© 2026 by Priya Anand. All rights reserved.

LEARNING SPARSE REPRESENTATIONS FOR LOW-RESOURCE SEQUENCE LABELLING

BY

PRIYA ANAND

B.S., Purdue University, 2018
M.S., University of Illinois at Urbana-Champaign, 2021

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Computer Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2026

Urbana, Illinois

Abstract

This thesis studies sparse representation learning for sequence labelling tasks in low-resource languages, where labelled training data is scarce and standard dense embeddings overfit quickly. We introduce a regularization scheme that encourages structured sparsity in the hidden layers of a recurrent encoder, and show on five typologically diverse languages that the resulting models generalize substantially better than their dense counterparts when fewer than one thousand labelled sentences are available. We further analyze the learned sparsity patterns and find they align with morphological boundaries the model was never given supervision for, suggesting the regularizer recovers structure that is genuinely present in the data rather than merely discarding capacity.

To my parents, who never once asked when I would finish.

Acknowledgments

I am deeply grateful to my advisor, Robert T. Halvorsen, for his patience and for pushing me to ask sharper questions than the ones I started with. Thanks also to my committee for their careful reading of early drafts, and to the members of the Language Processing Lab for years of useful arguments about what "low-resource" actually means. This work was supported in part by a Dissertation Completion Fellowship from the Graduate College. Finally, thanks to my family and friends for their patience during the long stretches when this thesis was the only thing I could talk about.

Table of Contents

List of Tables . . . . . **vii**

List of Figures . . . . . **viii**

List of Abbreviations . . . . . **ix**

Chapter 1 Introduction **1**

Chapter 2 Related Work **2**

Chapter 3 Model and Training Procedure **3**

Chapter 4 Experiments and Results **4**

Chapter 5 Conclusions **5**

Appendix A Hyperparameter Settings **6**

References **7**

Vita **8**

List of Tables

List of Figures

List of Abbreviations

NLP	Natural Language Processing.
POS	Part of Speech.
RNN	Recurrent Neural Network.

Chapter 1

Introduction

Sequence labelling underlies a wide range of natural language processing tasks, from part-of-speech tagging to named-entity recognition. Modern approaches rely on dense neural representations trained end to end [3], an approach that performs well given abundant labelled data but degrades sharply for the majority of the world's languages, where such data does not exist.

This thesis asks a narrower question: can a sparsity-inducing regularizer, applied to a modest recurrent architecture, close part of that gap without requiring any additional labelled data or cross-lingual supervision? We motivate the approach in Chapter 2, describe the model and training procedure in Chapter 3, and report experiments across five languages in Chapter 4.

Chapter 2

Related Work

Structured sparsity has a long history in statistical learning, predating its recent applications to neural networks by decades; early work on adaptive systems [2] already treated sparse, structured representations as a route to generalization under limited data. More recently, generative approaches to representation learning [1] demonstrated that architectural constraints, rather than raw data volume, can shape what a model learns to represent. We build on both lines of work, combining a structured sparsity penalty with a data-efficient recurrent encoder.

Chapter 3

Model and Training Procedure

We use a two-layer bidirectional LSTM encoder over subword units, with a group-lasso penalty applied to blocks of hidden units corresponding to fixed-width character windows. The penalty coefficient is tuned on a held-out set of 50 labelled sentences per language, and all remaining labelled data is reserved for the final evaluation reported in Chapter 4. Full hyperparameter settings and training curves are given in Appendix A.

Chapter 4

Experiments and Results

We evaluate on five typologically diverse languages selected for limited labelled data availability. Across all five, the sparsity- regularized model outperforms a dense baseline of matched capacity by a wide margin when fewer than 1,000 labelled sentences are available, with the gap narrowing as more data is added – consistent with the regularizer acting primarily as an inductive bias rather than an increase in effective capacity.

Chapter 5

Conclusions

The structured sparsity penalty proposed in this thesis substantially improves sequence-labelling accuracy under severe data scarcity, and the resulting sparsity patterns correlate with morphological structure the model was never told about. Future work should examine whether this correlation can be made explicit and used to guide annotation effort toward the sentences a model finds hardest to sparsify.

Appendix A

Hyperparameter Settings

All models were trained with the Adam optimizer, a learning rate of 3×10^{-4}, and early stopping on validation loss with a patience of ten epochs. The group-lasso coefficient was selected from $\{10^{-4}, 10^{-3}, 10^{-2}\}$ per language via the held-out tuning split described in Chapter 3.

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014.

[2] John H. Holland. Adaptation in natural and artificial systems. *University of Michigan Press*, 1975.

[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.

Vita

Priya Anand received her B.S. in Computer Science from Purdue University in 2018 and her M.S. from the University of Illinois at Urbana-Champaign in 2021, where she continued on to complete her Ph.D. in the Language Processing Lab.