Research Directions and Open Questions

1 Key Questions

1. What is a restrictive theory (less than FO) of long-distance transformations?
2. How can generalizations expressed logically (both constraints and transformations) be learned from examples?
3. What are the advantages and disadvantages of different linguistic representations?
4. How do insights like ISL help us conduct an automatic phonological analysis (identify URs and transformation)?
5. What are the morpho-phonological generalizations of an understudied language / phenomenon?
6. Can any of the previous research which partially provide some answers to these questions be coded and implemented?a

2 Long-distance Transformations

We have multiple characterizations for long-distance constraints based on precedence and successor on tiers that are below FO (PROP and CNL). However, we don't have yet characterizations of long-distance transformations based on precedence or tier-successor. What would these look like? How can they be learned? What kind of long-distance phenomena can they describe and not describe? Can we use weighted logic?

Some recent efforts in this directon include work on tier-strictly local transformations (Burness and McMullin, 2019) and least-fixed-point logics with successor (Chandlee and Jardine, 2019).

3 Learning

There is some recent work on learning constraints over arbitrary relational structures (Strother-Garcia et al., 2016; Vu et al., 2018; Chandlee et al., 2019). However, this work is theoretical and has not been implemented or tested on artificial or naturalistic corpora. Also, this work has not been extended to transformations or to weighted logics.

4 Representations

Previous research exists on:

1. Tone and autosegmental representations (Jardine, 2016, 2017)
2. Syllable structure (Strother-Garcia, 2018, 2019)
3. Phonological tiers (Heinz et al., 2011; Rogers and Lambert, 2019b, a.o.)

Those areas would benefit from additional study.

1. Complex Segments (long consonants & vowels, diphthongs, affricates, etc)
2. Features, binary vs unary vs scalar vs equipollent
3. Stress and Rhythm (grid vs feet vs neither; see also (Rogers and Lambert, 2019a))
4. Morpho-phonological interface (Dolatian, 2020)
5. Phonetics-phonology interface (some work in DCP)

5 Automating Morpho-phonological Analysis

One of the long-term goals has been to automate linguistic analysis. Given a morpho-phonological data set, of the kind we might see in textbooks, can we write a program that automatically identifies a lexicon of underlying forms, a morphological grammar (describing how affixation or other morphology present is performed), and a phonological grammar. There are some exciting developments at Rutgers on this front, and it would be great to get involved.

6 Understudied Languages & Phenomena

It is my belief that the kinds of logical grammars we have been writing offer unparallelled benefits in descriptive research. You can choose the representations you want. You describe precisely the generalizations you think are important. You can use something general (like FO or MSO) *because once written down other researchers can translate your analyses into other representations or check to see if they can be expressed with weaker logics automatically.*

Many of you are interested or working on understudied languages and phenomena. Dr. Neda Taherkhani (our department's Elahé Omidyar Mir-Djalali Postdoctoral Fellow) is a native speaker of Southern Tati, an understudied Indo-European language, and is working on aspects of its morpho-phonology and is open to collaboration.

7 Implementing Toolkits to Facilitate Research

We know how to write transformations from structures of model signature to another structures of a different model signature. But a toolkit faciltating this, and the translations between analyses that it would allow, has not been implemented as far as I know. Such a toolkit would allow one to translate an analysis given in FO(precedence,letters) to MSO(successor,letters) and FO(successor,features) to FO(successor,letters) and so on. It would allow us to implement the morpho-phonological analysis we write for a given phenomenon. It could be extended to translate logical analyses into finite-state transducers. It could be extended or interface with learning programs. It needs to be developed!!

References

Burness, P. and K. McMullin (2019, July). Efficient learning of output tier-based strictly 2-local functions. In *Proceedings of the 16th Meeting on the Mathematics of Language*, Toronto, Canada, pp. 78–90. Association for Computational Linguistics.

Chandlee, J., R. Eyraud, J. Heinz, A. Jardine, and J. Rawski (2019, 18–19 July). Learning with partially ordered representations. In *Proceedings of the 16th Meeting on the Mathematics of Language*, Toronto, Canada, pp. 91–101. Association for Computational Linguistics.

Chandlee, J. and A. Jardine (2019, July). Quantifier-free least fixed point functions for phonology. In *Proceedings of the 16th Meeting on the Mathematics of Language*, Toronto, Canada, pp. 50–62. Association for Computational Linguistics.

Dolatian, H. (2020). *Computational locality of cyclic phonology in Armenian.* Ph. D. thesis, Stony Brook University.

Heinz, J., C. Rawal, and H. G. Tanner (2011, June). Tier-based strictly local constraints for phonology. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, Oregon, USA, pp. 58–64. Association for Computational Linguistics.

Jardine, A. (2016). *Locality and non-linear representations in tonal phonology*. Ph. D. thesis, University of Delaware.

Jardine, A. (2017). The local nature of tone-association patterns. *Phonology 34*, 385–405.

Rogers, J. and D. Lambert (2019a). Extracting subregular constraints from regular stringsets. *Journal of Language Modelling 7*(2), 143–176.

Rogers, J. and D. Lambert (2019b, July). Some classes of sets of structures definable without quantifiers. In *Proceedings of the 16th Meeting on the Mathematics of Language*, Toronto, Canada, pp. 63–77. Association for Computational Linguistics.

Strother-Garcia, K. (2018). Imdlawn Tashlhiyt Berber syllabification is quantifier-free. In *Proceedings of the Society for Computation in Linguistics*, Volume 1. Article 16.

Strother-Garcia, K. (2019). *Using Model Theory in Phonology: A Novel Characterization of Syllable Structure and Syllabification*. Ph. D. thesis, University of Delaware.

Strother-Garcia, K., J. Heinz, and H. J. Hwangbo (2016, October). Using model theory for grammatical inference: a case study from phonology. In S. Verwer, M. van Zaanen, and R. Smetsers (Eds.), *Proceedings of The 13th International Conference on Grammatical Inference*, Volume 57 of *JMLR: Workshop and Conference Proceedings*, pp. 66–78.

Vu, M. H., A. Zehfroosh, K. Strother-Garcia, M. Sebok, J. Heinz, and H. G. Tanner (2018, July). Statistical relational learning with unconventional string models. *Frontiers in Robotics and AI 5*(76), 1–26.