

## University of Groningen

### One Model to Rule them All

Bjerva, Johannes

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*  
2017

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Bjerva, J. (2017). *One Model to Rule them All: multitask and Multilingual Modelling for Lexical Analysis*. [Thesis fully internal (DIV), University of Groningen]. Rijksuniversiteit Groningen.

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# Summary

When learning a new skill, you take advantage of your preexisting skills and knowledge. For instance, if you are a skilled violinist, you will likely have an easier time learning to play cello. Similarly, when learning a new language you take advantage of the languages you already speak. For instance, if your native language is Norwegian and you decide to learn Dutch, the lexical overlap between these two languages will likely benefit your rate of language acquisition. This thesis deals with the intersection of learning multiple tasks and learning multiple languages in the context of Natural Language Processing (NLP), which can be defined as the study of computational processing of human language. Although these two types of learning may seem different on the surface, we will see that they share many similarities.

The traditional approach in NLP is to consider a single task for a single language at a time. However, recent advances allow for broadening this approach, by considering data for multiple tasks and languages simultaneously. This is an important approach to explore further as the key to improving the reliability of NLP, especially for low-resource languages, is to take advantage of all relevant data whenever possible. In **Part I** of this thesis, we begin with an introduction to neural networks with a focus on NLP (Chapter 2), since such architectures are particularly well suited to combined learning of multiple tasks and languages. We will then look at some ways in which neural networks can consider multiple tasks and languages at the same time

(Chapter 3). Specifically, we will consider multitask learning (MTL), and several common multilingual approaches.

In **Part II** of this thesis, I look at exploiting the fact that many NLP tasks are highly related to one another. This is done by experimenting with MTL using hard parameter sharing, which has proven beneficial for a variety of NLP tasks. In spite of such successes, however, it is not clear *when* or *why* MTL is beneficial in NLP. Chapter 4 contains a case study in which semantic tagging is shown to be beneficial for POS tagging. This further highlights the question of when MTL is beneficial in NLP tagging tasks, which is explored using information-theoretic measures in Chapter 5.

Multilingual models can leverage the fact that many languages share commonalities with one another. These resemblances can occur on various levels, with languages sharing, for instance, syntactic, morphological, or lexical features. While there are many possibilities for exploiting these commonalities, the focus in this thesis is on using multilingual word representations, as they allow for straight-forward integration in a neural network. As with MTL, it is not clear in which cases it is an advantage to *go multilingual*. In **Part III**, I begin with presenting a case study on multilingual semantic textual similarity (Chapter 6). Following this, I explore how similar languages need to be, and in which way, in order to for it to be useful to go multilingual (Chapter 7).

In **Part IV** of this thesis, I experiment with a combined paradigm, in which a neural network is trained on several languages and tasks simultaneously (Chapter 8). Finally, the thesis is concluded in **Part V** (Chapter 9). The experiments in this thesis are run on a large collection of mainly lexically oriented tasks, both semantic and morphosyntactic in nature, and on a total of 60 languages, representing a relatively wide typological range.

While traditional NLP approaches consider a single task or language at a time, the aim of this thesis was to answer several research

questions dealing with pushing past this boundary. In doing so, the hope is that in the long term, low-resource languages can benefit from the advances made in NLP which are currently to a large extent reserved for high-resource languages. This, in turn, may then have positive consequences for, e.g., language preservation, as speakers of minority languages will have a lower degree of pressure to using high-resource languages. In the short term, answering the specific research questions posed should be of use to NLP researchers working towards the same goal.

