WRITING PLAN FOR DOCTORAL THESIS

PHONOLOGICAL FEATURES IN THE ERA OF DEEP LEARNING: A MULTI-DIMENSIONAL INVESTIGATION INTO OPTIMAL REPRESENTATIONAL UNITS FOR LANGUAGE MODELING

DOCTORAL THESIS WRITING QUALIFICATION REVIEW

Sora Nagano
Graduate School of Arts and Sciences
The University of Tokyo

s-oswld-n@g.ecc.u-tokyo.ac.jp

August 22, 2025

1 Comprehensive Writing Schedule and Research Implementation Plan

The doctoral thesis "Phonological Features in the Deep Learning Era" will be completed over a structured 36-month period following a systematic three-phase approach designed to ensure rigorous empirical investigation, theoretical contribution, and practical implementation. This writing plan outlines the detailed timeline, deliverables, quality assurance mechanisms, and resource allocation strategies necessary for successful completion.

1.1 Year 1: Foundation Phase (Months 1-12)

Quarter 1 (Months 1-3): The initial quarter focuses on establishing theoretical foundations through comprehensive literature review covering three core areas: computational phonology (Jarosz, 2019; Tesar, 1995), self-supervised learning in speech (Baevski et al., 2020; Mohamed et al., 2022), and neuro-symbolic integration methodologies (Panchendrarajan & Zubiaga, 2024). During this period, I will draft the theoretical framework chapter (approximately 40 pages) synthesizing insights from symbolic phonological theory and neural representation learning. Deliverables include an annotated bibliography of 150+ sources and a position paper outlining the research gap.

Quarter 2 (Months 4-6): Implementation of baseline models and evaluation framework design constitutes the primary focus. This includes setting up Docker containers for reproducibility, implementing baseline SSL models (wav2vec 2.0, HuBERT, WavLM), and developing probing methodologies (Venkateswaran et al., 2025). The methodology chapter (30 pages) will be completed, incorporating detailed descriptions of experimental protocols and evaluation metrics. A workshop paper submission to a computational linguistics venue is planned.

Quarter 3 (Months 7-9): Conducting Study 1 on representational landscape analysis comparing SSL models (S. Chen et al., 2022; Hsu et al., 2021) with vector quantization approaches (Higy et al., 2021). This involves systematic evaluation across phonological tasks using LibriSpeech and Common Voice datasets. The first empirical chapter draft (35 pages) will document findings. Results will be presented at the departmental seminar for feedback.

Quarter 4 (Months 10-12): Beginning hybrid architecture development for Study 2, integrating Maximum Entropy Harmonic Grammar (Hayes & Wilson, 2008) with neural networks. Technical implementation documentation will be maintained using version control. Conference submission preparation for ACL or INTERSPEECH is scheduled.

1.2 Year 2: Core Development Phase (Months 13-24)

Months 13-18: Complete implementation and evaluation of hybrid neuro-symbolic architectures following principles from (Begu, 2020; J. Chen & Elsner, 2023). Conduct comprehensive cross-linguistic evaluation using typologically diverse languages from PHOIBLE database. Draft second empirical chapter (40 pages) documenting architectural innovations and performance comparisons.

Months 19-24: Execute Study 3 on cognitive plausibility using CHILDES corpus (Cruz Blandón et al., 2023) for developmental trajectory simulations. Implement ABX discrimination tasks and cross-linguistic transfer experiments. Complete third empirical chapter (35 pages). Submit journal article to Computational Linguistics or TACL.

1.3 Year 3: Consolidation Phase (Months 25-36)

Months 25-30: Conduct additional experiments addressing gaps identified through peer review. Write introduction chapter (25 pages) synthesizing theoretical foundations and general discussion chapter (30 pages) integrating findings across studies. Complete first full dissertation draft.

Months 31-36: Incorporate feedback from supervisor and committee members. Write conclusion chapter (20 pages) addressing broader impacts and future directions. Format dissertation according to university guidelines. Prepare and deliver defense presentation. Complete post-defense revisions.

1.4 Quality Assurance and Milestones

Monthly progress meetings with primary supervisor ensure consistent advancement and timely problem resolution. Quarterly committee reviews provide comprehensive feedback on theoretical development and empirical progress. Regular presentations at lab meetings and reading groups facilitate peer feedback. All code and documentation maintained through Git version control ensures reproducibility. Target milestones include 2-3 conference papers (ACL, INTERSPEECH, NeurIPS) and 1 journal article during the thesis period.

1.5 Resource Planning and Infrastructure

Computational resources include secured GPU cluster access through department allocation for large-scale experiments. Data licensing agreements established for Common Voice (Parcollet et al., 2024) and other gated datasets. Software infrastructure includes Montreal Forced Aligner (McAuliffe et al., 2017) for phonetic alignment, Kaldi toolkit for acoustic modeling, PyTorch for neural network implementation, and Hugging Face libraries for pretrained models. Planned research visits to affiliated laboratories provide specialized training and collaboration opportunities.

1.6 Expected Outcomes and Deliverables

The completed dissertation will comprise 250-300 pages organized into 8 chapters: introduction, literature review, methodology, three empirical studies, general discussion, and conclusion. Publications target high-impact venues in computational linguistics, speech processing, and machine learning. Open-source release of developed tools and models ensures broader research community impact. The research aims to establish new benchmarks for phonological representation learning and provide practical improvements for speech technology applications.

References

Baevski, A., Zhou, Y., Mohamed, A., & Auli, M. (2020). Wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems*, *33*, 12449–12460.

Begu, G. (2020). Generative adversarial phonology: Modeling unsupervised phonetic and phonological learning with neural networks. *Frontiers in Artificial Intelligence*, *3*. https://doi.org/10.3389/frai.2020.00044

Chen, J., & Elsner, M. (2023). *Exploring how generative adversarial networks learn phonological representations* (arXiv:2305.12501). arXiv. https://doi.org/10.48550/arXiv.2305.12501

Chen, S., Wang, C., Chen, Z., Wu, Y., Liu, S., Chen, Z., Li, J., Kanda, N., Yoshioka, T., Xiao, X., Wu, J., Zhou, L., Ren, S., Qian, Y., Qian, Y., Wu, J., Zeng, M., Yu, X., & Wei, F. (2022). WavLM: Large-scale self-supervised pre-training for full stack speech processing. *IEEE Journal of Selected Topics in Signal Processing*, *16*(6), 1505–1518. https://doi.org/10.1109/JSTSP.2022.3188113

Cruz Blandón, M. A., Cristia, A., & Räsänen, O. (2023). Introducing meta-analysis in the evaluation of computational models of infant language development. *Cognitive Science*, *47*(7), e13307. https://doi.org/10.1111/cogs.13307

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, *39*(3), 379–440. https://doi.org/10.1162/ling.2008.39.3.379

Higy, B., Gelderloos, L., Alishahi, A., & Chrupaa, G. (2021). Discrete representations in neural models of spoken language. In J. Bastings, Y. Belinkov, E. Dupoux, M. Giulianelli, D. Hupkes, Y. Pinter, & H. Sajjad (Eds.), *Proceedings of the fourth BlackboxNLP workshop on analyzing and interpreting neural networks for NLP* (pp. 163–176). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.blackboxnlp-1.11

Hsu, W.-N., Bolte, B., Tsai, Y.-H. H., Lakhotia, K., Salakhutdinov, R., & Mohamed, A. (2021). *HuBERT: Self-supervised speech representation learning by masked prediction of hidden units* (arXiv:2106.07447). arXiv. https://doi.org/10.48550/arXiv.2106.07447

Jarosz, G. (2019). Computational modeling of phonological learning. *Annual Review of Linguistics*, *5*(1), 67–90. https://doi.org/10.1146/annurev-linguistics-011718-011832

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal forced aligner: Trainable text-speech alignment using kaldi. *Proceedings of the 18th Conference of the International Speech Communication Association (Interspeech)*, 498–502.

Mohamed, A., Lee, H., Borgholt, L., Havtorn, J. D., Edin, J., Igel, C., Kirchhoff, K., Li, S.-W., Livescu, K., Maaløe, L., Sainath, T. N., & Watanabe, S. (2022). Self-supervised speech representation learning: A review. *IEEE Journal of Selected Topics in Signal Processing*, *16*(6), 1179–1210. https://doi.org/10.1109/JSTSP.2022.3207050

Panchendrarajan, R., & Zubiaga, A. (2024). *Synergizing machine learning & symbolic methods: A survey on hybrid approaches to natural language processing* (arXiv:2401.11972). arXiv. https://doi.org/10.48550/arXiv.2401.11972

Parcollet, T., Nguyen, H., Evain, S., Boito, M. Z., Pupier, A., Mdhaffar, S., Le, H., Alisamir, S., Tomashenko, N., Dinarelli, M., Zhang, S., Allauzen, A., Coavoux, M., Esteve, Y., Rouvier, M., Goulian, J., Lecouteux, B., Portet, F., Rossato, S., … Besacier, L. (2024). *LeBenchmark 2.0: A standardized, replicable and enhanced framework for self-supervised representations of french speech* (arXiv:2309.05472). arXiv. https://doi.org/10.48550/arXiv.2309.05472

Tesar, B. B. (1995). *Computational optimality theory* [PhD thesis]. University of Colorado at Boulder.

Venkateswaran, N., Tang, K., & Wayland, R. (2025). *Probing for phonology in self-supervised speech representations: A case study on accent perception* (arXiv:2506.17542). arXiv. https://doi.org/10.48550/arXiv.2506.17542