

# Kenneth Steimel

---

## Education

2015–Present **Doctor of Philosophy**, *Indiana University*, Bloomington, 3.9.

Computational Linguistics

- Thesis: Swahili in the Universal Dependencies Framework
- Computer Science Minor

2011–2015 **Bachelor of Arts**, *University of Missouri*, Columbia, 3.9.

Linguistics – Summa Cum Laude with Honors

- Honors Thesis on Wanga Noun Tone
- Computer Science Minor

## Skills

Programming	Python, Julia, C, R, Powershell	NLP	Parsing, Sentiment analysis, Machine learning
OS	Linux, Windows	Markup Languages	T <sub>E</sub> X, HTML, XML, TEI
High Performance Computing	Practical knowledge of programming for a HPC system (MPI, OpenMP, Spark, Hadoop)	Machine Learning	Scikit-learn, Keras, Pytorch, Allennlp

## Experience

### Vocational

2021–Present **Assistant Research Engineer**, *Educational Testing Service*, Princeton, New Jersey.

- Developing interpretable neural content scoring models.
- Adhering to an AGILE research environment.

2020–2020 **Associate Instructor**, *Indiana University*, Bloomington, Indiana.

- Solo taught *Syntax: Grammar as Science*
- Developed lessons and assignments guiding students through the grammar creation process.

2020–2020 **X-Force Fellow**, *Department of Defense*, Virginia.

- Extracted large datasets from the web of science database for technology scouting.
- Cleaned extracted data using OpenRevise.
- Transformed data to produce burstiness visualizations (Kleinberg 2002).
- Trained and used topic modelers to gauge breadth of impact for highly cited publications.

2551 South Cargill Drive – Bloomington, Indiana – United States of America

📞 +1 (314) 775 4101 • ✉ ksteimel@outlook.com • 🌐 steimel.info

in Kenneth.Steimel • 🌐 ksteimel

- 2019–2020 **Intern**, *Educational Testing Service*, Princeton, New Jersey.
- Developed neural content scoring systems using transformer models.
  - Developed meta-learning infrastructure in the Allennlp framework.
    - MAML and Reptile integrated into a new Allennlp trainer.
  - Performed analysis to determine properties of ASAP-SAS prompts (average number of words, type-token ratio etc.)
  - Worked with ETS infrastructure including Sun Grid Engine, and in-house hyper parameter tuning modules.
- 2018–2019 **Associate Instructor**, *Indiana University*, Bloomington, Indiana.
- Led Friday lectures for **Introduction to the Study of Language**
  - Designed activities for teaching
  - Prepared a lecture on Computational Linguistics for an introductory audience.
- 2015–2018 **Information Technology Services Temp**, *Panera Bread*, St.Louis, Missouri.
- Development of scripts and platforms for integration of Hadoop into existing database ecosystem
  - Coordination of capacity planning data with Big Data consulting companies:
    - Creation of scripts to automate data flow into clusters using platform independent solutions
    - Preprocessing of data for machine-learning
  - Automation of server stand-up using tcl, bash and powershell.

### Miscellaneous Vocational Experience

- 2016–2018 **Student Editor**, *The LINGUIST List*, Bloomington, Indiana.  
I edited Calls for Papers, Conferences, Queries and other smaller posting areas for the website. In addition, I was involved with LINGUIST List projects like the LFG parser and the redesign of the website.
- 2011–2015 **Draftsman**, *Cardinal Surveying & Mapping*, Cottleville, Missouri.  
Drafted boundary surveys in Carlson Surveying software using raw data points and public land records.

## Publications

### Conference Proceedings

- Kenneth Steimel, Akbar Amat, Arienne Dwyer & Sandra Kübler. Fine-Grained Morpho Syntactic Analysis for the Under-Resourced Language Chaghatay. *Treebanks and Linguistic Theory (TLT 2020)*.
- Kenneth Steimel & Brian Riordan. Towards Instance-Based Content Scoring with Pre-trained transformer models. *Accepted submissions to the Workshop on Artificial Intelligence for Education at AAAI 2020, New York, New York*.
- Kenneth Steimel, Daniel Dakota, Yue Chen & Sandra Kübler. (2019). Investigating Multilingual Abusive Language Detection: A Cautionary Tale. *Proceedings of Recent Advances in Natural Language Processing (RANLP 2019), Varna, Bulgaria* (pp. 1151-1160).
- Kenneth Steimel. (2018). Part of Speech Tagging in Luyia: A Bantu Macrolanguage. *Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2018), Santa Fe, New Mexico* (pp. 46-54).
- Yue Chen, Kenneth Steimel, Everett Green, Nils Hjortnaes, Zuoyu Tian, Daniel Dakota, Sandra & Kübler. (2018). Towards Determining Textual Characteristics of High and Low Impact Publications. In Diesner, Jana, Rehm, Georg, Witt, Andreas (Eds.): *Proceedings of the 1st Workshop on Computational Impact Detection from Text Data, Miyazaki, Japan* (pp. 1–7). Paris, France:

2551 South Cargill Drive – Bloomington, Indiana – United States of America

📞 +1 (314) 775 4101 • ✉ ksteimel@outlook.com • 🌐 steimel.info

in Kenneth.Steimel • 🌐 ksteimel

European language resources association (ELRA).

- Damir Ćavar, Lwin Moe, Hai Hu & Kenneth Steimel. (2016). Preliminary results from the Free Linguistic Environment project. In Arnold, Doug, Butt, Miriam, Crysmann, Berthold, King, Tracy Holloway, & Müller, Stefan (Eds.): *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar, Polish Academy of Sciences, Warsaw, Poland* (pp. 161–181). Stanford, CA: CSLI Publications.
- Can Liu, Wen Li, Bradford Demarest, Yue Chen, Sara Couture, Daniel Dakota, Nikita Haduong, Noah Kaufman, Andrew Lamont, Manan Pancholi, Kenneth Steimel & Sandra Kübler. (2016). IUCL at SemEval-2016 Task 6: An Ensemble Model for Stance Detection in Twitter. *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), Association for Computational Linguistics, San Diego, California* (pp. 394–400). Stroudsburg, PA: Association for Computational Linguistics (ACL).

#### In review

- Hyunji Hayley Park, Coleman Hayley, Kenneth Steimel, Han Liu & Lane Schwarz. Morphology Matters: a Multilingual Language Modeling Analysis.

#### Preprints

- Lane Schwartz, Francis Tyers, Lori Levin, Christo Kirov, Patrick Littell, Chi-kiu Lo, Emily Prud'hommeaux, Hyunji Hayley Park, Kenneth Steimel, Rebecca Knowles, Jeffrey Micher, Lonny Strunk, Han Liu, Coleman Haley, Katherine J. Zhang, Robbie Jimmerson, Vasilisa Andriyanets, Aldrian Obaja Muis, Naoki Otani, Jong Hyuk Park, Zhisong Zhang. (2020). Neural Polysynthetic Language Modelling. *arXiv Preprint*.

## Projects

- I have a cluster of five servers that I run out of my house (homelab).
  - Use software defined storage (similar to the Gluster storage used at ETS)
  - Run a variety of virtual machines and containers for various open source servers as well as my own projects.
- I developed tools for working with Context Free Grammars (written in Julia) and used this to create a website for developing grammars using phrase structure rules (available at <https://parser.steimel.info>).
  - Earley parser with optionality and repetition.
  - Generation and parsing supported.
  - Tree diagrams created from scratch using Cairo compatibility package in Julia.
  - Test Driven Development and version control utilized heavily.

## References

### Academic References

- Sandra Kübler  
skuebler@indiana.edu  
*Indiana University*
- Damir Ćavar  
dcavar@indiana.edu  
*Indiana University*

### Professional References

- Brian Riordan  
bwrriordan@ets.org  
*Educational Testing Services*
- Amy Cooper  
amy.cooper@navy.mil  
*NSWC Crane*  
(more upon request)

2551 South Cargill Drive – Bloomington, Indiana – United States of America

📞 +1 (314) 775 4101 • ✉ ksteimel@outlook.com • 🌐 steimel.info

in Kenneth.Steimel • 🌐 ksteimel

**Institute for Digital Arts and Humanities**

*Herman B. Wells Library E170*

*1320 E. 10th Street*

*Bloomington, IN 47405*

April 15, 2020

To whom it may concern,

I am applying for a student academic appointment at the IDAH center. I am a PhD student in Computational Linguistics. My technical qualifications as they relate to digital humanities research are very strong: I am very proficient in programming, visualization, and machine learning. In addition, I have the tools to explore both the arts and humanities sides of IDAH: I work on corpus analysis, 3D and 2D drafting and 3D printing. These technical qualifications give me a palette of experience to pull from when consulting on projects or teaching through the use of technology. However, I would like to focus on my pedagogical qualifications in this letter.

I will discuss two cases where my technical aptitude augmented my teaching. While working as an associate instructor for **Introduction to the Study of Language**, I taught in a digital learning classroom in the Global and International Studies building. This building allowed for the unique opportunity to meld digital pedagogy and student involvement. I set up activities where students had to find videos showing imaging of the vocal tract to share with the class. Students then explained the pieces of the vocal tract being viewed in the video and shared one thing they found particularly interesting about the way the speech apparatus moved. Because this was at the beginning of the semester, this worked as a sort of icebreaker. The way a vocal tract moves really gets a lot of students excited. For example, many never realized that there's so much movement in the tongue. Showing off the video they found and explaining why it was interesting got them excited in the class but also has them see their peers as people or friends rather than faces.

To enable teaching of syntax for the advanced introduction to linguistics class, I built a website that allows for students to enter their own grammar rules into a text box, generate random sentences from these rules and display parse trees for sentences they provide (or explain where there are errors in their grammar if the sentences they provide cannot be parsed). This website is still up and running at <https://parser.duckdns.org>. This website allowed students to observe the generative principle of grammar design taught in linguistics in a very tangible sense. The generative principle says that the grammar of a language should generate all the sentences that are acceptable in a language and only sentences that are acceptable. By getting instant feedback for the sentences that their grammar can produce, they can see the connection between the grammar you write and the language you describe.

I have discussed two examples of how I have integrated digital media into teaching. While I may not have broad experience with pedagogy and digital media (both examples provided concern linguistics), I am very open to learning how to apply my technical abilities to other areas of study.

Sincerely,

**Kenneth Steimel**

**Kenneth Steimel**

*2551 South Cargill Drive – Bloomington, Indiana – United States of America*

📞 +1 (314) 775 4101 • ✉ [ksteimel@outlook.com](mailto:ksteimel@outlook.com) • 🌐 [steimel.info](http://steimel.info)

**in** [Kenneth.Steimel](https://www.linkedin.com/in/Kenneth.Steimel) • [ksteimel](https://github.com/ksteimel)

4/5