.1 ± 0.1	59.0 ± 0.1	77.5 ± 0.4	81.8 ± 0.3	78.7 ± 0.5	75.4 ± 0.6	$\mathbf{85.3 \pm 0.2}$
fr-cs	59.1 ± 0.3	70.0 ± 0.6	82.7 ± 0.6	83.9 ± 0.4	83.1 ± 0.2	80.2 ± 0.5	$\mathbf{87.2 \pm 0.3}$
fr-ja	59.1 ± 0.2	64.7 ± 0.5	61.5 ± 0.2	77.9 ± 0.1	69.5 ± 0.3	70.5 ± 0.3	$\mathbf{81.4 \pm 0.3}$
es-nl	59.8 ± 0.3	67.2 ± 0.2	82.3 ± 0.3	82.8 ± 0.9	81.3 ± 0.8	76.7 ± 0.5	$\mathbf{85.9 \pm 0.6}$
es-fr	57.4 ± 0.2	64.8 ± 0.3	81.0 ± 0.3	85.0 ± 0.3	82.2 ± 0.5	78.1 ± 0.8	$\mathbf{87.5 \pm 0.3}$
es-en	57.0 ± 0.1	61.7 ± 0.0	78.4 ± 0.3	84.7 ± 0.2	79.5 ± 0.2	74.8 ± 0.4	$\mathbf{88.8 \pm 0.3}$
es-cs	60.3 ± 0.2	72.2 ± 0.4	85.5 ± 0.5	85.6 ± 0.6	85.9 ± 0.5	82.7 ± 0.5	$\mathbf{88.0 \pm 0.4}$
es-ja	60.9 ± 0.1	67.0 ± 0.5	67.7 ± 0.1	78.3 ± 0.6	72.8 ± 0.8	71.2 ± 0.4	$\mathbf{83.1 \pm 0.6}$
cs-nl	57.6 ± 0.6	50.0 ± 0.0	78.5 ± 0.7	78.3 ± 0.9	78.1 ± 0.4	73.1 ± 0.4	$\mathbf{81.1 \pm 1.2}$
cs-fr	54.2 ± 0.2	60.0 ± 0.3	77.1 ± 1.0	78.5 ± 0.2	79.9 ± 0.7	75.0 ± 0.9	$\mathbf{81.4 \pm 0.3}$
cs-es	53.7 ± 0.4	56.9 ± 0.3	75.8 ± 0.3	77.7 ± 0.8	78.8 ± 0.4	73.7 ± 0.3	$\mathbf{81.6 \pm 0.9}$
cs-en	54.2 ± 0.2	57.1 ± 0.1	74.2 ± 0.2	78.9 ± 0.1	78.3 ± 0.4	73.4 ± 0.4	$\mathbf{84.5 \pm 0.4}$
cs-ja	58.6 ± 0.2	64.0 ± 0.3	65.8 ± 1.4	76.9 ± 0.1	72.2 ± 1.1	72.2 ± 1.1	$\mathbf{80.5 \pm 0.5}$
ja-nl	54.8 ± 0.4	61.6 ± 1.1	62.1 ± 0.4	72.8 ± 1.0	63.6 ± 0.9	68.3 ± 1.2	$\mathbf{76.9 \pm 0.3}$
ja-fr	53.3 ± 0.2	58.4 ± 0.2	50.6 ± 0.2	73.7 ± 0.6	58.4 ± 0.4	66.7 ± 0.3	$\mathbf{77.8 \pm 0.1}$
ja-es	52.7 ± 0.1	55.9 ± 0.4	56.1 ± 0.4	73.9 ± 0.4	60.0 ± 0.2	67.4 ± 0.4	$\mathbf{78.8 \pm 0.5}$
ja-cs	56.1 ± 0.5	65.7 ± 0.7	64.7 ± 0.6	77.5 ± 0.2	64.5 ± 0.5	73.3 ± 0.6	$\mathbf{80.7 \pm 0.4}$
ja-en	52.5 ± 0.1	55.8 ± 0.1	49.0 ± 0.1	75.6 ± 0.3	56.8 ± 1.0	64.2 ± 0.4	$\mathbf{81.6 \pm 0.8}$

Table 3: Macro-AUC scores (in %, best overall in bold, best locally underlined). The first part corresponds to the translation baselines; the second to the best distributed representations; the last to the retrofitted $\mathbf{FT}_{sif}$ vectors.

partial derivative of the retrofitting objective concerning i, two non-zero terms are corresponding to the related concept j: when i is the source and j is the target and vice-versa (Bengio et al., 2006; Saha et al., 2016). The further modifications that we make regard the parameters α and β. For each $i \in C$, Faruqui et al. (2015) fix α_i to 1, and β_{ij} to $\frac{1}{degree(i)}$ for $(i, j) \in E$ or 0 otherwise; $degree(i)$ is the number of related concepts i has in Ω.

While many embedding models can handle unknown words nowadays, concepts may still have unknown initial distributed vectors, depending on the model's choice. For this case, expanded retrofitting (Speer and Chin, 2016) has been proposed, considering $\alpha_i = 0$, for each concept i with unknown initial distributed vector, and $\alpha_i = 1$ for the rest. Thus, $\mathbf{q}_i$ is initialized to $\mathbf{0}_d$ and updated by averaging the embeddings of its related concepts at each iteration. Let us notice that, through retrofitting, representations are not only modified but also learned from scratch for some concepts.

We adopt the same approach to initialize α. Moreover, we also adjust the parameters β to weight the importance of each related concept em-

bedding depending on the relation semantics in our music genre ontology (Epure et al., 2020). Specifically, we distinguish between equivalence and relatedness as follows:

$$\overline{\beta}_{ij} = \begin{cases} 1 & : (i,j) \in E_\epsilon \subset E \\ \beta_{ij} & : (i,j) \in E - E_\epsilon \\ 0, & : (i,j) \notin E \end{cases}$$

where E_ϵ contains the equivalence relation types (*wikiPageRedirects*, *sameAs*); $E - E_\epsilon$ contains the relatedness relation types (*stylisticOrigin*, *musicSubgenre*, *derivative*, *musicFusionGenre*). We label this modified version of retrofitting as **Rfit**.

Finally, we want to highlight a crucial aspect of retrofitting. Previous works (Speer and Chin, 2016; Hayes, 2019; Fang et al., 2019) claim that, while the retrofitting updating procedure converges, the results depend on the order in which the updates are made. We prove in Appendix A that the retrofitting objective is strictly convex when at least one initial concept vector is known in each connected component. Hence, with this condition satisfied, retrofitting converges to the same solution always and independently of the updates' order.

5 Experiments

Cross-lingual music genre annotation, as formalized in Section 3, is a typical multi-label prediction task. For evaluation, we use the Area Under the receiver operating characteristic Curve (AUC, Bradley, 1997), macro-averaged. We report the mean and standard deviations of the macro AUC scores using 3-fold cross-validation. For each language, we apply an iterative split (Sechidis et al., 2011) of the test corpus that balances the number of samples and the tag distributions across the folds. We pre-process the music genres by either replacing special characters with space (_-/,) or removing them ((')':.!$). For Japanese, we introduce spaces between tokens obtained with Mecab[5]. Embeddings are then computed from pre-processed tags.

We test two translation baselines, one based on Google Translate[6] (**GTrans**) and one on the DBpedia *SameAs* relation (**DBpSameAs**). In this case, a source music genre is mapped on a single or no target music genre, its embedding being in the form $\{\mathbf{0,1}\}^{|T|}$. For **XLM**$_{Ctxt}$, we compute the sentence embedding by averaging the token embeddings obtained with mean pooling across all layers. For **mBERT**$_{Ctxt}$, we apply the same strategy, but by max pooling the token embeddings instead. We chose these representations as they showed the best performance experimentally compared to the other strategies described in Section 4.2.

When retrofitting language-specific music genre embeddings, we use the corresponding monolingual ontology (**Rfit**$_{u\Omega}$). When we learn multilingual embeddings from scratch with retrofitting, by knowing only music genre embeddings in one language (*la*), we use the partially aligned DBpedia ontologies which contain the *SameAs* relations (**Rfit**$_{a\Omega}^{la}$). For this case, we also propose a baseline representing a source concept embedding as a vector of geodesic distances in the partially aligned ontologies to each target concept (**DBp**$_{a\Omega}$**NNDist**).

Results. Table 3 shows the cross-lingual annotation results. The standard translation, **GTrans**, leads to the lowest results being over-performed by a knowledge-based translation, more adapted to this domain (**DBpSameAs**). Also, these results show that translation methods fail to capture the dissimilar cross-cultural music genre perception.

The second part of Table 3 contains only the

Pair	DBp$_{a\Omega}$NNDist	Rfit$_{a\Omega}^{en}$FT$_{sif}$
en-nl	83.7 ± 0.1	**90.7 ± 0.0**
en-fr	82.7 ± 0.3	**91.7 ± 0.1**
en-es	81.1 ± 0.3	**91.4 ± 0.2**
en-cs	86.6 ± 0.3	**91.6 ± 0.4**
en-ja	81.3 ± 0.1	**89.1 ± 0.2**
Pair	DBp$_{a\Omega}$NNDist	Rfit$_{a\Omega}^{ja}$FT$_{sif}$
ja-nl	68.5 ± 0.7	75.7 ± 0.3
ja-fr	71.9 ± 0.1	76.7 ± 0.5
ja-es	68.9 ± 0.4	76.1 ± 0.5
ja-cs	77.9 ± 0.8	**82.2 ± 0.8**
ja-en	70.4 ± 0.4	**82.3 ± 0.5**

Table 4: Macro-AUC scores (in %; those larger than **Rfit**$_{u\Omega}$**FT**$_{sif}$ in Table 3 in bold) with vectors learned by retrofitting to aligned monolingual ontologies.

best[7] music genre embeddings computed with each word/sentence pre-trained model or method. When averaging static multilingual word embeddings, those from **mBERT** often yield the most relevant cross-lingual annotations, while when applying the *sif* averaging, the aligned **FT** word vectors are the best choice. Between the two contextual word embedding models, **XLM**$_{Ctxt}$ significantly outperforms **mBERT**$_{Ctxt}$, thus we report only the former.

We can notice that all distributed representations of music genres can model quite well the varying music genre annotation across languages. **FT**$_{sif}$ results in the most relevant cross-lingual annotations consistently for 5 out of 6 languages as a source. For **cs** though, the embeddings from **XLM**$_{Ctxt}$ are sometimes slightly better. **LASER** under-performs for most languages but **ja**, for which the vectors obtained with **mBERT**$_{avg}$ are less suitable.

The last column of Table 3 shows the results of cross-lingual annotation when using the **FT**$_{sif}$ vectors retrofitted to monolingual music genre ontologies. The domain adaptation of concept embeddings, inferred with general-language pre-trained models, significantly improves music genre annotation modeling across all pairs of languages.

Table 4 shows the results when using retrofitting to learn music genre embeddings from scratch. Here, distributed vectors are known for one language (**en**, respectively **ja**[8]) and the monolingual ontologies are partially aligned. Even though not necessarily all music genres are linked to their equivalents in the other language, the concept representations learned in this way are more relevant for cross-lingual annotation, for all pairs involving **en** as the source and for **ja-cs** and **ja-en**. In fact, the

---

[5]https://taku910.github.io/mecab/
[6]https://translate.google.com

[7]The complete results are presented in Appendix B.
[8]The results for the other languages are in Appendix B.

baseline (**DBp**$_{a\Omega}$**NNDist**) reveals that the aligned ontologies stand-alone can model the cross-lingual annotation quite well, in particular for **en**.

Discussion. The results show that using translation to produce cross-lingual annotations is limited as it does not consider the culturally divergent perception of music genres. Instead, monolingual semantic representations can model this phenomenon rather well. For instance, from *Milton Cardona*'s music genres in **es**, *salsa* and *jazz*, it correctly predicts the Japanese equivalent of *fusion* (フュージョン) in **ja**. Yet, while a thorough qualitative analysis requires more work, preliminary exploration suggests that larger gaps in perception might still be inadequately modeled. For instance, for *Santana*'s album *Welcome* tagged with *jazz* in **es**, it does not predict *pop* in **fr**.

When comparing the distributed embeddings, a simple method that relies on a weighted average of multilingual aligned word vectors significantly outperforms the others. Although rarely used before, we question if we can notice such high performance with other multilingual data-sets. The cross-lingual annotations are further improved by retrofitting the distributed embeddings to monolingual ontologies. Interestingly, the vector alignment does not appear degraded by retrofitting to disjoint graphs. Or, the negative impact is limited and exceeded by introducing domain knowledge in representations. Further, as shown in Table 4, joining semantic representations in this way proves very suitable to learn music genre vectors from scratch.

Regarding the scores per language, we obtained the lowest ones for **ja** as the source. We could explain this by either a more challenging test corpus or still incompatible embeddings in **ja**, possibly because of the quality of the individual embedding models for this language and the completeness of the Japanese music genre ontology. Also, we did not notice any particular improvement for pairs of languages from the same language family, *e.g.* **fr** and **es**. However, we would need a sufficiently sizeable parallel corpus exhaustively annotated in all languages to reliably compare the performance for pairs of languages from the same language family or different ones.

Finally, by closely analysing the results in Table 3, we noticed that given two languages L_1 and L_2, with more music genre embeddings in L_1 than in L_2 (from both ontology and corpus), the results of mapping annotations from L_1 to L_1 seems al-

ways better than the results from L_2 to L_1. This observation explains two trends in Table 3. First, the scores achieved for **en** or **es** as the source, the languages with the largest number of music genres, are the best. Second, the results for the same pair of languages could vary a lot, depending on the role each language plays, source, or target.

One possible explanation is that the prediction from languages with fewer music genre tags such as L_2 towards languages with more music genre tags such as L_1 is more challenging because the target language contains more specific or rare annotations. For instance, when checking the results per tag from **cs** to **en** we observed that among the tags with the lowest scores, we found *moombahton*, *zeuhl*, or *candombe*. However, other common music genres, such as *latin music* or *hard rock*, were also poorly predicted, showing that other causes exist too. Is the unbalanced number of music genres used in annotations a cultural consequence? Related work (Ferwerda and Schedl, 2016) seems to support this hypothesis. Then could we design a better mapping function that leverages the unbalanced numbers of music genres in cross-cultural annotations? We will dedicate a thorough investigation of these questions as future work.

6 Conclusion

We have presented an extensive investigation on cross-lingual modeling of music genre annotation, focused on six languages, and two common approaches to semantically represent concepts: ontologies and distributed embeddings[9].

Our work provides a methodological framework to study the annotation behavior across language-bound cultures in other domains too. Hence, the effectiveness of language-specific concept representations to model the culturally diverse perception could be further probed. Then, we combined the semantic representations only with retrofitting. However, inspired by paraphrastic sentence embedding learning, one can also consider the music genre relations as paraphrasing forms with different strengths (Wieting et al., 2016). Finally, the models to generate cross-lingual annotations should be thoroughly evaluated in downstream music retrieval and recommendation tasks.

---

[9] https://github.com/deezer/ CrossCulturalMusicGenrePerception

References

Javier Franco Aixelá. 1996. Culture-specific items in translation. *Translation, power, subversion*, 8:52–78.

Sanjeev Arora, Yingyu Liang, and Tengyu Ma. 2017. A simple but tough-to-beat baseline for sentence embeddings. In *International Conference on Learning Representations*.

Mikel Artetxe and Holger Schwenk. 2019. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7:597–610.

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. Dbpedia: A nucleus for a web of open data. In *The Semantic Web*, pages 722–735, Berlin, Heidelberg. Springer.

Parsiad Azimzadeh and Peter A. Forsyth. 2016. Weakly chained matrices, policy iteration, and impulse control. *SIAM Journal on Numerical Analysis*, 54(3):1341–1364.

Yoshua Bengio, Olivier Delalleau, and Nicolas Le Roux. 2006. *Label Propagation and Quadratic Criterion*, semi-supervised learning edition, pages 193–216. MIT Press.

Dmitry Bogdanov, Alastair Porter, Hendrik Schreiber, Julián Urbano, and Sergio Oramas. 2019. The acousticbrainz genre dataset: Multi-source, multi-level, multi-label, and large-scale. In *Proceedings of the 20th Conference of the International Society for Music Information Retrieval (ISMIR 2019)*, pages 360–367, Delft, The Netherlands. International Society for Music Information Retrieval (ISMIR).

Andrew P. Bradley. 1997. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159.

Alexis Conneau, Ruty Rinott, Guillaume Lample, Adina Williams, Samuel Bowman, Holger Schwenk, and Veselin Stoyanov. 2018. XNLI: Evaluating cross-lingual sentence representations. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2475–2485, Brussels, Belgium. Association for Computational Linguistics.

Jon Dattorro. 2005. *Convex Optimization & Euclidean Distance Geometry*. Meboo Publishing.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Julian Eisenschlos, Sebastian Ruder, Piotr Czapla, Marcin Kadras, Sylvain Gugger, and Jeremy Howard. 2019. MultiFiT: Efficient multi-lingual language model fine-tuning. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5702–5707, Hong Kong, China. Association for Computational Linguistics.

Irene Eleta and Jennifer Golbeck. 2012. A study of multilingual social tagging of art images: Cultural bridges and diversity. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pages 695–704, New York, NY, USA. Association for Computing Machinery.

Elena V. Epure, Anis Khlif, and Romain Hennequin. 2019. Leveraging knowledge bases and parallel annotations for music genre translation. In *Conference of the International Society of Music Information Retrieval*, ISMIR 2019.

Elena V. Epure, Guillaume Salha, and Romain Hennequin. 2020. Multilingual music genre embeddings for effective cross-lingual music item annotation. In *Conference of the International Society of Music Information Retrieval*, ISMIR 2020.

Lanting Fang, Yong Luo, Kaiyu Feng, Kaiqi Zhao, and Aiqun Hu. 2019. Knowledge-enhanced ensemble learning for word embeddings. In *The World Wide Web Conference*, WWW '19, page 427–437, New York, NY, USA. Association for Computing Machinery.

Manaal Faruqui, Jesse Dodge, Sujay Kumar Jauhar, Chris Dyer, Eduard Hovy, and Noah A. Smith. 2015. Retrofitting word vectors to semantic lexicons. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1606–1615, Denver, Colorado. Association for Computational Linguistics.

Bruce Ferwerda and Markus Schedl. 2016. Investigating the relationship between diversity in music consumption behavior and cultural dimensions: A cross-country analysis. In *24th Conference on User Modeling, Adaptation, and Personalization (UMAP) Extended Proceedings: 1st Workshop on Surprise, Opposition, and Obstruction in Adaptive and Personalized Systems (SOAP)*, Halifax, NS, Canada.

Gene H Golub and Christian Reinsch. 1971. Singular value decomposition and least squares solutions. In *Handbook for Automatic Computation: Volume II: Linear Algebra*, pages 134–151. Springer Berlin Heidelberg, Berlin, Heidelberg.

Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2018. Learning word vectors for 157 languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*, Miyazaki,

Japan. European Languages Resources Association (ELRA).

Viktor Hangya, Fabienne Braune, Alexander Fraser, and Hinrich Schütze. 2018. Two methods for domain adaptation of bilingual tasks: Delightfully simple and broadly applicable. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 810–820, Melbourne, Australia. Association for Computational Linguistics.

Dmetri Hayes. 2019. What just happened? Evaluating retrofitted distributional word vectors. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1062–1072, Minneapolis, Minnesota. Association for Computational Linguistics.

Niels van der Heijden, Samira Abnar, and Ekaterina Shutova. 2019. A comparison of architectures and pretraining methods for contextualized multilingual word embeddings. *Computing Research Repository*, arXiv:1912.10169.

Romain Hennequin, Jimena Royo-Letelier, and Manuel Moussallam. 2018. Audio based disambiguation of music genre tags. In *Conference of the International Society of Music Information Retrieval*, ISMIR 2018.

Armand Joulin, Piotr Bojanowski, Tomas Mikolov, Hervé Jégou, and Edouard Grave. 2018. Loss in translation: Learning bilingual word mapping with a retrieval criterion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2979–2984, Brussels, Belgium. Association for Computational Linguistics.

Douwe Kiela, Felix Hill, and Stephen Clark. 2015. Specializing word embeddings for similarity or relatedness. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2044–2048, Lisbon, Portugal. Association for Computational Linguistics.

Joo-Kyung Kim, Marie-Catherine de Marneffe, and Eric Fosler-Lussier. 2016. Adjusting word embeddings with semantic intensity orders. In *Proceedings of the 1st Workshop on Representation Learning for NLP*, pages 62–69, Berlin, Germany. Association for Computational Linguistics.

Claire Kramsch and HG Widdowson. 1998. *Language and culture*. Oxford University Press.

Guillaume Lample and Alexis Conneau. 2019. Cross-lingual language model pretraining. In *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, Vancouver, Canada.

Jin Ha Lee, Kahyun Choi, Xiao Hu, and J. Stephen Downie. 2013. K-pop genres: A cross-cultural exploration. In *Conference of the International Society on Music Information Retrieval*, ISMIR 2013.

Jennifer C Lena. 2012. *Banding together: How communities create genres in popular music*. Princeton University Press.

Ben Lengerich, Andrew Maas, and Christopher Potts. 2018. Retrofitting distributional embeddings to knowledge graphs with functional relations. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2423–2436, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Pasquale Lisena, Konstantin Todorov, Cécile Cecconi, Françoise Leresche, Isabelle Canno, Frédéric Puyrenier, Martine Voisin, Thierry Le Meur, and Raphäel Troncy. 2018. Controlled vocabularies for music metadata. In *Conference of the International Society on Music Information Retrieval*, Paris, France.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, page 3111–3119, Red Hook, NY, USA. Curran Associates Inc.

George A. Miller. 1995. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41.

Peter Newmark. 1988. *A textbook of translation*, volume 66. Prentice hall New York.

Matthew Peters, Waleed Ammar, Chandra Bhagavatula, and Russell Power. 2017. Semi-supervised sequence tagging with bidirectional language models. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1756–1765, Vancouver, Canada. Association for Computational Linguistics.

Ulrike Pfeil, Panayiotis Zaphiris, and Chee Siang Ang. 2006. Cultural Differences in Collaborative Authoring of Wikipedia. *Journal of Computer-Mediated Communication*, 12(1):88–113.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual bert? *Computing Research Repository*, arXiv:1906.01502.

Edoardo Maria Ponti, Ivan Vulić, Goran Glavaš, Roi Reichart, and Anna Korhonen. 2019. Cross-lingual semantic specialization via lexical relation induction. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2206–2217, Hong Kong, China. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages

3982–3992, Hong Kong, China. Association for Computational Linguistics.

Sebastian Ruder, Ivan Vulić, and Anders Søgaard. 2019. A survey of cross-lingual word embedding models. *Journal of Artificial Intelligence Research*, 65(1):569–630.

Yousef Saad. 2003. *Iterative methods for sparse linear systems*, 2nd edition. Society for Industrial and Applied Mathematics, USA.

Tanay Kumar Saha, Shafiq Joty, Naeemul Hassan, and Mohammad Al Hasan. 2016. Dis-s2v: Discourse informed sen2vec. *Computing Research Repository*, arXiv:1610.08078.

Hendrik Schreiber. 2016. Genre ontology learning: Comparing curated with crowd-sourced ontologies. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 400–406, New York, USA.

Konstantinos Sechidis, Grigorios Tsoumakas, and Ioannis Vlahavas. 2011. On the stratification of multi-label data. In *Proceedings of the 2011 European Conference on Machine Learning and Knowledge Discovery in Databases - Volume Part III*, ECML PKDD'11, page 145–158, Berlin, Heidelberg. Springer-Verlag.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany. Association for Computational Linguistics.

Vered Shwartz and Ido Dagan. 2019a. Still a pain in the neck: Evaluating text representations on lexical composition. *Transactions of the Association for Computational Linguistics*, 7:403–419.

Vered Shwartz and Ido Dagan. 2019b. Still a pain in the neck: Evaluating text representations on lexical composition. *Transactions of the Association for Computational Linguistics*, 7:403–419.

Marcin Skowron, Florian Lemmerich, Bruce Ferwerda, and Markus Schedl. 2017. Predicting genre preferences from cultural and socio-economic factors for music retrieval. In *Advances in Information Retrieval*, pages 561–567, Cham. Springer International Publishing.

Mohamed Sordo, Oscar Celma, Martin Blech, and Enric Guaus. 2008. The Quest for Musical Genres: Do the Experts and the Wisdom of Crowds Agree? In *Conference of the International Society on Music Information Retrieval*.

Robyn Speer and Joshua Chin. 2016. An ensemble method to produce high-quality word embeddings. *Computing Research Repository*, arXiv:1604.01692.

Francisco Vargas, Kamen Brestnichki, Alex Papadopoulos Korfiatis, and Nils Hammerla. 2019. Multilingual factor analysis. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1738–1750, Florence, Italy. Association for Computational Linguistics.

John Wieting, Mohit Bansal, Kevin Gimpel, and Karen Livescu. 2016. Towards universal paraphrastic sentence embeddings. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*.

Shijie Wu and Mark Dredze. 2019. Beto, bentz, becas: The surprising cross-lingual effectiveness of BERT. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 833–844, Hong Kong, China. Association for Computational Linguistics.

George Kingsley Zipf. 1949. *Human behavior and the principle of least effort*. Cambridge, Addison-Wesley Press.

A Strict Convexity of Retrofitting

Theorem. Let V be a finite vocabulary with $|V| = n$. Let $\Omega = (V, E)$ be an ontology represented as a directed graph which encodes semantic relationships between vocabulary words. Further, let $\hat{V} \subseteq V$ be the subset of words which have non-zero initial distributed representations, $\hat{\mathbf{q}}_i$. The goal of retrofitting is to learn the matrix $\mathbf{Q} \in \mathbb{R}^d$, stacking up the new embeddings $\mathbf{q}_i \in \mathbb{R}^d$ for each $i \in V$. The objective function to be minimized is:

$$\Phi(\mathbf{Q}) = \sum_{i \in \hat{V}} \alpha_i ||\mathbf{q}_i - \hat{\mathbf{q}}_i||_2^2 + \sum_{i=1}^{n} \sum_{(i,j) \in E} \beta_{ij} ||\mathbf{q}_i - \mathbf{q}_j||_2^2,$$

where the α_i and β_{ij} are positive scalars. Assuming that each connected component of Ω includes at least one word from $\hat{V}$, the objective function Φ is strictly convex w.r.t. $\mathbf{Q}$.

Proof. First of all, let $\hat{\mathbf{Q}}$ denote the $n \times d$ matrix whose i-th row corresponds to $\hat{\mathbf{q}}_i$ if $i \in \hat{V}$, and to the d-dimensional null vector $\mathbf{0}_d$ otherwise. Let $\mathbf{A}$ denote the $n \times n$ diagonal matrix verifying $\mathbf{A}_{ii} = \alpha_i$ if $i \in \hat{V}$ and $\mathbf{A}_{ii} = 0$ otherwise. Let $\mathbf{B}$ denote the $n \times n$ symmetric matrix such as, for all $i, j \in \{1, ..., n\}$ with $i \neq j$, $\mathbf{B}_{ij} = \mathbf{B}_{ji} = -\frac{1}{2}(\beta_{ij} + \beta_{ji})$ and $\mathbf{B}_{ii} = \sum_{j=1, j \neq i}^{n} |\mathbf{B}_{ij}|$. With these notations, and with $Tr(\cdot)$ the trace operator for square matrices, we have:

$$\sum_{i \in \hat{V}} \alpha_i ||\mathbf{q}_i - \hat{\mathbf{q}}_i||_2^2$$
$$= Tr\Big((\mathbf{Q} - \hat{\mathbf{Q}})^T \mathbf{A} (\mathbf{Q} - \hat{\mathbf{Q}})\Big)$$
$$= Tr\Big(\mathbf{Q}^T \mathbf{A} \mathbf{Q} - \hat{\mathbf{Q}}^T \mathbf{A} \mathbf{Q} - \mathbf{Q}^T \mathbf{A} \hat{\mathbf{Q}} + \hat{\mathbf{Q}}^T \mathbf{A} \hat{\mathbf{Q}}\Big).$$

Also:

$$\sum_{i=1}^{n} \sum_{(i,j) \in E} \beta_{ij} ||\mathbf{q}_i - \mathbf{q}_j||_2^2 = Tr\Big(\mathbf{Q}^T \mathbf{B} \mathbf{Q}\Big).$$

Therefore, as the trace is a linear mapping, we have:

$$\Phi(\mathbf{Q}) = Tr\Big(\mathbf{Q}^T (\mathbf{A} + \mathbf{B}) \mathbf{Q}\Big) + Tr\Big(\hat{\mathbf{Q}}^T \mathbf{A} \hat{\mathbf{Q}} - \hat{\mathbf{Q}}^T \mathbf{A} \mathbf{Q} - \mathbf{Q}^T \mathbf{A} \hat{\mathbf{Q}}\Big).$$

Then, we note that $\mathbf{A} + \mathbf{B}$ is a weakly diagonally dominant matrix (WDD) as, by construction, $\forall i \in \{1, ..., n\}, |(\mathbf{A} + \mathbf{B})_{ii}| \geq \sum_{j \neq i} |(\mathbf{A} + \mathbf{B})_{ij}|$. Also, for all $i \in \hat{V}$, the inequality is strict, as $|(\mathbf{A} + \mathbf{B})_{ii}| = \alpha_i + \sum_{j \neq i} |\mathbf{B}_{ij}| > \sum_{j \neq i} |(\mathbf{A} + \mathbf{B})_{ij}| = \sum_{j \neq i} |\mathbf{B}_{ij}|$, which means that, for all $i \in \hat{V}$, row i of $\mathbf{A} + \mathbf{B}$ is strictly diagonally dominant (SSD). Assuming that each connected component of graph G includes at least one node from $\hat{V}$, we conclude that $\mathbf{A} + \mathbf{B}$ is a weakly chained diagonally dominant matrix (Azimzadeh and Forsyth, 2016), *i.e.* that:

- $\mathbf{A} + \mathbf{B}$ is WDD;

- for each $i \in V$ such that row i is not SSD, there exists a *walk* in the graph whose adjacency matrix is $\mathbf{A} + \mathbf{B}$ (two nodes i and j are connected if $(\mathbf{A} + \mathbf{B})_{ij} = (\mathbf{A} + \mathbf{B})_{ji} \neq 0$), starting from i and ending at a node associated to a SSD row.

Such matrices are nonsingular (Azimzadeh and Forsyth, 2016), which implies that $\mathbf{Q} \to \mathbf{Q}^T (\mathbf{A} + \mathbf{B}) \mathbf{Q}$ is a positive-definite quadratic form. As $\mathbf{A} + \mathbf{B}$ is a symmetric positive-definite matrix, there exists a matrix $\mathbf{M}$ such that $\mathbf{A} + \mathbf{B} = \mathbf{M}^T \mathbf{M}$. Therefore, denoting $|| \cdot ||_F^2$ the squared Frobenius matrix norm:

$$Tr\Big(\mathbf{Q}^T (\mathbf{A} + \mathbf{B}) \mathbf{Q}\Big)$$
$$= Tr\Big(\mathbf{Q}^T \mathbf{M}^T \mathbf{M} \mathbf{Q}\Big) = ||\mathbf{Q} \mathbf{M}||_F^2$$

which is strictly convex w.r.t. $\mathbf{Q}$ due to the strict convexity of the squared Frobenius norm (see *e.g.* 3.1 in Dattorro (2005)). Since the sum of strictly convex functions of $\mathbf{Q}$ (first trace in $\Phi(\mathbf{Q})$) and linear functions of $\mathbf{Q}$ (second trace in $\Phi(\mathbf{Q})$) is still strictly convex w.r.t. $\mathbf{Q}$, we conclude that the objective function Φ is strictly convex w.r.t. $\mathbf{Q}$.

Corollary. *The retrofitting update procedure is insensitive to the order in which nodes are updated.*

The aforementioned updating procedure for $\mathbf{Q}$ (Faruqui et al., 2015) is derived from Jacobi iteration procedure (Saad, 2003; Bengio et al., 2006) and converges for any initialization. Such a convergence result is discussed in Bengio et al. (2006). It can also be directly verified in our specific setting by checking that each irreducible element of $\mathbf{A} + \mathbf{B}$, *i.e.* each connected component of the underlying graph constructed from this matrix, is irreducibly diagonally dominant (see 4.2.3 in Saad (2003)) and then by applying Theorem 4.9 from Saad (2003) on each of these components. Besides, due to its strict convexity w.r.t. $\mathbf{Q}$, the objective function Φ admits a unique global minimum. Consequently, the retrofitting update procedure will converge to the same embedding matrix regardless of the order in which nodes are updated.

B Extended Results

The following Tables 5 and 6 provide more complete results from our experiments.

Pair	$\mathbf{FT}_{avg}$	$\mathbf{XLM}_{avg}$	$\mathbf{mBERT}_{avg}$	$\mathbf{FT}_{sif}$	$\mathbf{XLM}_{sif}$	$\mathbf{mBERT}_{sif}$	$\mathbf{XLM}_{ctxt}$	$\mathbf{mBERT}_{ctxt}$
en-nl	75.2 ± 0.2	83.7 ± 0.1	86.2 ± 0.2	86.5 ± 0.1	86.6 ± 0.2	88.3 ± 0.2	85.4 ± 0.3	85.1 ± 0.2
en-fr	78.6 ± 0.2	84.4 ± 0.2	85.2 ± 0.3	87.4 ± 0.3	87.7 ± 0.3	87.6 ± 0.2	86.6 ± 0.2	84.9 ± 0.3
en-es	77.5 ± 0.1	82.5 ± 0.1	83.8 ± 0.1	86.9 ± 0.2	86.7 ± 0.2	86.2 ± 0.2	85.4 ± 0.3	84.7 ± 0.2
en-cs	74.5 ± 0.6	87.2 ± 0.6	88.2 ± 0.4	88.6 ± 0.4	90.6 ± 0.5	90.6 ± 0.3	89.0 ± 0.5	87.2 ± 0.1
en-ja	69.6 ± 0.5	70.9 ± 0.1	72.4 ± 0.2	80.8 ± 0.3	76.1 ± 0.3	70.5 ± 0.3	74.2 ± 0.3	68.9 ± 0.2
nl-en	73.8 ± 0.5	71.9 ± 0.4	74.6 ± 0.3	79.8 ± 0.4	77.7 ± 0.4	78.2 ± 0.3	74.2 ± 0.7	73.5 ± 0.5
nl-fr	63.9 ± 0.5	72.7 ± 0.9	76.1 ± 0.3	79.3 ± 0.8	78.4 ± 1.0	79.5 ± 0.3	74.9 ± 1.0	75.1 ± 0.7
nl-es	63.7 ± 0.4	71.6 ± 0.6	75.0 ± 0.4	77.7 ± 0.5	76.8 ± 0.3	77.6 ± 0.6	73.5 ± 0.3	75.0 ± 0.3
nl-cs	65.1 ± 0.3	77.4 ± 0.4	80.8 ± 0.4	80.6 ± 0.2	82.0 ± 0.4	83.2 ± 0.4	79.0 ± 0.4	79.6 ± 0.2
nl-ja	64.8 ± 0.2	65.2 ± 0.8	65.8 ± 2.5	74.9 ± 1.0	70.1 ± 0.4	67.6 ± 0.2	66.0 ± 0.6	65.0 ± 0.2
fr-nl	67.7 ± 1.0	77.5 ± 0.4	79.4 ± 0.1	81.9 ± 0.4	81.6 ± 0.3	82.0 ± 0.4	78.6 ± 0.1	77.7 ± 0.2
fr-en	76.2 ± 0.2	74.7 ± 0.6	77.2 ± 0.5	83.0 ± 0.2	81.5 ± 0.3	80.8 ± 0.1	78.8 ± 0.5	77.2 ± 0.5
fr-es	71.0 ± 0.2	75.6 ± 0.3	77.5 ± 0.4	81.8 ± 0.3	81.0 ± 0.5	80.0 ± 0.4	78.7 ± 0.5	78.2 ± 0.3
fr-cs	70.4 ± 0.8	80.1 ± 0.4	82.7 ± 0.6	83.9 ± 0.4	85.3 ± 0.3	85.5 ± 0.5	83.1 ± 0.2	82.0 ± 0.4
fr-ja	71.1 ± 0.3	67.4 ± 0.5	61.5 ± 0.2	77.9 ± 0.1	73.1 ± 0.1	66.9 ± 0.2	69.5 ± 0.3	64.8 ± 0.5
es-nl	68.5 ± 0.5	80.3 ± 0.6	82.3 ± 0.3	82.8 ± 0.9	83.8 ± 0.7	84.8 ± 0.2	81.3 ± 0.8	80.7 ± 0.4
es-fr	70.8 ± 0.4	79.9 ± 0.5	81.0 ± 0.3	85.0 ± 0.3	84.6 ± 0.4	84.5 ± 0.4	82.2 ± 0.5	80.3 ± 0.4
es-en	75.3 ± 0.1	76.4 ± 0.2	78.4 ± 0.3	84.7 ± 0.2	83.2 ± 0.4	82.9 ± 0.3	79.5 ± 0.2	78.1 ± 0.4
es-cs	68.9 ± 0.8	83.2 ± 0.7	85.5 ± 0.5	85.6 ± 0.6	88.2 ± 0.7	88.6 ± 0.6	85.9 ± 0.5	84.2 ± 0.4
es-ja	65.2 ± 0.4	70.4 ± 0.2	67.7 ± 0.1	78.3 ± 0.6	74.4 ± 0.2	68.6 ± 0.5	72.8 ± 0.8	65.1 ± 0.6
cs-nl	68.5 ± 1.0	75.7 ± 0.9	78.5 ± 0.7	78.3 ± 0.9	80.7 ± 0.9	80.5 ± 1.0	78.1 ± 0.4	76.8 ± 0.3
cs-fr	64.5 ± 0.9	74.7 ± 0.9	77.1 ± 1.0	78.5 ± 0.2	80.7 ± 0.3	80.7 ± 0.9	79.9 ± 0.7	76.2 ± 1.2
cs-es	65.3 ± 0.9	73.8 ± 0.7	75.8 ± 0.3	77.7 ± 0.8	79.8 ± 0.2	78.7 ± 0.3	78.8 ± 0.4	75.9 ± 0.8
cs-en	70.3 ± 0.5	70.9 ± 0.0	74.2 ± 0.2	78.9 ± 0.1	78.9 ± 0.5	79.2 ± 0.5	78.3 ± 0.4	74.0 ± 0.3
cs-ja	67.1 ± 1.1	70.8 ± 0.2	65.8 ± 1.4	76.9 ± 0.1	72.7 ± 0.3	67.3 ± 0.7	72.2 ± 1.1	68.3 ± 0.1
ja-nl	62.0 ± 0.5	59.5 ± 1.0	62.1 ± 0.4	72.8 ± 1.0	68.0 ± 0.6	67.1 ± 0.2	63.6 ± 0.9	52.4 ± 0.3
ja-fr	66.0 ± 1.1	47.6 ± 0.2	50.6 ± 0.2	73.7 ± 0.6	69.4 ± 0.7	65.6 ± 0.3	58.4 ± 0.4	43.9 ± 0.7
ja-es	63.2 ± 0.3	53.6 ± 0.3	56.1 ± 0.4	73.9 ± 0.4	67.5 ± 1.0	63.6 ± 0.6	60.0 ± 0.2	48.7 ± 0.6
ja-cs	61.7 ± 1.3	64.8 ± 0.8	64.7 ± 0.6	77.5 ± 0.2	73.3 ± 0.6	69.2 ± 0.4	64.5 ± 0.5	56.8 ± 0.9
ja-en	72.1 ± 1.0	46.3 ± 0.4	49.0 ± 0.1	75.6 ± 0.3	66.8 ± 0.6	64.1 ± 0.7	56.8 ± 1.0	44.0 ± 0.2

Table 5: Macro-AUC scores (in %, best locally underlined). The first two parts correspond to averaging or applying *sif* averaging to static multilingual word embeddings; the third part corresponds to the contextual sentence embeddings.

Pair	$\mathbf{DBp}_{a\Omega}\mathbf{NNDist}$	$\mathbf{Rfit}_{u\Omega}\mathbf{FT}_{sif}$	$\mathbf{Rfit}_{a\Omega}^{en}\mathbf{FT}_{sif}$
en-nl	83.7 ± 0.1	90.0 ± 0.1	$\mathbf{90.7 \pm 0.0}$
en-fr	82.7 ± 0.3	90.8 ± 0.2	$\mathbf{91.7 \pm 0.1}$
en-es	81.1 ± 0.3	89.9 ± 0.1	$\mathbf{91.4 \pm 0.2}$
en-cs	86.6 ± 0.3	90.4 ± 0.3	$\mathbf{91.6 \pm 0.4}$
en-ja	81.3 ± 0.1	86.7 ± 0.3	$\mathbf{89.1 \pm 0.2}$
Pair	$\mathbf{DBp}_{a\Omega}\mathbf{NNDist}$	$\mathbf{Rfit}_{u\Omega}\mathbf{FT}_{sif}$	$\mathbf{Rfit}_{a\Omega}^{nl}\mathbf{FT}_{sif}$
nl-en	72.4 ± 0.7	84.3 ± 0.1	$\mathbf{86.9 \pm 0.1}$
nl-fr	72.9 ± 0.1	81.5 ± 0.7	$\mathbf{85.0 \pm 0.7}$
nl-es	69.2 ± 0.1	80.5 ± 0.4	$\mathbf{83.3 \pm 0.2}$
nl-cs	68.0 ± 0.5	$\mathbf{83.4 \pm 0.5}$	$\mathbf{83.4 \pm 0.5}$
nl-ja	72.3 ± 0.4	$\mathbf{80.0 \pm 0.7}$	$\mathbf{81.6 \pm 1.1}$
Pair	$\mathbf{DBp}_{a\Omega}\mathbf{NNDist}$	$\mathbf{Rfit}_{u\Omega}\mathbf{FT}_{sif}$	$\mathbf{Rfit}_{a\Omega}^{fr}\mathbf{FT}_{sif}$
fr-nl	75.8 ± 0.1	84.7 ± 0.3	$\mathbf{87.5 \pm 0.3}$
fr-en	74.5 ± 0.4	87.7 ± 0.1	$\mathbf{89.1 \pm 0.2}$
fr-es	72.1 ± 0.3	85.3 ± 0.2	$\mathbf{86.4 \pm 0.2}$
fr-cs	82.0 ± 0.8	87.2 ± 0.3	$\mathbf{89.6 \pm 0.3}$
fr-ja	76.7 ± 0.6	81.4 ± 0.3	$\mathbf{84.0 \pm 0.4}$
Pair	$\mathbf{DBp}_{a\Omega}\mathbf{NNDist}$	$\mathbf{Rfit}_{u\Omega}\mathbf{FT}_{sif}$	$\mathbf{Rfit}_{a\Omega}^{es}\mathbf{FT}_{sif}$
es-nl	78.3 ± 0.9	85.9 ± 0.6	$\mathbf{88.6 \pm 0.5}$
es-fr	78.5 ± 0.4	87.4 ± 0.3	$\mathbf{89.0 \pm 0.3}$
es-en	76.0 ± 0.1	88.8 ± 0.3	$\mathbf{90.6 \pm 0.5}$
es-cs	82.5 ± 0.9	88.0 ± 0.4	$\mathbf{90.4 \pm 0.3}$
es-ja	77.8 ± 0.7	83.1 ± 0.6	$\mathbf{84.7 \pm 1.0}$
Pair	$\mathbf{DBp}_{a\Omega}\mathbf{NNDist}$	$\mathbf{Rfit}_{u\Omega}\mathbf{FT}_{sif}$	$\mathbf{Rfit}_{a\Omega}^{cs}\mathbf{FT}_{sif}$
cs-nl	65.1 ± 0.2	$\mathbf{81.1 \pm 1.2}$	$\mathbf{81.1 \pm 1.2}$
cs-fr	75.4 ± 0.5	81.4 ± 0.3	$\mathbf{85.1 \pm 0.6}$
cs-es	72.6 ± 0.6	81.6 ± 0.9	$\mathbf{84.4 \pm 1.0}$
cs-en	75.6 ± 0.5	84.5 ± 0.4	$\mathbf{88.2 \pm 0.3}$
cs-ja	77.1 ± 0.9	80.5 ± 0.5	$\mathbf{84.5 \pm 0.7}$
Pair	$\mathbf{DBp}_{a\Omega}\mathbf{NNDist}$	$\mathbf{Rfit}_{u\Omega}\mathbf{FT}_{sif}$	$\mathbf{Rfit}_{a\Omega}^{ja}\mathbf{FT}_{sif}$
ja-nl	68.5 ± 0.7	$\mathbf{76.9 \pm 0.3}$	75.7 ± 0.3
ja-fr	71.9 ± 0.1	$\mathbf{77.8 \pm 0.1}$	76.7 ± 0.5
ja-es	68.9 ± 0.4	$\mathbf{78.8 \pm 0.5}$	76.1 ± 0.5
ja-cs	77.9 ± 0.8	80.7 ± 0.4	$\mathbf{82.2 \pm 0.8}$
ja-en	70.4 ± 0.4	$\mathbf{81.6 \pm 0.8}$	$\mathbf{82.3 \pm 0.5}$

Table 6: Macro-AUC scores (in %) with vectors learned by leveraging aligned monolingual ontologies. The first column shows the results by relying on the aligned ontologies only. The second and third columns show the results obtained by retrofitting $\mathbf{FT}_{sif}$ embeddings to monolingual ontologies, respectively, aligned monolingual ontologies.

Embedding Model	Dimension
FT	300
mBERT	768
XLM	2048
LASER	1024

Table 7: Embedding dimensions for pre-trained models used in our study, corresponding to the values provided and optimized by each model's authors.