Morphological Paradigm Completion with LSTM Neural Models

Conor Stuart Roe

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Arts in Linguistics and Computer Science

October 25, 2019

Version: Draft 1

Advisor: Jane Chandlee

Contents

1 Introduction

In linguistics, morphology refers to alterations to words to reflect changes in meaning or grammatical category. For example, English verbs have differing morphological forms to indicate the simple present and simple past tenses, e.g., *show* → *showed*, *see* → *saw*, etc. (Dreyer et al., 2008). Grammatical inflection in particular has a tendency to be structured into paradigms - sets of all possible morphological forms that words of a certain type can take on, often shown arrayed in tables. The table below gives part of the paradigm for the verb *to see*:

	simple		progressive	
	3rd singular	3rd plural	3rd singular	3rd plural
present	(she) **sees**	(they) **see**	(she) **is seeing**	(they) **are seeing**
past	(she) **saw**	(they) **saw**	(she) **was seeing**	(they) **were seeing**

Historically, in language technologies and modeling, morphology has been somewhat under-emphasized. This is probably due at least in part to the dominance of English in language technology research, and its below-average morphological complexity (Cotterell and Heigold, 2017). English lexemes tend to have few grammatically inflected forms compared to in most other languages. This means that in machine learning models which are given an English training corpus and then tested on new material, the occurrence of *out-of-vocabulary* (OOV) inflected forms - that is, forms in the test data that never occur in the training data - are less frequent.

Moreover, as can be seen above, many English grammatical forms are periphrastic, using several words to construct a single inflected form. This means that even with grammatical information removed via lemmatization, the process of reducing words to their citation form, an English sentence may still be fairly understandable. (The *citation form* of a lexeme is the form under which it would appear in a dictionary: *see*, *sees*, *seeing*, and *saw* all belong to the same lexeme, with the citation form *see*.)

For instance, given a sentence *she have see it* whose words have been lemmatized, a proficient reader of English can guess that the original sentence was *she has seen it* or *she had seen it*; not much information is lost.

However, in many languages grammatical inflection carries much more semantic burden, and the number of possible inflected forms can be much greater. For instance, a single verb in the Archi language can be inflected in 1,725 ways (**Kibrik1994**). For such highly inflected languages, data sparsity is a significant problem for language models naive to morphology. In languages with a high number of possible forms per lexeme, a much larger number OOV forms are inevitably encountered in test data, requiring reliance on a model's representation of OOV words. Even inflected forms that do occur in training data may not appear often enough for a naive model to understand their semantic content. A model that considers grammatical categories separately is better able to understand the semantic content of inflected words (Cotterell, Kirov, Sylak-Glassman, Yarowsky, et al., 2016). It has been empirically shown that comprehending grammatical categories and inflection improves accuracy rates in language modeling and machine translation (Faruqui et al., 2015).

To this end, an ongoing body of research in the last few years has attempted to apply machine learning to various tasks related to morphological analysis and generation. Very high accuracy rates have been achieved in tasks related to predicting grammatical inflection for many languages, but gaps in current capabilities exist.

2 Literature Review

2.1 Machine learning of morphology: sub-problems and related problems

Within the realm of machine understanding of morphology, there are many sub-problems and related problems. The most basic areas of research involve predicting the morphology of words in isolation - transforming a word into a specific morphological form, or the inverse, tagging a form with its morphological categories. Typically, this has been done with words in regular grammatical paradigms, such as number and case marking on nouns, or tense, aspect, mood, and/or argument agreement patterns on verbs.

2.1.1 Core supervised learning problems

Some of the earliest work in computational morphology involves making specific morphological transformations. That is, given a particular form of a lexeme (often, but not necessarily, a citation form), predicting another form. An example would be learning to transform English verbs from present to past tense, e.g., *show* → *showed*, *see* → *saw*, etc. (Dreyer et al., 2008).

The natural extension of this is aiming to be able to predict any inflected form given one specific form of a lexeme and an arbitrary set of morphological categories. For instance, given a lexeme *see* and the categories `3rd person singular, simple present`, generating the correct form *sees*. Generally speaking, a citation form has been used as input (Durrett and DeNero, 2013) (Faruqui et al., 2015) (Cotterell and Heigold, 2017). The related "reinflection" problem involves being given any inflected form as input, and transforming it into any other (Cotterell, Kirov, Sylak-Glassman, Yarowsky, et al., 2016).

A further extension of the morphology generation problem is the generation of complete inflection tables. The exact nature of this problem depends on the type of training and test data used. If a model is only trained on a sparse, random sampling of forms for each lexeme, then a task may consist of filling out the rest of an inflection table for those lexemes. For instance, a model may be given the forms *sees* and *seeing* among its training data, and be required to fill out the remaining forms of that paradigm, including *see* and *saw*.

If a model is instead trained using entire inflection tables, e.g., all forms of the verb *see*, then test data must consist of new lexemes. Another dimension along which paradigm completion tasks differ is whether only a single citation form is provided as a prompt, or whether any form or even multiple forms may be provided as a prompt for producing a single table (Hulden et al., 2014) (Ahlberg et al., 2015) (Cotterell and Heigold, 2017).

2.1.2 Inflection types

Overall, a diverse set of inflection shapes have been worked with in the most recent efforts of this subfield. Since 2016, the Special Interest Group on Computational Morphology and Phonology (SIGMORPHON), a research collective focused on computational morphology and related problems, has fielded "shared tasks" in which several research teams globally are given training data and a task definition, and attempt to create models which are subsequently compared. The SIGMORPHON shared task 2018 included training data from 103 typologically diverse languages, and paradigms using suffixing, prefixing, infixing, reduplication, and non-concatenative morphology.

However, all SIGMORPHON work has been with well-defined, tabular paradigms of grammatical inflection of fairly moderate size (≤ 200 forms) (Cotterell, Kirov, Sylak-Glassman, Walther, et al., 2018). Other types of morphology such as derivational morphology and cliticization, which have the potential to greatly expand the number of possible forms and in a less organized fashion, have been less well explored. What Mattissen, 2004 calls "compositionally polysynthetic" languages, those which make extensive use of more expansive types of word building including incorporation and productive derivational morphology, are ripe for future work (Cotterell, Kirov, Sylak-Glassman, Yarowsky, et al., 2016). Such languages include Chukchi, Cherokee, Ainu, Nahuatl, and others (Mattissen, 2004).

The chief challenges with approaching these less paradigmatic types of inflection are data procurement and structuring. For grammatical inflection paradigms, the set of potential forms is simply combinatorial from the set of grammatical categories, and can be specifically enumerated. Concretely, the vast majority of grammatical paradigm data used in SIGMORPHON comes from Wiktionary, a free and collaborative online multilingual dictionary which provides full or partial tables of grammatical inflection along with lexeme definitions; these tables define the space of inflected forms for the task. In contrast, with derivational morphology, noun incorporation, and other less paradigmatic types of inflection, the possibility space is considerably less well-defined. How can a model know a priori that *likeable* is a valid word of English but *hateable* is not, or that the opposite of *accessible* is *inaccessible* and not *unaccessable*? There are no tables for the set of, e.g., derivational forms that a lexeme may take on, since it is constrained by semantics and idiosyncratic usage, and there is considerably more variability among lexemes of the same category (say, verbs) in which derivational categories they may take on than which grammatical categories. Consequently, labeled data is much harder to come by, and it is harder to demonstrate that a dataset of derivational forms is truly exhaustive.

2.1.3 Related problems

Within only the last two or so years, there has been work on predicting morphology in context. In the 2018 and 2019 SIGMORPHON shared tasks, to which several teams of researchers submitted solutions, a sub-task was dedicated to cloze challenges, a type of test in which one word in a sentence, given in citation form, was to be inflected based on context (Cotterell, Kirov, Sylak-Glassman, Walther, et al., 2018) (McCarthy et al., 2019). This work is essentially a synthesis of morphology generation and morphosyntactic modeling.

Since 2017, there has been some work done on learning curves for computational morphology. The datasets published for SIGMORPHON 2017 and 2018 include partitions into low ($\sim$100 forms), medium ($\sim$1000 forms), and high ($\sim$10,000 forms) data training sets for the express purpose of assessing the learning curve of different models. Evidence suggests that learning curve varies by model type; LSTM neural models, a type of model generally considered to be state of the art, despite being the most successful models with high-data training sets, often fare worse than more baseline string transduction models with low data training sets (Cotterell and Heigold, 2017) (Cotterell, Kirov, Sylak-Glassman, Walther, et al., 2018). Improving performance with small training sets is of interest, as much of the applicability of computational morphology models is to languages which don't already have high-quality technical tools or datasets.

The most recent new challenge that SIGMORPHON has tried to address is that of transfer learning of morphology, in the shared task earlier this year. Given a state of the art model trained on a language with a high volume of training data, teams were asked to alter it into a model that would perform well on a new language, given a smaller amount of training data for that language. 80% of the pairs of languages were closely related, while 20% were distantly or not at all related. Gains of transfer learning models between closely related languages generally performed better than transfer learning models between more distant languages (McCarthy et al., 2019).

2.2 Non-neural approaches

2.2.1 Vector embedding

A technique that has found success in a variety of computational linguistics tasks is that of representing words in relatively low-dimensional vector spaces. That is, words are represented as a vector, a series of numbers of fixed length; the length of the vector is typically much smaller than the number of total known words.

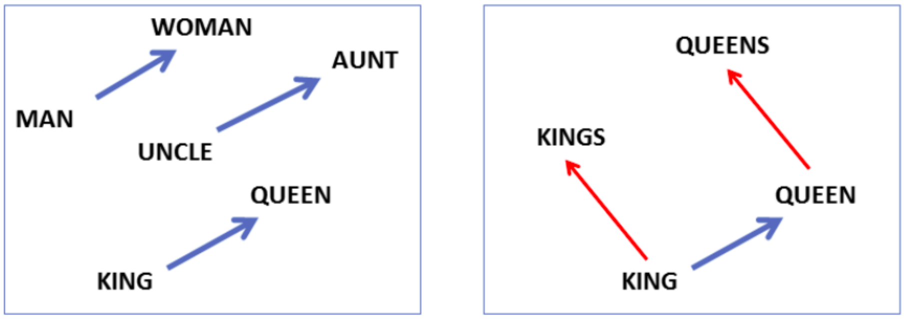

Fig. 2.1: Regular spatial transformations encode semantic or grammatical content (Mikolov et al., 2013)

This has the intention of capturing semantic and syntactic content in a principled way - similarity between the number of two words are expected to signify actual similarity in their meaning, and regular linear transformations between vectors should roughly correspond to specific semantic or grammatical changes. These vector representations can be generated via various unsupervised learning methods (Bilmes and Kirchhoff, 2003) (Alexandrescu and Kirchhoff, 2006).

Regularities in the relative location of semantically related words have been exploited for semantic analysis tasks (Alexandrescu and Kirchhoff, 2006). Similarly, morphological changes may appear as spatial transformations in vector space, and work has been done on discovering morphological relationships between in-vocabulary words based on their relative spatial locations (Mikolov et al., 2013) (Soricut and Och, 2015) (Dos Santos and Zadrozny, 2014). Figure 2.1 illustrates this idea in a slightly simplified way: once a vector embedding model has been trained on English, it can be discovered that semantic transformations (male to female) and grammatical transformations (singular to plural) roughly correspond to regular spatial translations in vector space.

Vector embedding has the limitation that it cannot extend to words for which a vector representation has not been trained, and so it cannot directly provide understanding of the many OOV forms encountered in test data of highly inflected languages (Soricut and Och, 2015) (Cotterell and Schütze, 2019). However, it can be a means to discover relationships between words in an unsupervised manner, which may support labeling tasks in support of computational morphology and other tasks.

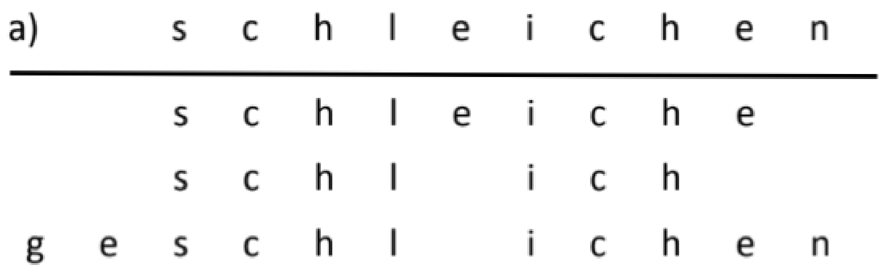

Fig. 2.2: Character alignment for various forms of the German verb *schleichen* (Nicolai et al., 2015).

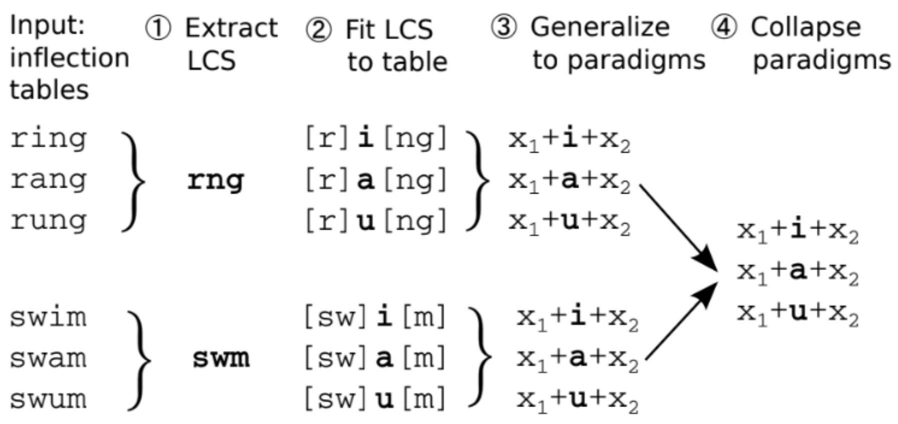

Fig. 2.3: Conceptual depiction of a typical example of a string transduction method of morphology learning (Hulden et al., 2014).

2.2.2 String transduction

Earlier work specifically focused on the problem of morphology prediction made use of iteratively improving methods of string transduction - in essence, pattern matching on the written representations of words (Durrett and DeNero, 2013) (Hulden et al., 2014) (Nicolai et al., 2015) (Ahlberg et al., 2015). Typical steps of string transduction methods include character alignment (depicted in Figure 2.2), identification of characters that are inserted or deleted based on grammatical form, and generalization of lexemes which are inflected by the same sets of insertions or deletions (depicted in Figure 2.3).

A crucial limitation of string transduction methods are their general assumption that most lexemes have exactly the same set of characterwise transformations as a large group of other lexemes, and that a manageably small number of such inflection classes exist. There are paradigms with such a limited set of inflection classes, such as Spanish *-ar*, *-er*, and *-ir* verbs. However, when multiple morpholonological processes are at play, individual lexemes may be nearly unique in their exact set of transformations.

For example, Finnish noun declension has processes of vowel harmony, consonant gradation, and vowel alternation and lengthening operating to produce final inflected forms (Ranta, 2008). As an illustration, consider the Finnish nouns *puku* "suit" and *kenkä* "shoe", which have the inessive singular forms *puvussa* "in the suit" and *kengässä* "in the shoe", respectively. In both forms, the letter *k* is transformed via consonant gradation, but the letter it becomes depends on the surrounding letters. The final vowel of the forms may be *a* or *ä*, depending on vowel harmony. In other inflected forms, the final vowel of the words may be doubled (*English Wiktionary* n.d.). A model that naively seeks to match these words with other words using the same set of character transformations across the paradigm may need to assign nearly every word to its own category, failing to generalize the patterns at work.

The poorer performance of string transduction relative to neural models has led to a move of the field away from string transduction since about 2016 (Cotterell, Kirov, Sylak-Glassman, Walther, et al., 2018).

2.3 LSTM and other neural approaches

Since 2016, almost all work on paradigm completion has made use of long short-term memory (LSTM) or related gated recurrent network (GRU) models (Faruqui et al., 2015) (Cotterell, Kirov, Sylak-Glassman, Yarowsky, et al., 2016) (Cotterell and Heigold, 2017) (Cotterell, Kirov, Sylak-Glassman, Walther, et al., 2018) (McCarthy et al., 2019). An LSTM network is a variation on a recurrent neural network (RNN), a variant of neural network (**Hochreiter1997**).

Neural networks are a type of function approximation model inspired by the connection of neurons in animal brains. They have come to be implemented in a variety of forms, and underlie state of the art machine learning models in a variety of applications. The common feature of neural models is repeated matrix multiplication

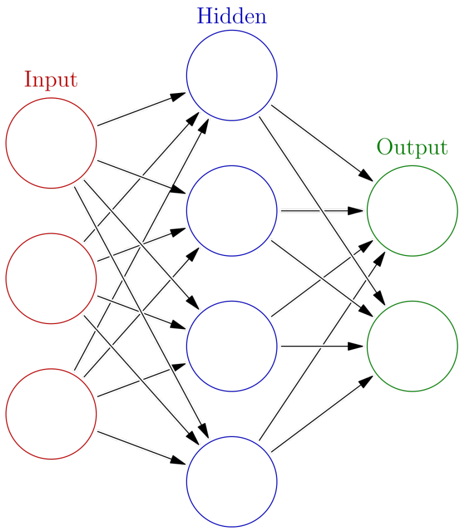

Fig. 2.4: The structure of a feed-forward neural network with three inputs, four hidden cells, and two outputs. (commons.wikimedia.org/wiki/File:Colored_neural_network.svg)

followed by the application of a non-linear "activation" function. Figure 2.4 illustrates a simple feed-forward network, in which a vector of three inputs is multiplied by some 3×4 matrix and activated to produce a vector of four intermediate values, which are again multiplied by some 4×2 matrix and activated to produce a vector of two outputs.

An RNN is a type of neural network that operates over sequences of inputs, typically with unknown length. Figure 2.5 illustrates the general structure: each input is represented by a vector, which is multiplied by a vector U to modify state, and the modified state is then multiplied by a vector W to produce an output vector and a matrix V to produce the next state. An LSTM network utilizes a specific means of using the inputs to modify state, represented by Figure 2.6, which includes the ability to modify the rate at which information is forgotten. LSTMs are intended to solve the forgetfulness of earlier RNN types, which tend to be unable to recall information from more than a few iterations prior (**Hochreiter1997**).

LSTMs have become the dominant model type in a variety of language tasks, including syntactic and morphological tasks. They significantly outperformed other types of models in SIGMORPHON 2016, since which time they have come to underlie nearly

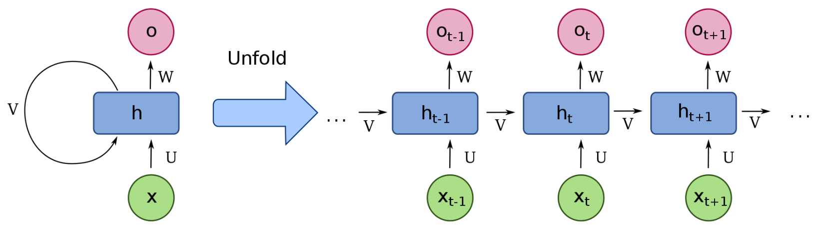

Fig. 2.5: The generalized structure of a recurrent neural network. (commons.wikimedia.org/wiki/File:Recurrent_neural_network_unfold.svg)

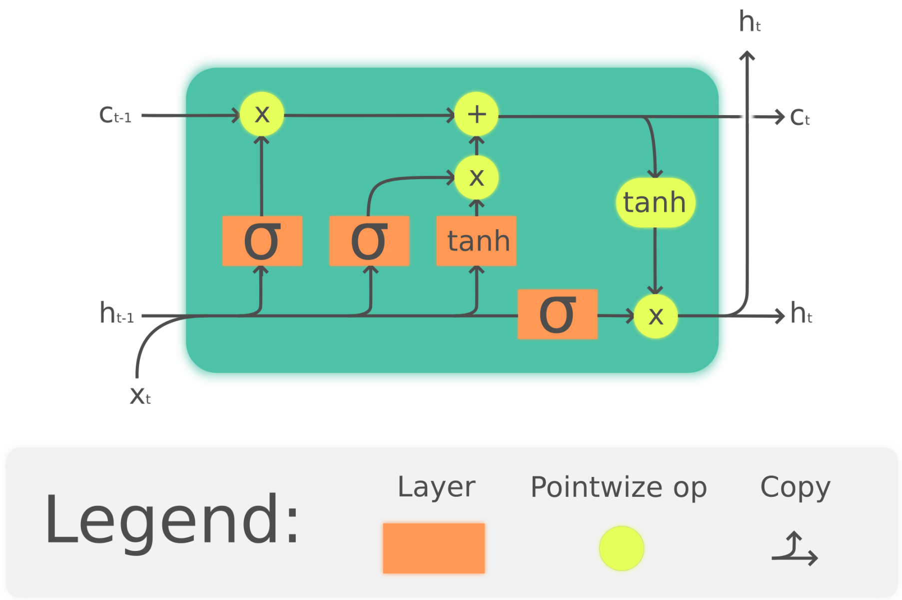

Fig. 2.6: The state cell of an LSTM model. (commons.wikimedia.org/wiki/File:The_LSTM_cell.png)

all morphology prediction models (Cotterell, Kirov, Sylak-Glassman, Yarowsky, et al., 2016) (Cotterell and Heigold, 2017).

Data

3

3.1 Structured paradigm data

The English Wiktionary, a collaborative online dictionary, has become something of a standard source of supervised morphological data. It provides full or partial inflection tables alongside lexeme definitions; the structure of tables is consistent for a given language and part of speech. An example table is given in figure 3.1. For some highly inflected languages (e.g., Navajo), Wiktionary only provides a fixed subset of forms. For some relationships between words that could be considered grammatical, it may simply offer them as separate lexical entries; for example, Russian perfect and imperfect forms are given as separate entries, as are Navajo verb forms that vary by aspect or thematic classifier (*English Wiktionary* n.d.).

Specific Wiktionary dumps have come to be used as standard datasets, used by multiple authors as a benchmark to compare model performance. Durrett and DeNero, 2013 published a Wiktionary dataset of five paradigms from three languages (Finnish, Spanish, and German) which have been used in much future work (Hulden et al., 2014) (Nicolai et al., 2015) (Ahlberg et al., 2015) (Faruqui et al., 2015).

SIGMORPHON 2017 published a dataset partitioned into three training levels (100, 1000, and 10,000 tables), containing both sparse and full inflection tables for one or more parts of speech for each of 52 languages. Most of that data was derived from a January 2017 Wiktionary dump. (Cotterell and Heigold, 2017)

The most complete structured dataset to date was published for the SIGMORPHON shared task 2018, a superset of the SIGMORPHON 2017 data, which is partitioned in the same manner and includes data from 103 languages. For most of the languages,

Declension [edit]

Inflection of *puku* (Kotus type 1/valo, *k-v* gradation)		[more ▼]
nominative	**puku**	puvut
genitive	puvun	pukujen
partitive	pukua	pukuja
illative	pukuun	pukuihin

Fig. 3.1: The English Wiktionary partial inflection table for the Finnish word *puku*.

```
vapiti  vapiteitta      N;PRIV;PL
fasistipuolue   fasistipuolueitta       N;PRIV;PL
tasata  ette tasanneet  V;ACT;PST;NEG;IND;2;PL
kauneudenhoito  kauneudenhoito  N;NOM;SG
kysellä älkööt olko kyselleet   V;ACT;PRS;PRF;NEG;IMP;3;PL
maalaisliitto   maalaisliitolla N;AT+ESS;SG
harrastua       ei ole harrastuttu      V;PASS;PRS;PRF;NEG;IND
mukavuus        mukavuudet      N;NOM;PL
```

Fig. 3.2: A sample of the SIGMORPHON 2018 data for Finnish, scraped from Wiktionary and provided on GitHub (https://github.com/sigmorphon/conll2018/blob/master/task1/all/finnish-train-high).

data was scraped from Wiktionary. An example of some Finnish forms provided for SIGMORPHON 2018 is shown in Figure 3.2.

3.2 Text corpora

3.3 Representation of morphology

In earlier work, morphology is encoded in a language-specific and model-specific way.

(Luong et al., 2013) The UniMorph project, initially published in 2015, makes an effort to encode morphological categories uniformly cross-linguistically. (Sylak-Glassman, Kirov, Post, et al., 2015) (Sylak-Glassman, Kirov, Yarowsky, et al., 2015) (Sylak-Glassman, 2016) The SIGMORPHON data since 2018 has been published in UniMorph format.

4 Potential research directions

4.1 Research gaps

The research so far has been conducted on a fairly comprehensive set of languages, covering many language families and types of grammatical inflection. However, there are areas that stand out as gaps, many of which have been noted as potential research directions for the field.

4.1.1 Highly synthetic languages and large paradigms

Although certain highly synthetic languages such as Navajo have been modeled, the grammatical paradigms of those languages have not been modeled in their entirety. The data sources used, primarily Wiktionary, segment paradigms, and certain grammatical dimensions are either treated as lexical or simply not listed. For instance, in Wiktionary's Navajo data, only subject agreement patterns are explicitly given; tables may be listed for a few aspects, but most aspectual categories, as well as object agreement, thematic and classifier distinctions are not given. The paradigms trained in existing SIGMORPHON tasks are limited to 100-200 forms at the most, even though there are paradigms in some languages an order of magnitude or more greater. The chief challenges in using larger paradigms include producing them manually and handling the much larger sets of individual forms they entail.

4.1.2 Less paradigmatic morphology

There has also been no real exploration into less paradigmatic types of morphology, including cliticization and more compositional or derivational morphology. Highly synthetic languages such as the Inuit languages, of a type that rely more on compositional morphology and other strategies like noun incorporation, would have no way of being effectively modeled given the current scope of research.

The major hurdle in modeling less paradigmatic morphology, similar to very large paradigms, is acquiring and structuring data. Existing datasets of these types are much harder to find, and some reliance on unannotated text corpora is probably necessary. Thus, any research in these directions must concern itself with new strategies for data acquisition and labeling.

4.1.3 Transfer learning pairs

A last area is more pointedly investigating the effectiveness of transfer learning pairs. Only 79 transfer pairs have been tested, while the possibility space of transfer learning pairs is much larger than that of individual languages. A further advantage of this research question is that data already exists - it's only a matter of combining pairs of the many existing structured paradigm training sets. Since the 2019 shared task explicitly focused on related language pairs, there is room to conduct testing of unrelated pairs, and hopefully learn more about what factors other than genetic relation can make for effective transfer learning pairs. Such a research question has a real-world justification as well - many currently under-resourced languages are not closely related to major world languages, and if modeling of those languages can benefit from large existing datasets, the language technology toolkits for those language communities could see considerable improvement. Such an investigation would be aided by a dataset of typological information about the languages for which training data exists, so that regular patterns in transfer learning effectiveness can be identified. Much of this information can probably be derived from typological databases such as WALS, though the dataset may need to be manually constructed to some degree. Fortunately, it would be much smaller than most NLP datasets!

4.2 Research proposal

My choice of research direction will ultimately depend on the nature of the data that I can locate, but investigating transfer learning is the most straightforward direction from a data acquisition perspective. In the coming week, I will conduct a concerted search for useful datasets, but in the meantime I tentatively propose constructing a typological dataset specific to the SIGMORPHON 2018 languages, and using it to identify gaps in the transfer learning tests to date and explore which types of such pairs are most effective.

Bibliography

Ahlberg, Malin, Markus Forsberg, and Mans Hulden (Jan. 2015). „Paradigm classification in supervised learning of morphology“. In: pp. 1024–1029 (cit. on pp. 4, 8, 12).

Alexandrescu, Andrei and Katrin Kirchhoff (Jan. 2006). „Factored Neural Language Models.“ In: (cit. on p. 7).

Bilmes, Jeff A. and Katrin Kirchhoff (2003). „Factored Language Models and Generalized Parallel Backoff“. In: *Companion Volume of the Proceedings of HLT-NAACL 2003 - Short Papers*, pp. 4–6 (cit. on p. 7).

Cotterell, Ryan and Georg Heigold (2017). „Cross-lingual, Character-Level Neural Morphological Tagging“. In: *CoRR* abs/1708.09157. arXiv: 1708.09157 (cit. on pp. 1, 3, 4, 6, 9, 11, 12).

Cotterell, Ryan, Christo Kirov, John Sylak-Glassman, Géraldine Walther, et al. (Oct. 2018). „The CoNLL–SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection“. In: *Proceedings of the CoNLL–SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection*. Brussels: Association for Computational Linguistics, pp. 1–27 (cit. on pp. 4–6, 9).

Cotterell, Ryan, Christo Kirov, John Sylak-Glassman, David Yarowsky, et al. (Aug. 2016). „The SIGMORPHON 2016 Shared Task—Morphological Reinflection“. In: (cit. on pp. 2, 4, 5, 9, 11).

Cotterell, Ryan and Hinrich Schütze (2019). „Morphological Word Embeddings“. In: *CoRR* abs/1907.02423. arXiv: 1907.02423 (cit. on p. 7).

Dos Santos, Cıcero Nogueira and Bianca Zadrozny (2014). „Learning Character-level Representations for Part-of-speech Tagging“. In: ICML’14, pp. II-1818–II-1826 (cit. on p. 7).

Dreyer, Markus, Jason Smith, and Jason Eisner (Jan. 2008). „Latent-Variable Modeling of String Transductions with Finite-State Methods.“ In: pp. 1080–1089 (cit. on pp. 1, 3).

Durrett, Greg and John DeNero (2013). „Supervised Learning of Complete Morphological Paradigms“. In: *Proceedings of the North American Chapter of the Association for Computational Linguistics* (cit. on pp. 3, 8, 12).

English Wiktionary (n.d.). https://en.wiktionary.org. Accessed: 2019-10-25 (cit. on pp. 9, 12).

Faruqui, Manaal, Yulia Tsvetkov, Graham Neubig, and Chris Dyer (2015). „Morphological Inflection Generation Using Character Sequence to Sequence Learning“. In: *CoRR* abs/1512.06110. arXiv: 1512.06110 (cit. on pp. 2, 3, 9, 12).

Hulden, Mans, Markus Forsberg, and Malin Ahlberg (Apr. 2014). „Semi-supervised learning of morphological paradigms and lexicons“. In: *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*. Gothenburg, Sweden: Association for Computational Linguistics, pp. 569–578 (cit. on pp. 4, 8, 12).

Luong, Thang, Richard Socher, and Christopher Manning (Aug. 2013). „Better Word Representations with Recursive Neural Networks for Morphology“. In: pp. 104–113 (cit. on p. 13).

Mattissen, Johanna (2004). „A structural typology of polysynthesis“. In: *WORD* 55.2, pp. 189–216. eprint: https://doi.org/10.1080/00437956.2004.11432546 (cit. on p. 5).

McCarthy, Arya D., Ekaterina Vylomova, Shijie Wu, et al. (Aug. 2019). „The SIGMORPHON 2019 Shared Task: Morphological Analysis in Context and Cross-Lingual Transfer for Inflection“. In: *Proceedings of the 16th Workshop on Computational Research in Phonetics, Phonology, and Morphology*. Florence, Italy: Association for Computational Linguistics, pp. 229–244 (cit. on pp. 5, 6, 9).

Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig (June 2013). „Linguistic Regularities in Continuous Space Word Representations“. In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Atlanta, Georgia: Association for Computational Linguistics, pp. 746–751 (cit. on p. 7).

Nicolai, Garrett, Colin Cherry, and Grzegorz Kondrak (Jan. 2015). „Inflection Generation as Discriminative String Transduction“. In: pp. 922–931 (cit. on pp. 8, 12).

Ranta, Aarne (2008). „How predictable is Finnish morphology? an experiment on lexicon construction“. In: *Resourceful Language Technology: Festschrift in Honor of Anna Sågvall Hein*, pp. 130–148 (cit. on p. 9).

Soricut, Radu and Franz Och (Jan. 2015). „Unsupervised Morphology Induction Using Word Embeddings“. In: pp. 1627–1637 (cit. on p. 7).

Sylak-Glassman, John (2016). „The Composition and Use of the Universal Morphological Feature Schema (UniMorph Schema)“. In: (cit. on p. 13).

Sylak-Glassman, John, Christo Kirov, Matt Post, Roger Que, and David Yarowsky (2015). „A Universal Feature Schema for Rich Morphological Annotation and Fine-Grained Cross-Lingual Part-of-Speech Tagging“. In: *Systems and Frameworks for Computational Morphology*. Ed. by Cerstin Mahlow and Michael Piotrowski. Cham: Springer International Publishing, pp. 72–93 (cit. on p. 13).

Sylak-Glassman, John, Christo Kirov, David Yarowsky, and Roger Que (July 2015). „A Language-Independent Feature Schema for Inflectional Morphology“. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Beijing, China: Association for Computational Linguistics, pp. 674–680 (cit. on p. 13).