some crucial issues with prior efforts focused on the creation of multi-lingual semantic resources, namely: i) limited coverage; ii) heterogeneous annotation guidelines; and iii) concept pairs which are semantically incomparable across different languages.

3) We offer to the community manually annotated evaluation sets of 1,888 concept pairs across 12 typologically diverse languages, and 66 large cross-lingual evaluation sets. To the best of our knowledge, Multi-SimLex is the most comprehensive evaluation resource to date focused on the relation of semantic similarity.

4) We benchmark a wide array of recent state-of-the-art monolingual and cross-lingual word representation models across our sample of languages. The results can serve as strong baselines that lay the foundation for future improvements.

5) We present a first large-scale evaluation study on the ability of encoders pretrained on language modeling (such as BERT (Devlin et al. 2019) and XLM (Conneau and Lample 2019)) to reason over word-level semantic similarity in different languages. To our own surprise, the results show that monolingual pretrained encoders, even when presented with word types out of context, are sometimes competitive with static word embedding models such as fastText (Bojanowski et al. 2017) or word2vec (Mikolov et al. 2013). The results also reveal a huge gap in performance between massively multilingual pretrained encoders and language-specific encoders in favor of the latter: our findings support other recent empirical evidence related to the "curse of multilinguality" (Conneau et al. 2019; Bapna and Firat 2019) in representation learning.

6) We make all of these resources available on a website which facilitates easy creation, submission and sharing of Multi-Simlex-style datasets for a larger number of languages. We hope that this will yield an even larger repository of semantic resources that inspire future advances in NLP within and across languages.

In light of the success of Universal Dependencies (Nivre et al. 2019), we hope that our initiative will instigate a collaborative public effort with established and clear-cut guidelines that will result in additional Multi-SimLex datasets in a large number of languages in the near future. Moreover, we hope that it will provide means to advance our understanding of distributional and lexical semantics across a large number of languages. All monolingual and cross-lingual Multi-SimLex datasets–along with detailed translation and annotation guidelines–are available online at: https://multisimlex.com/.

2. Lexical Semantic Similarity

2.1 Similarity and Association

The focus of the Multi-SimLex initiative is on the lexical relation of pure *semantic similarity*. For any pair of words, this relation measures whether their referents share the same features. For instance, *graffiti* and *frescos* are similar to the extent that they are both forms of painting and appear on walls. This relation can be contrasted with the cognitive *association* between two words, which often depends on how much their referents interact in the real world, or are found in the same situations. For instance, a *painter* is easily associated with *frescos*, although they lack any physical commonalities. Association is also known in the literature under other names: relatedness (Budanitsky and Hirst 2006), topical similarity (McKeown et al. 2002), and domain similarity (Turney 2012).

Semantic similarity and association overlap to some degree, but do not coincide (Kiela, Hill, and Clark 2015; Vulić, Kiela, and Korhonen 2017). In fact, there exist plenty of pairs that are intuitively associated but not similar. Pairs where the converse is true can